

## DIFFERENCE WAVELET — THEORY AND A COMPARISON STUDY \*

I-LIANG CHERN<sup>†</sup> AND CHIEN-CHANG YEN<sup>‡</sup>

**Abstract.** Wavelet methods with polynomial filters are usually favored in applications for their fast wavelet transforms and compact support. However, wavelet methods with rational filters have more freedom to achieve smaller condition numbers, more regularity and better efficiency. Such methods can be attractive if they also possess fast algorithms and have fast decay (as if the corresponding wavelets had compact support). In the first part of this paper, we propose a new wavelet method with rational filters which do have these properties. We call it the difference wavelet method. It is a generalization of Butterworth wavelets. The analysis part is simply averaging and finite differencing. The wavelet coefficients measure the finite differences of the averages of an input data sequence. Its synthesis part involves rational filters, which can be performed with linear computational complexity by the cyclic reduction method. Their Riesz basis property, biorthogonality, decay and regularity are investigated.

In the second part of this paper, we perform comparison studies of the difference wavelet method (Diff) with three other popular wavelet methods: the Cohen-Daubechies-Feauveau biorthogonal wavelets (CDF), the Daubechies orthogonal wavelets (Daub) and the Chui-Wang semi-orthogonal wavelets (CW). Natural criteria in designing good wavelet methods for representing functions and operators are *speed*, *stability* and *efficiency*. Therefore, the items of our first comparison include (i) operation counts for performing transformations, (ii) condition numbers of the wavelet transformations, (iii) compression ratios, by some numerical experiments, for representing (smooth or non-smooth) data sequences and matrices (smooth or non-smooth kernels). The results show that (i) Diff, Daub and CDF have about the same operation counts, and CW has more; (ii) Diff has about the same condition numbers as those of CDF and CW; (iii) Diff has better compression ratio for both (smooth or non-smooth) data sequences and matrices (smooth or non-smooth kernels).

The items of our second comparison include regularity, approximation power (the constant in the approximation estimate), approximation errors for non-smooth functions (where Gibbs phenomena appear) and the “essential supports.” The results show that Diff has better regularity and better approximation ability with only slightly bigger essential supports. It is evident that the better efficiency of Diff for smooth functions is due to its regularity. It is surprising that, even for non-smooth functions, Diff is comparable to, sometimes even superior to, other methods, despite its infinite-support property.

This paper is organized as follows. Sec. 1 is preliminary. Sec. 2 provides the theory of the difference wavelet method. Sec. 3 contains the comparison studies. Experts are suggested to read Sec. 3 directly.

**1. Preliminary.** A wavelet expansion method decomposes data (or functions) into fluctuations at various resolutions. It depends on four sets of filter coefficients:  $\{h_k\}_{k \in \mathbb{Z}}$ ,  $\{g_k\}_{k \in \mathbb{Z}}$ ,  $\{\tilde{h}_k\}_{k \in \mathbb{Z}}$  and  $\{\tilde{g}_k\}_{k \in \mathbb{Z}}$ . The first two are called the analysis filters, and the latter two the synthesis filters. A data sequence  $c_j = \{c_{j,i}\}_{i \in \mathbb{Z}}$  at resolution level  $j$  can be decomposed, through the analysis filters, into two sets of data sequences at level  $j - 1$ :

$$(1.1) \quad \begin{cases} \text{the low-pass data: } c_{j-1,i} = \sqrt{2} \sum_k h_k c_{j,2i-k}, \\ \text{the high-pass data: } d_{j-1,i} = \sqrt{2} \sum_k g_k c_{j,2i+1-k}. \end{cases}$$

Here,  $\sqrt{2}$  is a normalized scale factor [12]. By applying the transform:  $c_j \mapsto (c_{j-1}, d_{j-1})$  recursively for  $j = J, J - 1, \dots, 1$ , one can decompose a given data sequence  $c_J$  at the finest resolution  $J$  into  $(c_0, d_0, d_1, \dots, d_{J-1})$ , the averages at coarsest

---

\*Received May 10, 2001; accepted for publication December 10, 2001. This work was supported by the National Science Council of the Republic of China under Contract NSC86-2115-M002-008.

<sup>†</sup>Department of Mathematics, National Taiwan University, Taipei, Taiwan (chern@math.ntu.edu.tw).

<sup>‡</sup>Institute of Astronomy and Astrophysics, Academia Sinica, Condensed Matter Science and Physics Building, National Taiwan University, Taipei, Taiwan 106 (yen@asiaa.sinica.edu.tw).

resolution and the fluctuations at various resolutions. The mapping  $T_J : c_J \mapsto (c_0, d_0, d_1, \dots, d_{J-1})$  is called a discrete wavelet transformation.

The data sequence  $c_j$  can be reconstructed from  $(c_{j-1}, d_{j-1})$  through the synthesis filters  $\{\tilde{h}_k\}_{k \in \mathbb{Z}}$  and  $\{\tilde{g}_k\}_{k \in \mathbb{Z}}$  by:

$$(1.2) \quad c_{j,i} = \sqrt{2} \sum_k \left[ \tilde{h}_{i-2k} c_{j-1,k} + \tilde{g}_{i-2k-1} d_{j-1,k} \right].$$

By applying this inverse transform recursively for  $j = 1, \dots, J-1$ ,  $c_J$  can be recovered from  $(c_0, d_0, \dots, d_{J-1})$ .

Let us define the  $z$ -transform of  $\{h_k\}_{k \in \mathbb{Z}}$  by  $h(z) = \sum_k h_k z^k$ , and still call it a filter. It can be a polynomial (i.e. a filter with finite length), or a rational function (infinite length), or in general, a Laurent series.

In order to have  $c_j$  be reconstructed from (1.2), the filter bank (i.e. the collection of analysis and synthesis filters) needs to satisfy the following perfect reconstruction condition [9, 16]:

$$(1.3) \quad h(z)\tilde{h}(z) + g(z)\tilde{g}(z) = 1,$$

$$(1.4) \quad h(-z)\tilde{h}(z) - g(-z)\tilde{g}(z) = 0.$$

A formal calculation gives

$$(1.5) \quad h(z)\tilde{h}(z) + h(-z)\tilde{h}(-z) = 1,$$

$$(1.6) \quad g(z) = \tilde{h}(-z)P(z^2),$$

$$(1.7) \quad \tilde{g}(z) = h(-z)/P(z^2),$$

for some Laurent series  $P(z)$ . If both analysis and synthesis filters are polynomials, then  $P(z) = z^m$  for some integer  $m$ . In this case, we can normalize to  $P(z) = 1$  [9]. In general, for stability consideration, we should choose  $P(z)$  to be in the Wiener class, that is,  $P(z) = \sum_k p_k z^k$  with  $\sum_k |p_k| < \infty$  and  $P(z) \neq 0$  for all  $|z| = 1$  (see Chui [3]).

So, a general procedure to construct filter banks is to find  $h(z)$  and  $\tilde{h}(z)$  to satisfy (1.5). We may normalize them by

$$(1.8) \quad h(1) = \tilde{h}(1) = 1.$$

Then we define  $g(z)$  and  $\tilde{g}(z)$  by (1.6) and (1.7) with a proper function  $P(z)$  chosen in the Wiener class.

A general principle to design filters in applications is to have the wavelet transform to be *fast*, *stable* and the wavelet approximation to be *efficient*. The term “fast” means that the wavelet transform  $T_J$  and its inverse are of linear computational complexity. Usually, polynomial filters are favored. However, a rational filter which can be performed with linear complexity is also acceptable in many applications. The term “stable” means that the forward and inverse wavelet transforms are unconditionally stable in  $\ell^2$ , independent of the resolution level  $J$ , that is, both  $\|T_J\|, \|T_J^{-1}\| = O(1)$ . The term “efficient” means that only a small amount of wavelet coefficients  $d_{j,i}$  plus the averages  $c_{0,i}$  are sufficient to approximate the original data accurately.

It is well-known that the stability condition can be characterized by the Riesz basis property of a wavelet function  $\psi$ , see [9, 10]. More precisely, associated with the

analysis filter bank, one can define a refinable function (or called scaling function)  $\phi$  and a wavelet function  $\psi$  as follows:

$$(1.9) \quad \phi(x) = 2 \sum_k h_{-k} \phi(2x - k),$$

$$(1.10) \quad \psi(x) = 2 \sum_k g_{1-k} \phi(2x - k).$$

The function  $h(z)$  is called the mask of the refinable function  $\phi$ . Let us define  $\psi_{j,i}(\cdot) = 2^{j/2} \psi(2^j \cdot -i)$ . Then the forementioned stability condition is equivalent to that  $\{\psi_{j,i}\}_{i,j \in \mathbb{Z}}$  forms a Riesz basis in  $L^2(\mathbb{R})$ . That is, there exist constants  $0 < \gamma, \Gamma < \infty$  such that for any sequence  $\{d_{j,i}\}_{i,j \in \mathbb{Z}}$ , we have

$$\gamma \sum_{i,j \in \mathbb{Z}} |d_{j,i}|^2 \leq \left\| \sum_{i,j \in \mathbb{Z}} d_{j,i} \psi_{j,i} \right\|_{L^2}^2 \leq \Gamma \sum_{i,j \in \mathbb{Z}} |d_{j,i}|^2.$$

For the synthesis filter bank, one also defines

$$(1.11) \quad \tilde{\phi}(x) = 2 \sum_k \tilde{h}_k \tilde{\phi}(2x - k),$$

$$(1.12) \quad \tilde{\psi}(x) = 2 \sum_k \tilde{g}_{k-1} \tilde{\phi}(2x - k).$$

Then  $\{\tilde{\psi}_{j,i}\}_{i,j \in \mathbb{Z}}$  forms the dual Riesz basis of  $\{\psi_{j,i}\}_{i,j \in \mathbb{Z}}$  in  $L^2(\mathbb{R})$ , i.e.  $(\psi_{j,i}, \tilde{\psi}_{\ell,k}) = \delta_{j,\ell} \delta_{i,k}$ . These two wavelets are called biorthogonal [9]. It was commented by Cohen-Daubechies-Feauveau [9] and Dahmen [10] that biorthogonal wavelets are more flexible to use than orthogonal wavelets in practice.

Three popular wavelets are the Daubechies' orthogonal wavelet (Daub) [11], Cohen-Daubechies-Feauveau's biorthogonal wavelet (CDF) [9] and Chui-Wang's semi-orthogonal wavelet (CW) [4]. The filter bank of the CDF wavelet is defined as follows:

$$(1.13) \quad h(z) = \tilde{g}(-z) = z^{-[r/2]} \left( \frac{1+z}{2} \right)^r,$$

$$(1.14) \quad \tilde{h}(z) = g(-z) = z^{-[(\tilde{r}+1)/2]} \left( \frac{1+z}{2} \right)^{\tilde{r}} Q_K(z).$$

Here,  $r + \tilde{r} = 2K$ ,  $K > 0$  is an integer parameter, and

$$(1.15) \quad Q_K(z) = \sum_{n=0}^{K-1} \binom{K-1+n}{n} \left( \frac{2-z-z^{-1}}{4} \right)^n.$$

We call the parameter  $r$  the averaging order and  $\tilde{r}$  the differencing order, or the number of vanishing moments.

The filter bank of Daubechies' orthogonal wavelet is defined by

$$(1.16) \quad h(z) = \tilde{h}(z^{-1}) = \tilde{g}(-z) = g(-z^{-1}) = \left( \frac{1+z}{2} \right)^K Q(z),$$

where  $Q(z)Q(z^{-1}) = Q_K(z)$ . The filter bank of Chui-Wang's semi-orthogonal wavelet is defined by

$$\begin{aligned} h(z) &= \left(\frac{1+z}{2}\right)^K G_K(z)/G_K(z^2), \\ g(z) &= \left(\frac{1-z}{2}\right)^K /G_K(z^2), \\ \tilde{h}(z) &= \left(\frac{1+z}{2}\right)^K, \\ \tilde{g}(z) &= \left(\frac{1-z}{2}\right)^K G_K(-z), \end{aligned}$$

where  $G_K(z) := \sum_k N_{2K}(k)z^k$  and  $N_{2K}$  is the B-spline of order  $2K - 1$ , that is the  $2K$  times self convolution of the step function  $1_{[0,1]}$ .

It is well-known that the regularity of a wavelet is related to its orders  $r, \tilde{r}$  and the magnitude of its amplification factor (i.e.  $Q_K(e^{i\xi})$  in CDF wavelet and  $G_K(e^{i\xi})/G_K(e^{i2\xi})$  in CW wavelet) [12]. The more regularity a wavelet has, the better approximation ability it can have [12]. We can look for rational filters that have smaller amplification factors. The gain is that the wavelet function is smoother. The prices to pay are that (i) it has infinite support, and (ii) a linear system needs to be solved to perform a wavelet transform. Problem (i) is not severe if the essential support (e.g. the region where  $|\tilde{\phi}| > \epsilon$ ) is still small. Problem (ii) is also solvable if a fast algorithm for solving this linear system is available. Below, we propose the difference wavelet method which has a relatively small amplification factor.

**2. Difference Wavelets.**

**2.1. The filter bank of the difference wavelet method.** Given a positive integer  $K$ , let us define the filter bank of the difference wavelet method to be

(2.1) 
$$h(z) = z^{-\lfloor \frac{r}{2} \rfloor} \left(\frac{1+z}{2}\right)^r,$$

(2.2) 
$$g(z) = z^{-\lfloor \frac{\tilde{r}+1}{2} \rfloor} \left(\frac{1-z}{2}\right)^{\tilde{r}},$$

(2.3) 
$$\tilde{h}(z) = z^{-\lfloor \frac{\tilde{r}+1}{2} \rfloor} \left(\frac{1+z}{2}\right)^{\tilde{r}} / P_K(z^2),$$

(2.4) 
$$\tilde{g}(z) = z^{-\lfloor \frac{r}{2} \rfloor} \left(\frac{1-z}{2}\right)^r / P_K(z^2),$$

(2.5) 
$$P_K(z^2) = z^{-K} \left(\frac{1+z}{2}\right)^{2K} + (-z)^{-K} \left(\frac{1-z}{2}\right)^{2K},$$

where  $r + \tilde{r} = 2K$ . Roughly speaking, the difference of this method from the CDF method is to replace the filter  $Q_K(z)$  by  $1/P_K(z^2)$ . We call this method the difference wavelet method because its high-pass filter is simply a finite difference. In the case of  $r = 0$  and  $\tilde{r} = 2K$ ,  $\tilde{h}(z)$  is the Butterworth filter, which was well-known in the field of signal processing [15]. To justify this method to be valuable, we shall show that

1. the operation count to perform  $1/P_K(z^2)$  is almost the same as that of  $Q_K(z)$ ;

2. it is stable in  $L^2(\mathbb{R})$ ; in fact the condition numbers of the difference wavelet transforms are comparable to those of the CDF transform;
3. the corresponding  $\tilde{\phi}$  decays exponentially at far field, in fact, the lengths of the “essential supports” of the difference wavelets are about twice those of the CDF wavelets;
4. it is more efficient in the sense that it has better approximation ability and better compression ratio.

**2.2. Fast algorithm for difference wavelet transform.** We adopt the cyclic reduction method [14] to perform the filter  $1/P_K(z^2)$ . First, we factor  $P_K(z^2)$  into

$$\begin{aligned} P_K(z^2) &= \prod_{k=1}^{[K/2]} \frac{1}{1+2\alpha_k} (\alpha_k z^{-2} + 1 + \alpha_k z^2) \\ (2.6) \quad &\equiv \prod_{k=1}^{[K/2]} P^k(z^2). \end{aligned}$$

Here,

$$(2.7) \quad 0 < \alpha_k = 1/(\tan^2 \theta_k + 1/\tan^2 \theta_k) < 1/2,$$

$$(2.8) \quad \theta_k = \begin{cases} \frac{(2k-1)\pi}{4K} & \text{if } K \text{ is even,} \\ \frac{k\pi}{2K} & \text{if } K \text{ is odd,} \end{cases}$$

where  $k = 1, \dots, 2K$  for even  $K$ , and  $k = 1, \dots, K-1, K+1, \dots, 2K-1$  for odd  $K$ . This factorization can be derived from the fact that  $z = i \tan \theta_k$  are the roots of  $P_K(z^2) = 0$ .

Secondly, for each  $k = 1, \dots, [K/2]$ , we perform the filter  $1/P^k(z^2)$  by solving a tridiagonal system  $Av = f$ , where  $A = \text{diag}(\alpha_k, 1, \alpha_k)$  with  $|\alpha_k| < 1/2$ . This system can be solved by the cyclic reduction method [14]. We briefly describe it below for reader's convenience.

A one-step cyclic reduction reduces this system to half size with the same structure. We apply this reduction recursively until a small system is met or until  $A$  becomes almost diagonal. Then this reduced system can be solved directly. To describe this reduction procedure, we assume the current linear system is of the form:

$$(2.9) \quad a_\ell v_{i-1}^\ell + v_i^\ell + a_\ell v_{i+1}^\ell = f_i^\ell.$$

Here,  $\ell$  is the index of the recursion procedure. We eliminate  $v_{2i\pm 1}^\ell$  terms to obtain

$$-a_\ell^2 v_{2i-2}^\ell + (1 - 2a_\ell^2) v_{2i}^\ell - a_\ell^2 v_{2i+2}^\ell = f_{2i}^\ell - a_\ell (f_{2i-1}^\ell + f_{2i+1}^\ell).$$

We rename the variables:

$$\begin{aligned} v_i^{\ell-1} &= v_{2i}^\ell, \\ f_i^{\ell-1} &= \frac{1}{1-2a_\ell^2} (f_{2i}^\ell - a_\ell (f_{2i-1}^\ell + f_{2i+1}^\ell)), \\ (2.10) \quad a_{\ell-1} &= \frac{-a_\ell^2}{1-a_\ell^2}. \end{aligned}$$

Then  $v_i^{\ell-1}$  satisfies

$$a_{\ell-1}v_{i-1}^{\ell-1} + v_i^{\ell-1} + a_{\ell-1}v_{i+1}^{\ell-1} = f_i^{\ell-1}.$$

Thus, we arrive a system of half size with the same structure. Notice that, from (2.10),  $a_{\ell-m}$  converges to zero quadratically as  $m \rightarrow \infty$ . This is because  $|a_\ell| < 1/2$  initially. Usually,  $m = 5$  (i.e. five-level reductions) will make  $a_{\ell-m}$  go down to  $10^{-14}$ . Once  $\{v_i^{\ell-m}\}$  are found, we can reconstruct  $\{v_i^{\ell-m+1}\}$  from

$$\begin{aligned} v_{2i}^{\ell-m+1} &= v_i^{\ell-m}, \\ v_{2i+1}^{\ell-m+1} &= f_{2i+1}^{\ell-m+1} - a_{\ell-m+1} (v_i^{\ell-m} + v_{i+1}^{\ell-m}). \end{aligned}$$

We can continue this reconstruction procedure recursively from  $\ell - m + 1$  to  $\ell$  to obtain  $v_i^\ell$ .

By a direct calculation, the work for performing  $1/P_K(z^2)$  is less than  $(4A + 3M)[K/2]$  per each datum. Here,  $A$  is the addition operation and  $M$  is the multiplication operator. Thus, the procedure to perform the rational filters  $\tilde{h}(z)$  and  $\tilde{g}(z)$  is of linear complexity. In the next section, a comparison study shows that the operation counts for the difference wavelet method are about the same as those of the Daubechies' orthogonal wavelet method and of the CDF method.

**2.3. Stability of the difference wavelet method.** For the difference wavelet method, from (2.1), the corresponding scaling function  $\phi$  is the B-spline:

$$\phi(x) = 1_{[-1,0]}^{*r}(x - [r/2]).$$

The corresponding dual scaling function  $\tilde{\phi}$  has the following properties.

PROPOSITION 1.

1.  $\tilde{\phi}$  is symmetric.
2. The Fourier transform of  $\tilde{\phi}$  satisfies

$$(2.11) \quad |\hat{\tilde{\phi}}(\xi)| = O\left(|\xi|^{-(\tilde{r}-r)/2-1}\right).$$

3.  $\tilde{\phi}$  decays at  $\pm\infty$  exponentially for  $\tilde{r} \geq r$ :

$$(2.12) \quad \int \left| e^{\sigma'|x|} \tilde{\phi}(x) \right|^2 dx < \infty, \quad \forall \sigma' < \sigma,$$

where

$$\sigma = 2 \left| \ln \tan \left( \frac{K+1}{4K} \pi \right) \right|.$$

*Proof.*

1. The symmetry of  $\tilde{\phi}$  follows from  $\tilde{h}(z) = \tilde{h}(-z)$ .
2. Using Fourier transform and (1.11), we have

$$(2.13) \quad \begin{aligned} \hat{\tilde{\phi}}(\xi) &= \prod_{\ell=1}^{\infty} \tilde{h}(e^{-i\xi/2^\ell}) \\ &= e^{-i\xi/2} \left( \frac{\sin \xi/2}{\xi/2} \right)^{\tilde{r}} \prod_{\ell=1}^{\infty} \frac{1}{P_K(e^{-i2\xi/2^\ell})}. \end{aligned}$$

Therefore, to prove (2.11) we should show that the infinite product  $\prod_{\ell=1}^{\infty} 1/P_K(e^{-2i\xi/2^\ell})$  converges uniformly and absolutely on every compact subsets in the complex plane and has the following asymptotic estimate:

$$(2.14) \quad \prod_{\ell=1}^{\infty} \frac{1}{P_K(e^{-i2\xi/2^\ell})} = O(|\xi|^{K-1}) \text{ for } |\xi| \gg 1.$$

To see this, first, from the continuity of  $P_K$  and  $P_K(1) = 1$ , there exist constants  $\xi_0, C > 0$  such that

$$(2.15) \quad \left| \frac{1}{P_K(e^{-2i\xi})} - 1 \right| < C|\xi|$$

for all  $|\xi| \leq \xi_0$ .

Given any compact subset  $B \subset \mathbb{C}$ , there exists an integer  $L_0$  such that  $B \subset \{\xi \in \mathbb{C} \mid |\xi| \leq 2^{L_0}\xi_0\}$ . For any  $\xi \in B$ , there exists a positive integer  $L$  such that  $2^{L-1}\xi_0 < |\xi| < 2^L\xi_0$ . We split the above infinite product  $\prod_{\ell=1}^{\infty}$  into  $\prod_{\ell=1}^L \cdot \prod_{\ell=L+1}^{\infty}$ . From (2.15), the second term has the estimate:

$$\begin{aligned} \prod_{\ell=L+1}^{\infty} \left| \frac{1}{P_K(e^{-i2\xi/2^\ell})} \right| &\leq \prod_{\ell=L+1}^{\infty} \left( 1 + C \frac{|\xi|}{2^\ell} \right) \\ &\leq \exp(C\xi_0) \equiv C_1. \end{aligned}$$

By the dominant convergence theorem, this infinite product converges absolutely and uniformly for  $|\xi| \leq 2^{L_0}\xi_0$ .

Next, from (2.5),

$$(2.16) \quad P_K(e^{-i2\xi}) = \cos^{2K} \xi + \sin^{2K} \xi.$$

It is easy to see that its minimum is at  $\xi = \pi/4$  with minimal value  $2^{-(K-1)}$ . This yields  $\max_{\xi \in \mathbb{R}} |1/P_K(e^{-i2\xi})| = 2^{K-1}$ . Hence,

$$\prod_{\ell=1}^L \left| \frac{1}{P_K(e^{-i2\xi/2^\ell})} \right| \leq 2^{(K-1)L} \leq \left( \frac{2|\xi|}{\xi_0} \right)^{K-1}.$$

Therefore, we obtain  $\prod_{\ell=1}^{\infty} 1/P_K(e^{-i2\xi/2^\ell}) = O(|\xi|^{K-1})$ .

3. First, we show that  $\widehat{\phi}(\xi_1 + i\xi_2)$  is analytic for  $|\xi_2| < \sigma$ . We claim that  $\widehat{\phi}(\xi_1 + i\xi_2)$  is a meromorphic function with poles at:

$$2^\ell (\pm\pi/2 + i \ln |\tan \theta_k|),$$

for  $\ell = 1, 2, \dots, k = 1, \dots, K$  for even  $K$  and for  $k = 1, \dots, K - 1$  for odd  $K$ . To see this, from (2.13), the poles of  $\widehat{\phi}$  are the roots of  $\prod_{\ell=1}^{\infty} P_K(e^{-i2\xi/2^\ell})$ . From (2.6), (2.7), the roots of  $P_K(z^2) = 0$  are  $z = e^{-i\xi} = \pm i \tan \theta_k$  ( $k = 1, \dots, K$  for even  $K$  and  $k = 1, \dots, K - 1$  for odd  $K$ ). That is,  $\xi = \pm\pi/2 + i \ln |\tan \theta_k|$ . The claim follows immediately. With this, the poles of  $\widehat{\phi}(\xi)$  with the smallest imaginary part are  $2(\pm\pi/2 \pm i \ln \tan(\frac{K+1}{4K}))$ . Therefore,  $\widehat{\phi}(\xi_1 + i\xi_2)$  is analytic for  $|\xi_2| < \sigma$ .

Next, we show  $e^{\sigma'|x|}\tilde{\phi} \in L^2(\mathbb{R})$  for any  $\sigma' < \sigma$ . When  $x \geq 0$ , we move the integration line in the Fourier inversion formula from the real line to  $\{\xi_1 + i\sigma' \mid \xi_1 \in \mathbb{R}\}$ :

$$\tilde{\phi}(x) = \frac{1}{\sqrt{2\pi}} \int e^{ix(\xi_1 + i\sigma')} \tilde{\phi}(\xi_1 + i\sigma') d\xi_1.$$

Here, we have used  $\tilde{r} \geq r$ , (2.11) and the fact that  $\tilde{\phi}$  is analytic for  $|\xi| < \sigma$ . From  $\tilde{\phi}(\cdot \pm i\sigma') \in L^2(\mathbb{R})$  for  $\tilde{r} \geq r$  and the Planchel equality, we obtain  $e^{\sigma'|x|}\tilde{\phi} \in L^2(\mathbb{R})$ . When  $x < 0$ , we can move the integration line to  $\{\xi_1 - i\sigma' \mid \xi_1 \in \mathbb{R}\}$  and prove similarly.  $\square$

DEFINITION 2.1. 1. A function  $\phi$  is said to satisfy the Riesz basis property if  $\phi \in L^2(\mathbb{R})$  and there exist constants  $0 < A, B < \infty$  such that for any finite sequence  $\{c_k\}$  we have

$$A \sum_k |c_k|^2 \leq \left\| \sum_k c_k \phi(\cdot - k) \right\|_{L^2}^2 \leq B \sum_k |c_k|^2.$$

2. Two refinable functions  $\phi$  and  $\tilde{\phi}$  are said to be biorthogonal if both of them satisfy the Riesz basis property and they are dual to each other, namely,  $\int \phi(x - i)\tilde{\phi}(x - k) dx = \delta_{i,k}$ .

It is known [20, 9, 6] that a refinable function  $\phi$  with mask  $h(z)$  satisfies the Riesz basis property if and only if  $\phi \in L^2$  and  $h(z)$  satisfies the Cohen criterion [6]. That is, there exist a compact set  $K$  and finite many disjoint closed intervals  $K_i$ , associated with an integer  $n_i$  such that

- (1)  $K$  contains 0 as an interior point,
- (2)  $K = \cup_i K_i$  and  $[-\pi, \pi] = \cup(2n_i\pi + K_i)$ , and  $2n_i\pi + K_i$  can intersect each other at most at their boundaries,
- (3)  $h(e^{i\xi/2^j}) \neq 0$  for all  $j > 0$  and for all  $\xi \in K$ .

PROPOSITION 2. The refinable functions  $\phi$  and  $\tilde{\phi}$  constructed from the difference wavelet method are biorthogonal for  $\tilde{r} \geq r \geq 1$ .

Proof. Since  $\phi$  is a spline for  $r \geq 1$ , it is in  $L^2(\mathbb{R})$ . It also satisfies Cohen's criterion trivially with  $K = [-\pi, \pi]$ . For  $\tilde{\phi}$ , we notice from (2.16) that  $1/P(e^{i2\xi}) > 0$  for  $\xi \in [-\pi, \pi]$ . Hence,  $\tilde{\phi}$  also satisfies Cohen's criterion trivially with  $K = [-\pi, \pi]$ . From (2.11), we see that  $\tilde{\phi} \in L^2(\mathbb{R})$  as  $\tilde{r} \geq r$ . The duality of  $\phi$  and  $\tilde{\phi}$  follows from (1.5) [9].  $\square$

THEOREM 2.1. The difference wavelets  $\psi$  and  $\tilde{\psi}$  with  $\tilde{r} \geq r \geq 1$  are biorthogonal in  $L^2(\mathbb{R})$ .

Proof. Our theorem follows from a theorem of Chui [5] which says that  $\psi$  and  $\tilde{\psi}$  are biorthogonal if and only if  $\phi$  and  $\tilde{\phi}$  are biorthogonal and  $g(z) = \tilde{h}(-z)P(z^2)$  and  $\tilde{g}(z) = h(-z)/P(z^2)$  with  $P$  being in the Wiener class. We choose  $P = P_K$  here.  $P_K(z)$  is a Laurent polynomial and, from (2.16),  $P_K(z) \neq 0$  for all  $|z| = 1$ . Hence  $P_K$  is in the Wiener class.  $\square$

**3. A Comparison Study.** In this section, we compare four wavelet methods: Cohen-Daubechies-Feauveau wavelet (CDF), Daubechies orthogonal wavelet (Daub), Chui-Wang's semi-orthogonal wavelet (CW), and the difference wavelet (Diff). Natural criteria of a good wavelet method are fast, stable and efficient. Therefore our comparisons include

- (1) operation counts,
- (2) condition numbers,
- (3) compression ratio for data sequence and matrices.

Roughly speaking, the results below show that the difference wavelet method is more efficient than other methods. In addition, we also compare

- (4) regularity,
- (5) approximation power,
- (6) approximation error for functions with jumps,
- (7) length of “essential support,”

to demonstrate that this better efficiency is probably due to its better regularity with slightly bigger “essential support.” It is surprising that even for non-smooth functions, Diff has better representation in the sense that it is smoother and less overshoots and undershoots.

**3.1. Operation Counts.** In this comparison, we shall show that the operation cost of Diff is about the same as those of CDF or Daub, even though it is a rational filter. We compute the number of operations per each datum in a one-level wavelet transform. The operation counts include both forward and inverse transforms (i.e. (1.1) and (1.2)). The common part of the four methods is  $(\frac{1+z}{2})^{2K}$ . A factorization of Laurant polynomials does not change its operation counts. Thus, the differences among these four methods are the operations for  $Q_K(z)$ ,  $G_K(z)/G_K(z^2)$  and  $1/P_K(z)$ . The inversion of  $1/P_K(z^2)$  and  $1/G_K(z^2)$  are performed by the cyclic reduction method. Table 1 shows that Diff has about the same number of operations as those of CDF or Daub, while CW has more.

Diff	CDF, Daub	CW
$1/P_K(z)$	$Q_K(z)$	$G_K(z)/G_K(z^2)$
$< (4A + 3M)[K/2]$	$(2A + 1M)K$	$(6A + 5M)K$

TABLE 1

*Comparison of operation counts for performing filters  $1/P_K(z)$  (Diff),  $Q_K(z)$  (CDF and Daub) and  $G_K(z)/G_K(z^2)$  (CW) per each datum. Here,  $A$  is the addition operation and  $M$  is the multiplication.*

**3.2. Condition numbers.** We compute the condition number of  $T_J : c_J \mapsto (c_0, d_0, \dots, d_{J-1})$  by Matlab to study the sensitivity of various wavelet transforms. The matrix size is  $1024 \times 1024$ . Table 2 shows the following:

1. It is clear that orthogonal wavelet transform has the smallest condition number, which is 1.
2. Among the biorthogonal wavelet methods considered here, the difference wavelet method has the smallest condition numbers for the cases  $r = \tilde{r}$ . In general, the condition numbers of difference wavelet transforms are reasonably small for application.
3. Notice that the condition numbers of CDF are big for the cases  $r = \tilde{r} \geq 4$ . This is because the corresponding  $\tilde{\phi}$ 's are not in  $L^2(\mathbb{R})$ , and the corresponding wavelets do not form a Riesz basis in  $L^2(\mathbb{R})$ .

**3.3. Compression ratio for data sequences.** We demonstrate by numerical tests to show that Diff method does have better compression ratio for data sequences (both smooth and nonsmooth) and matrices (smooth kernel and singular kernel).

$r$	$\tilde{r}$	Diff	CDF	CW	Daub
1	5	5.2	2.1	—	—
2	4	3.5	2.5	—	—
3	3	3.1	9.1	5.1	1.0
1	7	10.1	2.4	—	—
2	6	7.0	2.5	—	—
3	5	5.5	5.9	—	—
4	4	4.5	35.4	10.0	1.0
1	9	19.5	2.5	—	—
2	8	14.0	2.5	—	—
3	7	11.0	5.5	—	—
4	6	8.6	14.7	—	—
5	5	7.0	154.9	19.3	1.0

TABLE 2

Comparison of the condition numbers of various wavelet transforms  $T_J : c_J \rightarrow (c_4, d_4, \dots, d_{J-1})$ . Here,  $r$  is the average order,  $\tilde{r}$  is the difference order and  $J = 10$ . The matrix size is  $1024 \times 1024$ . The results are computed by Matlab.

We measure the efficiency of a wavelet representation for a data sequence  $c_J$  by a quantity  $C_2$  defined by

$$(3.1) \quad C_2(\epsilon, J, r, \tilde{r}) = N_2(\epsilon, J, r, \tilde{r})/N,$$

where  $N = 2^J$  is the total number of data. The number  $N_2$  is defined as follows. Firstly, we transform the discrete data  $c_J$  to  $(c_0, d_0, \dots, d_{J-1})$ . Next, we truncate  $(c_0, d_0, \dots, d_{J-1})$  by a threshold  $\delta$  to yield  $(\bar{c}_0, \bar{d}_0, \dots, \bar{d}_{J-1})$ . The threshold  $\delta$  is chosen so that the inverse transform of  $(\bar{c}_0, \bar{d}_0, \dots, \bar{d}_{J-1})$  (denoted by  $\bar{c}_J$ ) is within  $\epsilon$  neighborhood of  $c_J$  in  $\ell^2$ , i.e.

$$\left( \sum_i (c_{J,i} - \bar{c}_{J,i})^2 2^{-J} \right)^{1/2} \leq \epsilon.$$

Then we define  $N_2(\epsilon, J, r, \tilde{r})$  to be the number of nonzero elements in  $(\bar{c}_0, \bar{d}_0, \dots, \bar{d}_{J-1})$ .

We perform two tests: one is a smooth data, the other is a nonsmooth data. Namely, we choose  $c_{J,i} = u(2^{-J}i)$ ,  $i = 1, \dots, 2^J$ , where  $u(x) = \sin 4\pi x + \sin 6\pi x$  for the first test, and  $u(x) = \chi_{[0,1/2]}(x)$  for the second test.

Table 3 and 4 show the value of  $C_2$  corresponding to the above two tests, where  $N = 2^{10}$  and  $\delta = 10^{-6}$ . We should compare various methods with fixed  $K$ , because they have about the same amount of operation cost. We observe that the Diff with  $r = 1$  is the best in both tests.

**3.4. Compression ratio for matrices.** Fast matrix-vector multiplication is important in many applications. Following Beylkin-Coifman-Rokhlin [2], we use “standard method” for matrix compression. That is, given a  $2^J \times 2^J$  matrix  $G$ , we transform it to  $T_J G T_J^t$ . A truncation is applied to every entries of  $T_J G T_J^t$  with threshold  $\delta$ . The number  $N_2$  is the total number of nonzero entries. The error of  $\bar{G}$  (i.e. the inverse transform of the truncated  $T_J G T_J^t$ ) is measured by a matrix  $\ell^2$  norm. The quantity  $C_2$  is defined by  $N_2/N$ . Roughly speaking,  $C_2$  is the number of operation needed per each datum for a matrix-vector multiplication. Notice that

$r$	$\tilde{r}$	Diff	CDF	CW	Daub
1	5	0.105	0.119	–	–
2	4	0.203	0.330	–	–
3	3	0.423	0.916	0.423	0.746
1	7	0.057	0.063	–	–
2	6	0.061	0.111	–	–
3	5	0.113	0.213	–	–
4	4	0.223	0.455	0.111	0.234
1	9	0.031	0.053	–	–
2	8	0.031	0.061	–	–
3	7	0.057	0.100	–	–
4	6	0.061	0.143	–	–
5	5	0.117	0.262	0.057	0.119

TABLE 3

Comparison of  $C_2$ , where  $C_2 = N_2/N$ ,  $N_2$  is the number of nonzero truncated wavelet coefficients,  $N$  is the total number of data. The test data is  $c_{J,k} = u(2^{-J}k)$ , where  $u(x) = \sin 4\pi x + \sin 6\pi x$ . Here,  $N = 2^{10}$ , the threshold  $\delta = 10^{-6}$ ,  $r$ , the averaging order and  $\tilde{r}$ , the differencing order.

$r$	$\tilde{r}$	Diff	CDF	CW	Daub
1	5	0.031	0.059		
2	4	0.031	0.061		
3	3	0.033	0.061	0.266	0.059
1	7	0.059	0.082		
2	6	0.059	0.084		
3	5	0.057	0.084		
4	4	0.057	0.084	0.334	0.082
1	9	0.059	0.102		
2	8	0.059	0.104		
3	7	0.057	0.107		
4	6	0.061	0.107		
5	5	0.059	0.105	0.387	0.094

TABLE 4

Comparison of the  $C_2$  for non-smooth data:  $c_{J,k} = u(2^{-J}k)$  and  $u(x) = \chi_{[0,1/2]}$ ,  $N = 2^{10}$ ,  $\delta = 10^{-6}$ .

the compression ratio mentioned in other literatures is  $N^2/N_2$ , which is  $N/C_2$  in our language.

We perform three tests. The first one is the heat kernel on periodic domain. The second one is a singular kernel which is basically the Green's function of the Laplacian in 2-d. The third one is the matrix which converts the coefficients of finite Chebyshev expansion into the coefficients of a finite Legendre expansion.

**3.4.1. Heat kernel.** We evaluate the integral:

$$u(x, t) = \int_{-1}^1 G(x, y, t) f(y) dy$$

where

$$G(x, y, t) = \sum_{\ell=-\infty}^{\infty} \frac{1}{\sqrt{4\pi t}} \exp\left(-\frac{[2\pi(x-y-\ell)]^2}{4t}\right),$$

that is, the fundamental solution of the heat equation on the periodic domain  $[0, 1]$ . This integral appears commonly for solving convection-diffusion equations [8]. In our test, we choose  $t = 0.1$ , and the kernel is obtained by summing the heat kernel over 20 periods. We discretize the above integral by trapezoidal rule on a uniform grid. Table 5 shows the quantity  $C_2$  and Table 6 gives the corresponding  $L^2$ -errors and the thresholds. In the comparison, we fix the threshold  $\delta$  used for the Diff. The thresholds used for other methods are chosen so that the corresponding errors are no less than the errors of the Diff produced by using the fixed threshold  $\delta$ . This is to guarantee that the Diff produces the least error in our test.

The result shows that the difference wavelet method with  $r = 1$  is the most efficient method.

$r$	$\tilde{r}$	Diff	CDF	CW	Daub
1	5	1.86	4.59		
2	4	4.41	12.22		
3	3	14.23	50.88	15.19	30.31
1	7	0.72	1.95		
2	6	1.50	2.00		
3	5	1.86	6.38		
4	4	4.31	19.22	2.48	4.98
1	9	0.70	1.09		
2	8	0.75	2.13		
3	7	0.72	2.16		
4	6	1.59	5.38		
5	5	1.86	14.59	2.5	2.09

TABLE 5

This table shows the values of  $C_2$ , the number of operations per each datum in matrix-vector multiplication. The matrix  $G$  is a finite approximation of the fundamental solution of heat equation over period domain  $(0, 1)$ . The matrix size is  $1024 \times 1024$ . The corresponding errors and thresholds are tabulated in Table 6. The result demonstrates that the Diff method with  $r = 1$  is the most efficient method.

**3.4.2. Singular kernel.** Singular kernels often appear in boundary integral methods. Here, we consider the kernel corresponding to the Laplacian in 2-dimension [2]. Namely,

$$(3.2) \quad \Lambda_{ij} = \begin{cases} \ln(i-j)^2 & \text{if } i \neq j, \\ 0 & \text{if } i = j. \end{cases}$$

Table 7 shows the value of  $C_2$  and Table 8 gives the corresponding errors  $\epsilon$  and thresholds  $\delta$ . The result demonstrates that the difference wavelet method is the most efficient method.

**3.4.3. Fast Legendre transform.** It is well-known that a fast Legendre transform can be achieved through a fast Chebyshev transform followed by a compressed

$r$	$\tilde{r}$	Diff	CDF	CW	Daub
1	5	1.4e-07	4.2e-07		
2	4	5.2e-07	2.0e-06		
3	3	2.3e-06	4.0e-06	2.9e-06	2.6e-06
1	7	1.1e-08	1.2e-07		
2	6	6.4e-07	2.9e-06		
3	5	3.3e-07	5.3e-07		
4	4	1.2e-06	2.0e-06	1.5e-06	2.0e-06
1	9	1.1e-08	1.9e-07		
2	8	9.1e-10	6.7e-08		
3	7	2.6e-08	8.9e-07		
4	6	1.5e-06	2.0e-06		
5	5	1.1e-06	5.3e-06	1.5e-06	1.4e-06
1	5	1e-06	2e-06		
2	4	1e-06	1e-06		
3	3	1e-06	1e-06	5e-06	1e-06
1	7	1e-06	1e-06		
2	6	1e-06	1.7e-05		
3	5	1e-06	1e-06		
4	4	1e-06	1e-06	3.9e-05	1e-05
1	9	1e-06	1e-06		
2	8	1e-06	1e-06		
3	7	1e-06	2e-06		
4	6	1e-06	1e-06		
5	5	1e-06	1e-06	3e-05	1e-05

TABLE 6

The top subtable shows the  $L^2$ -error  $\epsilon$  of the truncated matrix  $\tilde{G}$  in Table 5. The bottom one gives the corresponding threshold  $\delta$  used for truncation.

matrix which converts the coefficients of a finite Chebyshev expansion into the coefficients of a finite Legendre expansion of the same polynomial [1, 2]. The matrix is given by

$$\Lambda_{ij} = \begin{cases} \frac{1}{2}\Lambda^2(j) & \text{if } 0 = i \leq j < N, \\ \frac{2}{\pi}\Lambda(j-i)\Lambda(j+i) & \text{if } 0 < i \leq j < N, \\ 0 & \text{otherwise,} \end{cases}$$

where,  $\Lambda(z) = \Gamma(z + 1/2)/\Gamma(z + 1)$  and  $\Gamma(z)$  is the gamma function. Table 9 is the value of  $C_2$  and Table 10 is the corresponding errors and thresholds. The result shows that the difference wavelet method is again the most efficient method.

**3.5. Regularity.** In this subsection, we compare the smoothness of  $\tilde{\phi}$  for various methods. We shall find  $\tilde{s}$  such that  $\tilde{\phi} \in H^s(\mathbb{R})$  for all  $s < \tilde{s}$ . Here,  $H^s(\mathbb{R})$  denotes the Sobolev space of regularity order  $s$ .

Firstly, for semi-orthogonal wavelet,  $\tilde{r} = K$  and  $\tilde{\phi}$  is the spline  $1_{[0,1]}^{*K}$ . Hence,  $\tilde{s} = \tilde{r} - 1/2$ . For other methods, basically we compute the spectral radius of the transition matrix to determine its regularity. First, we use the following lemma [9, 20] to normalize our comparison.

$r$	$\tilde{r}$	Diff	CDF	CW	Daub
1	5	48.7	64.6		
2	4	55.0	85.1		
3	3	75.9	149.4	141.7	86.5
1	7	50.3	68.2		
2	6	49.0	71.1		
3	5	48.5	81.4		
4	4	55.8	128.1	136.8	75.5
1	9	57.0	63.5		
2	8	54.5	70.6		
3	7	47.6	75.4		
4	6	46.9	89.4		
5	5	49.0	121.4	134.2	68.9

TABLE 7

Compression ratio  $C_2$  for the matrix (defined in equation 3.2), the matrix size is  $N = 1024$ . The result shows that the Diff is the most efficient one.

LEMMA 3.1. Suppose  $\tilde{h}(z) = \left(\frac{1+z}{2}\right)^{\tilde{r}} \tilde{H}(z)$  with  $\tilde{H}(-1) \neq 0$  and  $\tilde{H}(1) = 1$ . Let  $\tilde{\phi}$  and  $\tilde{\Phi}$  be the scaling functions associated with  $\tilde{h}(z)$  and  $\tilde{H}(z)$ . Then  $\tilde{\phi} \in H^{s+\tilde{r}}(\mathbb{R})$  if and only if  $\tilde{\Phi} \in H^s(\mathbb{R})$ .

The proof of this lemma follows easily from

$$\widehat{\tilde{\phi}}(\xi) = e^{-i\tilde{r}\xi/2} \left(\frac{\sin \frac{\xi}{2}}{\frac{\xi}{2}}\right)^{\tilde{r}} \widehat{\tilde{\Phi}}(\xi),$$

see [20]. Now, we apply this lemma to our cases:

$$\begin{aligned} \tilde{h}_{\text{CDF}}(z) &= \left(\frac{1+z}{2}\right)^{\tilde{r}} \tilde{H}_{\text{CDF}}(z), \\ \tilde{h}_{\text{Diff}}(z) &= \left(\frac{1+z}{2}\right)^{\tilde{r}} \tilde{H}_{\text{Diff}}(z). \end{aligned}$$

Thus, we only need to compare the regularity of the scaling functions  $\tilde{\Phi}_{\text{CDF}}$  and  $\tilde{\Phi}_{\text{Diff}}$  associated with  $\tilde{H}_{\text{CDF}}$  and  $\tilde{H}_{\text{Diff}}$ , respectively. Table 11 gives the regularity for these functions. It is basically quoted from Cohen-Daubechies [7]. They gave a sharp estimates on  $s_0$  by using transition matrix method, where  $s_0$  is the best Sobolev exponent of the Sobolev space in which these functions can live. The case of Diff in the first row is the Butterworth case in their paper. See also Fan and Sun [13].

We can transfer this table to a regularity table for  $\tilde{\phi}$ , see Table 12. We observe that

1. Diff is more regular than CDF and Daub.
2. With  $\tilde{r}$  fixed,  $\phi_{\text{CW}}$  is the B-spline, which is the most regular function with given  $\tilde{r}$ . However, with  $K$  fixed, which means it takes about same amount of time to take wavelet transformation, Diff with  $r = 1$  is even more regular than CW.

$r$	$\tilde{r}$	Diff	CDF	CW	Daub
1	5	1.1e-06	1.3e-06		
2	4	1.0e-06	1.2e-06		
3	3	2.1e-06	2.5e-06	2.2e-06	2.3e-06
1	7	3.4e-06	3.6e-06		
2	6	2.9e-06	2.9e-06		
3	5	3.7e-06	3.9e-06		
4	4	3.6e-06	5.8e-06	4.2e-06	3.8e-06
1	9	1.5e-05	1.9e-05		
2	8	5.1e-06	5.2e-06		
3	7	5.8e-06	6.0e-06		
4	6	5.1e-06	5.7e-06		
5	5	9.3e-06	2.6e-05	1.6e-05	9.7e-06
1	5	1e-06	5e-06		
2	4	1e-06	2e-06		
3	3	1e-06	1e-06	6.4e-05	7e-06
1	7	1e-06	1.9e-05		
2	6	1e-06	7e-06		
3	5	1e-06	4e-06		
4	4	1e-06	1e-06	0.000256	1.9e-05
1	9	1e-06	7.5e-05		
2	8	1e-06	2.1e-05		
3	7	1e-06	6e-06		
4	6	1e-06	3e-06		
5	5	1e-06	1e-06	0.001024	4.5e-05

TABLE 8

The matrix error  $\epsilon$  and the thresholds for the matrix (defined in equation 3.2) in Table 7

**3.6. Approximation power.** Given any function  $u \in L^2(\mathbb{R})$ , it can be approximated by the projection

$$\mathcal{P}^J u := \sum_k (u, \phi_{J,k}) \tilde{\phi}_{J,k}$$

as  $J \rightarrow \infty$  [9]. When  $u \in L^2(\mathbb{R}) \cap C^{\tilde{r}+1}(\mathbb{R})$ , this approximation has the following sharp estimate [17, 18, 19]:

$$\|\mathcal{P}^J u - u\|_{L^2} = C_{\tilde{\phi}} 2^{-\tilde{r}J} \|u^{(\tilde{r})}\|_{L^2} + O(2^{-(\tilde{r}+1)J}) \text{ as } J \rightarrow \infty,$$

where

$$C_{\tilde{\phi}} = \frac{1}{\tilde{r}!} \left( \sum_{k \neq 0} |\hat{\tilde{\phi}}^{(\tilde{r})}(2k\pi)|^2 \right)^{1/2}.$$

$r$	$\tilde{r}$	Diff	CDF	CW	Daub
1	5	29.3	39.6		
2	4	28.1	42.3		
3	3	29.1	55.3	69.4	40.1
1	7	39.1	37.3		
2	6	33.3	32.1		
3	5	29.6	46.6		
4	4	28.2	62.1	73.3	41.4
1	9	44.0	31.4		
2	8	35.7	28.3		
3	7	31.3	43.8		
4	6	29.4	49.2		
5	5	28.5	67.4	82.3	39.5

TABLE 9

Compression ratio  $C_2$  for matrix which converts the finite Chebyshev expansion to finite Legendre expansion. The matrix size is  $1024 \times 1024$ . The result shows that the different wavelet method has the best compression ratio (see the cases  $(r, \tilde{r}) = (2, 4), (4, 4)$  and  $(5, 5)$ ).

The smaller  $C_{\tilde{\phi}}$  is, the better the approximation power is. Following Unser [19], the constant  $C_{\tilde{\phi}}$  has the following expression:

$$C_{\tilde{\phi}} = \begin{cases} \frac{Q_K(-1)}{\sqrt{(4^{\tilde{r}}-1)2^{\tilde{r}}}} D & \text{for CDF} \\ \frac{1/P_K(1)}{\sqrt{(4^{\tilde{r}}-1)2^{\tilde{r}}}} D & \text{for Diff} \\ \frac{1}{\sqrt{(4^K-1)2^K}} D & \text{for CW} \\ \frac{Q(-1)}{\sqrt{(4^K-1)2^K}} D & \text{for Daub} \end{cases}$$

$$D = \left( \sum_{k \in \mathbb{Z}} |\hat{\phi}((2k+1)\pi)|^2 \right)^{1/2}.$$

In Table 13, we compute  $C_{\tilde{\phi}}$  for various methods. We observe the following things.

1. When  $r = \tilde{r} = K$  fixed, CW has the best approximation power and Diff is the second. This is due to that the approximate space  $\tilde{V}^J$  of CW is spanned by splines which are the most regular scaling function with given  $K$ .
2. However, if we fix  $K$  (this means the amount of works for wavelet transform is of the same order), then we see Diff has the best approximation power. This table is consistent to the previous regularity table.

**3.7. Approximation error for non-smooth functions.** When  $u$  is not smooth, the representation  $\mathcal{P}^J u = \sum (u, \phi_{J,k}) \tilde{\phi}_{J,k}$  exhibits the Gibbs phenomenon. We choose the test function to be  $\chi_{[1/4, 3/4]}$  on the periodic domain  $[0, 1]$ . We compare the  $L^1$  and  $L^2$  errors and the heights of overshoots and undershoots. In this comparison, the coefficients  $(u, \phi_{J,k})$  are computed as follows. For Diff and CDF,  $\phi$  is the splines, we compute  $(u, \phi)$  exactly by using Maple. For CW,  $\mathcal{P}^J u$  is the orthogonal projection of  $u$  onto the space spanned by splines, therefore we find the coefficients

$r$	$\tilde{r}$	Diff	CDF	CW	Daub
1	5	8.8e-07	1.0e-06		
2	4	1.0e-06	1.6e-06		
3	3	1.6e-06	2.2e-06	1.8e-06	1.9e-06
1	7	5.4e-06	5.7e-06		
2	6	5.5e-06	5.9e-06		
3	5	3.0e-06	4.0e-06		
4	4	3.5e-06	7.8e-06	4.9e-06	3.7e-06
1	9	1.8e-05	1.8e-05		
2	8	1.5e-05	1.5e-05		
3	7	7.6e-06	8.0e-06		
4	6	6.5e-06	6.6e-06		
5	5	7.7e-06	6.0e-05	1.41e-05	7.8e-06
1	5	1e-06	4e-06		
2	4	1e-06	3e-06		
3	3	1e-06	1e-06	3.2e-05	6e-06
1	7	1e-06	1.6e-05		
2	6	1e-06	1.3e-05		
3	5	1e-06	3e-06		
4	4	1e-06	1e-06	0.000128	1.2e-05
1	9	1e-06	6.1e-05		
2	8	1e-06	3.9e-05		
3	7	1e-06	8e-06		
4	6	1e-06	3e-06		
5	5	1e-06	1e-06	0.000512	2.5e-05

TABLE 10

The top subtable shows the errors  $\epsilon$  and the bottom shows the thresholds  $\delta$  corresponding to Table 9.

K	2	3	4	5	6	7
$s_0(\text{Diff})$	-1.5	-1.9	-2.3	-2.7	-3.1	-3.5
$s_0(\text{CDF})$	-2.0	-3.2	-4.4	-5.8	-7.2	-8.6
$s_0(\text{Daub})$	-1.0	-1.6	-2.2	-2.9	-3.6	-4.3

TABLE 11

Regularity table.  $K = (r + \tilde{r})/2$ ,  $\prod Q_K(e^{-i\xi/2^\ell}) \in H^s(\mathbb{R})$  for all  $s < s_0(\text{CDF})$ , and  $\prod 1/P_K(e^{-2i\xi/2^\ell}) \in H^s(\mathbb{R})$  for all  $s < s_0(\text{Diff})$ , quoted from [7]. Those with less negative values are of better regularity.

$c_{J,k} = (u, \phi_{J,k})$  by solving the linear system

$$\sum_k (\tilde{\phi}_{J,i} \tilde{\phi}_{J,k}) c_{J,k} = (u, \tilde{\phi}_{J,i}),$$

where  $\tilde{\phi}_{J,k}$  is the splines. For Daub, we do not have an analytic formula for  $\phi_{J,k}$ . Therefore we compute  $(u, \phi_{J,k})$  by direct numerical integration. We first transform the integral  $(u, \phi_{J,k})$  to  $2^{-J/2} \int u(2^{-J}(k+y))\phi(y) dy$ . Since  $u$  is a characteristic function, the integral only involves integration of  $\phi$ . A trapezoidal rule is then applied for this

$r$	$\tilde{r}$	Diff	CDF	CW	Daub
1	5	3.1	1.8	—	—
2	4	2.1	0.8	—	—
3	3	1.1	-0.2	2.5	1.4
1	7	4.7	2.6	—	—
2	6	3.7	1.6	—	—
3	5	2.7	0.6	—	—
4	4	1.7	-0.4	3.5	1.8
1	9	6.3	3.2	—	—
2	8	5.3	2.2	—	—
3	7	4.3	1.2	—	—
4	6	3.3	0.2	—	—
5	5	2.3	-0.8	4.5	2.1

TABLE 12

Comparison of regularity of  $\tilde{\phi}$ . The table shows the value of  $\tilde{s}$ , where  $\tilde{\phi} \in H^s(\mathbb{R})$  for all  $s < \tilde{s}$ .

$r$	$\tilde{r}$	Diff	CDF	CW	Daub
1	5	1.1e-03	1.1e-02	—	—
2	4	6.8e-03	6.8e-02	—	—
3	3	4.3e-02	6.2e-01	5.8e-03	3.0e-01
1	7	6.8e-05	2.4e-03	—	—
2	6	4.3e-04	1.5e-02	—	—
3	5	2.7e-03	9.8e-02	—	—
4	4	1.7e-02	1.8e+00	9.1e-04	5.6e-01
1	9	4.2e-06	5.3e-04	—	—
2	8	2.7e-05	3.3e-03	—	—
3	7	1.7e-04	2.1e-02	—	—
4	6	1.1e-03	1.6e-01	—	—
5	5	6.6e-03	1.1e+01	1.4e-04	1.3e+00

TABLE 13

Comparison of the constant  $C_{\tilde{\phi}}$  in the estimate of the approximation power. The result shows that with  $K$  fixed, the difference wavelet method has smallest constant  $C_{\tilde{\phi}}$ , i.e. the best approximation power.

integration using  $N = 2^{13}$  grid points. Thus, the numerical error for this integration is of order  $O(2^{-J/2-13})$ , which is relatively small compared with the approximation errors.

Table 14–15 show the  $L^1$  and  $L^2$  norms of the error  $\mathcal{P}^J u - u$ . Here,  $u = \chi_{[1/4, 3/4]}$ ,  $J = 10$  and the numerical errors are computed using  $2^{13}$  grid points. Table 16 shows the heights of the corresponding overshoots or undershoots (i.e.  $\max_{x \in (1/4, 3/4)} \mathcal{P}^J u(x) - 1$ , or  $-\min_{x \notin [1/4, 3/4]} \mathcal{P}^J u$ ). We do not show the results of CDF for the case (3, 3), (4, 4) and (5, 5), because the corresponding errors are too large. In the comparison, for each  $K = 3, 4, 5$ , we choose the best case for each method. For instance, for Diff, we choose  $(r, \tilde{r}) = (3, 3), (2, 6), (2, 8)$ . For CDF, we select  $(r, \tilde{r}) = (1, 5), (2, 6), (2, 8)$ . For CW and Daub, we have no other choices but  $(r, \tilde{r}) = (3, 3), (4, 4), (5, 5)$ . It is surprising that the two non-compact supported

wavelet methods, Diff and CW, have the least errors of overshoots and undershoots. Indeed, Diff is slightly better than CW. It only has 9% overshoots or undershoots, which is about the same as those using finite Fourier method. Figure 1 are the corresponding graphs of  $\mathcal{P}^J u$  in the interval  $[0.24, 0.26]$ . We observe that Diff does have better representation in the sense of smoothness and smaller overshoots/undershoots.

$r$	$\tilde{r}$	Diff	CDF	CW	Daub
1	5	0.0045	0.0034	–	–
2	4	0.0046	0.0050	–	–
3	3	0.0027	–	0.0034	0.0046
1	7	0.0053	0.0038	–	–
2	6	0.0045	0.0047	–	–
3	5	0.0036	0.0031	–	–
4	4	0.0052	–	0.0050	0.0034
1	9	0.0057	0.0041	–	–
2	8	0.0045	0.0046	–	–
3	7	0.0042	0.0031	–	–
4	6	0.0052	0.0153	–	–
5	5	0.0044	–	0.0041	0.0033

TABLE 14

The  $L^1$  error of  $u - \mathcal{P}^J u$ , where  $u = \chi_{[1/4, 3/4]}$  on the periodic domain  $[0, 1]$  and  $J = 10$ .

$r$	$\tilde{r}$	Diff	CDF	CW	Daub
1	5	0.00074	0.00072	–	–
2	4	0.00077	0.00083	–	–
3	3	0.00067	–	0.00068	0.00077
1	7	0.00077	0.00073	–	–
2	6	0.00074	0.00080	–	–
3	5	0.00068	0.00072	–	–
4	4	0.00078	–	0.00075	0.00060
1	9	0.00078	0.00074	–	–
2	8	0.00073	0.00078	–	–
3	7	0.00070	0.00068	–	–
4	6	0.00073	0.00190	–	–
5	5	0.00071	–	0.00069	0.00069

TABLE 15

The  $L^2$  error of  $u - \mathcal{P}^J u$ , where  $u = \chi_{[1/4, 3/4]}$  on the periodic domain  $[0, 1]$  and  $J = 10$ .

**3.8. Support.** We have shown that  $\tilde{\phi}_{Diff}$  decays exponentially in the previous section. Below, the length of the region where  $|\tilde{\phi}(x)| \geq 10^{-3}$ . Table 17 shows a comparison of this “essential support.” We observe that the essential support of  $\tilde{\phi}_{Diff}$  is about twice of the support of  $\tilde{\phi}_{CDF}$ . Figures 2, 3 give the profiles of  $\tilde{\phi}$ ,  $\psi$  and  $\tilde{\psi}$  of the difference wavelet method for  $K = 3$  and 4, respectively, with various  $r$ . One observes that their “essential supports” are indeed quite compact.

$r$	$\tilde{r}$	Diff	CDF	CW	Daub
1	5	0.19	0.17	—	—
2	4	0.10	0.19	—	—
3	3	0.09	—	0.10	0.13
1	7	0.20	0.18	—	—
2	6	0.09	0.11	—	—
3	5	0.10	0.13	—	—
4	4	0.14	—	0.10	0.16
1	9	0.20	0.19	—	—
2	8	0.09	0.10	—	—
3	7	0.12	0.11	—	—
4	6	0.13	1.00	—	—
5	5	0.12	—	0.09	0.11

TABLE 16

This table shows the magnitudes of overshoot or undershoot of  $\mathcal{P}^J u$ , where  $u = \chi_{1/4, 3/4}$  and  $J = 10$ .

$r$	$\tilde{r}$	Diff	CDF	Daub
1	5	12.8	6.2	—
2	4	13.0	6.6	—
3	3	13.0	6.7	4.3
1	7	15.4	7.9	—
2	6	17.0	8.2	—
3	5	18.6	8.5	—
4	4	18.8	9.6	5.2
1	9	19.3	8.5	—
2	8	21.1	8.7	—
3	7	22.8	9.5	—
4	6	23.0	11.0	—
5	5	24.8	12.7	6.2

TABLE 17

Comparison of the length of the “essential support” (where  $|\tilde{\phi}(x)| \geq 10^{-3}$ )

**Acknowledgement.** The authors thank Professor Gilbert Strang for his helpful comments.

## REFERENCES

- [1] B.K. ALPERT AND V. ROKHLIN, *A fast algorithm for the evaluation of Legendre expansions*, SIAM J. Sci. Stat. Comput., 12:1 (1991), pp. 158–179.
- [2] G. BEYLKIN, R. COIFMAN, AND V. ROKHLIN, *Fast wavelet transforms and numerical algorithms I*, Communications on Pure and Applied Mathematics, Vol. XLIV (1991), pp. 141–183.
- [3] C. K. CHUI, *An Introduction of Wavelets*, Academic Press, San Diego, CA (1992).
- [4] C.K CHUI AND J. Z. WANG, *A cardinal spline approach to wavelets*, Proc. Amer. Math. Soc., 113 (1991), pp. 785–793.
- [5] CHARLES K. CHUI, *Wavelets: A Mathematical Tool for Signal Analysis*, SIAM Monographs on Mathematical Modeling and Computation, 1997.
- [6] A. COHEN, *Ondelettes, analyses multirésolutions et traitement numérique du signals*, thesis, Université Paris IX Dauphine, Dauphine, France, (1990).
- [7] A. COHEN AND I. DAUBECHIES, *A new technique to estimate the regularity of refinable functions*, Revista Matemática Iberoamericana, 12:2 (1996), pp. 527–593.
- [8] B. ENGQUIST, S. OSHER AND S. ZHONG, *Fast wavelet based algorithms for linear evolution equations*, SIAM J. Sci. Comput. Vol., 15:4 (1994), pp. 755–775.
- [9] A. COHEN, I. DAUBECHIES AND J. FEAUVEAU, *Biorthogonal bases of compactly supported wavelets*, Comm. Pure Appl. Math., 45 (1992), pp. 485–560.
- [10] W. DAHMEN, *Stability of multiscale transformations*, J. Fourier Analysis and Applications.
- [11] I. DAUBECHIES, *Orthonormal bases of compactly supported wavelets*, Comm. Pure Appl. Math., 41 (1988), pp. 909–996.
- [12] I. DAUBECHIES, *Ten Lectures on Wavelets*, CBMS-NSF Regional Conf. Series in Appl. Math. Vol. 61, SIAM, Philadelphia, PA (1992).
- [13] A. H. FAN AND QIYU SUN, *Regularity of Butterworth refinable functions*, preprint, Laboratoire d'Analyse Numérique, Université de Pierre et Marie Curie, 1997.
- [14] G. GOLUB AND C. F. VAN LOAN, *Matrix Computations*, Johns Hopkins University Press, (1984)
- [15] A. OPPENHEIM AND R. SCHAEFER, *Digital Signal Processing*, Prentice Hall, New York, 1975.
- [16] G. STRANG AND T. NGUYEN, *Wavelets and Filter Banks*, Wellesley-Cambridge Press (1996).
- [17] G. STRANG AND G. FIX, *A Fourier analysis of the finite element variational method*, in Constructive Aspects of Functional Analysis, (1971), pp. 796–830.
- [18] W. SWELDENS AND R. PIESENS, *Quadrature formulae and asymptotic error expansions for wavelet approximations of smooth functions, I, II*, SIAM J. Numer. Anal., 31 (1994), pp. 1240–1264; Numer. Math., 68 (1994), pp. 377–401.
- [19] M. UNSER, *Approximation power of biorthogonal wavelet expansions*, IEEE Trans. SP., 44:3 (1996), pp. 519–527.
- [20] L. VILLEMOES, *Energy moments in time and frequency of two-scale difference equation solutions and wavelets*, SIAM J. Math. Anal., 23 (1992), pp. 1519–1543.

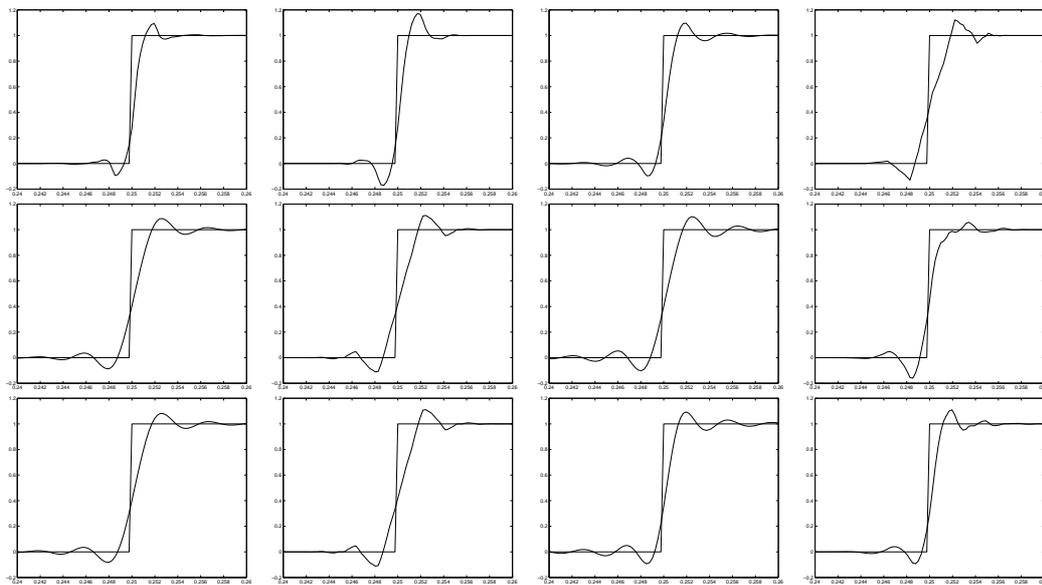


FIG. 1. The graphs of  $\mathcal{P}^J u$  on the interval  $[0.24, 0.25]$ , where  $u = \chi_{[1/4, 3/4]}$  and  $J = 10$ . The graphs from left to right are the representations by Diff, CDF, CW and Daub, respectively. For each method, the graphs from top to bottom corresponding to the best case for each  $K = 3, 4, 5$ . More precisely, from top to bottom,  $(r, \bar{r}) = (3, 3), (2, 6), (2, 8)$  for the Diff,  $(r, \bar{r}) = (1, 5), (2, 6), (2, 8)$  for CDF, and  $r = \bar{r} = K$  for both CW and Daub.

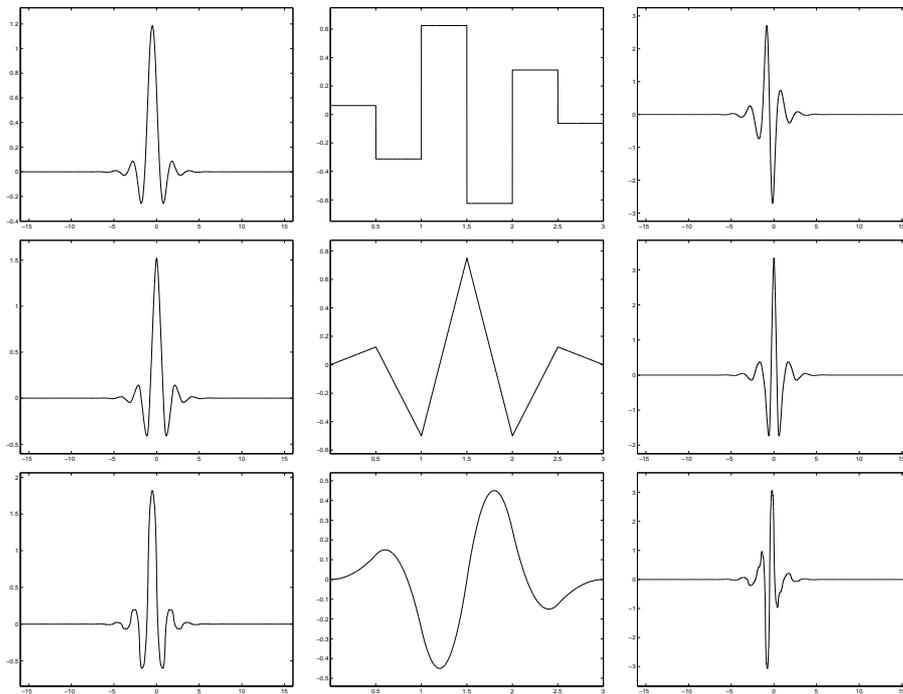


FIG. 2. The profiles of  $\tilde{\phi}$ ,  $\psi$  and  $\tilde{\psi}$  (left to right) of the difference wavelet for  $K = 3$  and  $r = 1, 2, 3$  (top to bottom).

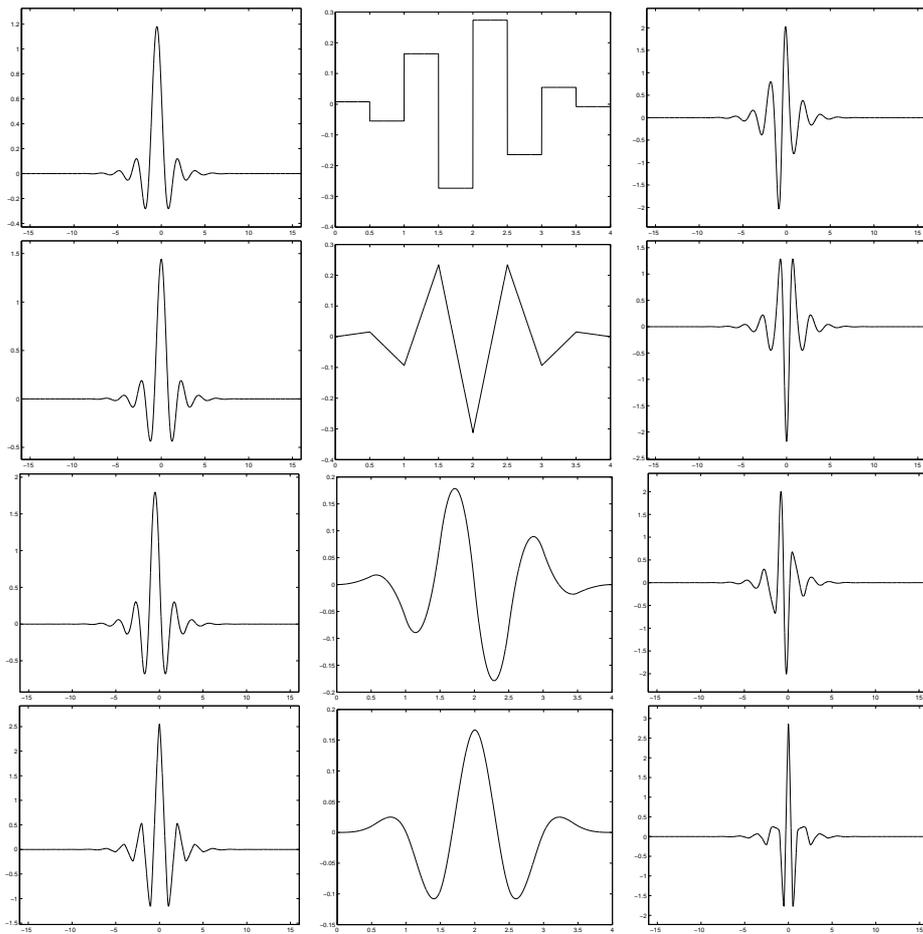


FIG. 3. The profiles of  $\tilde{\phi}$ ,  $\psi$  and  $\tilde{\psi}$  (left to right) of the difference wavelet for  $K = 4$  and  $r = 1, 2, 3, 4$  (top to bottom).