

# BIORTHOGONAL WAVELETS FOR FAST MATRIX COMPUTATIONS

FRITZ KEINERT  
DEPARTMENT OF MATHEMATICS  
IOWA STATE UNIVERSITY  
AMES, IA 50011  
KEINERT@IASTATE.EDU

ABSTRACT. In [1], Beylkin et al. introduced a wavelet-based algorithm that approximates integral or matrix operators of a certain type by highly sparse matrices, as the basis for efficient approximate calculations. The wavelets best suited for achieving the highest possible compression with this algorithm are Daubechies wavelets, while Coiflets lead to a faster decomposition algorithm at slightly lesser compression. We observe that the same algorithm can be based on biorthogonal instead of orthogonal wavelets, and derive two classes of biorthogonal wavelets that achieve high compression and high decomposition speed, respectively. In numerical experiments, these biorthogonal wavelets achieved both higher compression and higher speed than their wavelet counterparts, at comparable accuracy.

## 1. INTRODUCTION

In [1], Beylkin et al. observed that wavelet decomposition can be used to approximate linear operators of a certain type by highly sparse matrices. The decomposition and subsequent calculations can often be done in  $O(N)$  or  $O(N \log N)$  operations for an  $N \times N$  matrix.

For best results, the wavelets should have several vanishing moments, and a support as small as possible. Faster algorithms are possible if the scaling function also has several vanishing moments (except for the zeroth moment, which must be 1). Daubechies wavelets are expected to give the best compression ratio, while Coiflets lead to faster decomposition algorithms at slightly reduced compression.

The algorithm of Beylkin et al. can also be based on biorthogonal wavelets. The algorithm itself and all the relevant estimates from [1] carry over to the biorthogonal case. The main purpose of this paper is to derive biorthogonal counterparts to Daubechies wavelets and Coiflets and compare their performance in operator approximation.

For all six examples in [1], the biorthogonal wavelets exhibited both higher compression factors and faster execution than corresponding orthogonal wavelets, at comparable accuracy. Thus, they present a highly competitive alternative to Daubechies wavelets and Coiflets for these applications.

The organization of the paper is as follows.

In section 2, we briefly review the algorithm of [1] and motivate the search for short biorthogonal wavelets with  $M$  vanishing moments. The equations that these wavelets need to satisfy are listed in section 3.

---

*Date:* April 3, 1996.

*1991 Mathematics Subject Classification.* 65F30, secondary 42C15.

*Key words and phrases.* Fast matrix computations, wavelets, biorthogonal wavelets.

In section 4, we derive the counterpart to Daubechies wavelets: biorthogonal wavelets with a given number  $M$  of vanishing moments, and the shortest possible support. We call them *dual B-spline wavelets*, since the dual scaling function is a B-spline of order  $M$ .

In section 5, we derive a counterpart to Coiflets: biorthogonal wavelets for which the scaling function  $\phi$  and the wavelet  $\psi$  both have  $M$  vanishing moments (except for the zeroth moment of  $\phi$ ), and the shortest possible support. Instead of the most general case, we concentrate on a particular subclass we call *complementary wavelets* for which the operation count for decomposition is cut in half.

Section 6 lists some of the properties of these new wavelets, concentrating on the properties relevant to numerical performance.

In section 7, we repeat the numerical examples from [1] with all four types of wavelets (Daubechies, Coiflets, dual B-spline, complementary).

In section 8, we discuss the results of the first numerical example in more detail, and explain why the biorthogonal case gives better compression.

Finally, section 9 contains a brief summary.

## 2. FAST MATRIX COMPUTATIONS USING WAVELETS

The fundamentals of wavelet theory have been described in many publications, so we refer to the literature for most of the background, and only introduce the specific notation and results we need. A standard reference for wavelets is [7]. Biorthogonal wavelets are discussed in [2].

For orthogonal wavelets, the scaling function  $\phi$  and mother wavelet  $\psi$  are given by the recursion relations

$$\begin{aligned}\phi(x) &= \sqrt{2} \sum_k h_k \phi(2x - k), \\ \psi(x) &= \sqrt{2} \sum_k g_k \phi(2x - k).\end{aligned}$$

Their scaled translates are denoted by

$$\begin{aligned}\phi_k^n(x) &= 2^{n/2} \phi(2^n x - k), \\ \psi_k^n(x) &= 2^{n/2} \psi(2^n x - k).\end{aligned}$$

Throughout this paper, the coefficients are normalized so that

$$\sum h_k = \sum (-1)^k g_k = \sqrt{2}.$$

In the case of biorthogonal wavelet, the dual scaling function and mother wavelet are denoted by  $\tilde{\phi}$  and  $\tilde{\psi}$ , respectively, their recursion coefficients by  $\{\tilde{h}_j\}$ ,  $\{\tilde{g}_j\}$ . The  $m$ th moments of  $\{h_k\}$ ,  $\{g_k\}$  are

$$\begin{aligned}G_m &= \sum_k k^m g_k, \\ H_m &= \sum_k k^m h_k.\end{aligned}$$

If these discrete moments vanish for  $m = 0, 1, \dots, M - 1$  (except for  $H_0$ , which is always  $\sqrt{2}$ ), then so do the corresponding continuous moments of  $\phi, \psi$ :

$$\begin{aligned} \int x^m \psi(x) dx &= 0, \quad m = 0, \dots, M - 1, \\ \int x^m \phi(x) dx &= 0, \quad m = 1, \dots, M - 1. \end{aligned}$$

In [1], Beylkin et al. show that wavelets with several vanishing moments can be used to approximate certain types of linear operators by sparse matrices. The class of operators to which the algorithm applies includes all Calderon-Zygmund and pseudo-differential operators. In many cases, the numerical approximation and subsequent calculations can be achieved in  $O(N)$  operations for the non-standard decomposition, or  $O(N \log N)$  operations for the standard decomposition of an  $N \times N$  matrix.

The basis of the algorithm is the following observation: If  $f$  is any sufficiently smooth function, and if  $\psi$  has  $M$  vanishing moments, then a Taylor expansion of  $f$  around  $x = 2^{-n}j$  yields

$$\begin{aligned} \int f(x) \psi_j^n(x) dx &= 2^{-n/2} \int f(2^{-n}j + 2^{-n}y) \psi(y) dy \\ &\approx 2^{-(n+1)(M+1/2)} \frac{G_M}{M!} f^{(M)}(2^{-n}j) = O(2^{-(n+1)(M+1/2)}). \end{aligned} \quad (2.1)$$

Likewise, if  $\phi$  has  $M$  vanishing moments, then

$$\int f(x) \phi_j^n(x) dx \approx f(2^{-n}j) + 2^{-(n+1)(M+1/2)} \frac{H_M}{M!} f^{(M)}(2^{-n}j). \quad (2.2)$$

Let  $A$  be an integral operator on  $L^2$  of the form

$$Af(x) = \int k(x, y) f(y) dy,$$

where the partial derivatives of  $k$  decrease rapidly away from a small set of singularities. If  $A$  is discretized in a wavelet basis, the calculation of most entries in the resulting matrix involves one or two  $\psi$  integrations in the region where  $k$  is smooth and has small partial derivatives. These entries can be set to 0 with very little error. If  $\phi$  also has several vanishing moments, most of the remaining entries can be replaced by point evaluation to high accuracy, by (2.2). The only entries that need to be explicitly calculated are in a neighborhood of the singularities of  $k$ . Exact estimates for Calderon-Zygmund and pseudo-differential operators can be found in [1].

Similar arguments can be made in the discrete case, where  $A$  is an  $N \times N$  matrix of the form

$$A_{ij} = k(ih, jh)$$

for some stepsize  $h$ , and  $k$  as above. We list the discrete counterparts to (2.1), (2.2) for later reference:

$$\begin{aligned} \sum_j f(jh) h_{j-2i} &= f(2ih) H_0 + f^{(M)}(2ih) \frac{h^M}{M!} H_M + O(h^{M+1}), \\ \sum_j f(jh) g_{j-2i} &= f^{(M)}(2ih) \frac{h^M}{M!} G_M + O(h^{M+1}). \end{aligned} \quad (2.3)$$

An application of these estimates to a specific example can be found in section 8.

What types of wavelets are suitable for this application?

If two different types of wavelets with the same number of vanishing moments are used on the same operator  $A$ , then (2.1) or (2.3) show that the wavelet with smaller  $|G_M|$  will result in a sparser matrix.

The wavelet with fewer coefficients will have the faster execution time. Also, the singularities of  $k$  will affect matrix entries in a band of width  $\approx L$  around the singularities, where  $L$  is the number of wavelet coefficients. Depending on the cutoff value, this may limit achievable compression. Other bands of width  $(L - 2)/2$  will appear along some edges if end-point corrections are not used. Thus, a shorter wavelet may have slightly better compression as well.

For fixed  $M$ , Daubechies wavelets are the shortest and therefore well suited for this application. If execution time is to be minimized by using wavelets with vanishing moments of  $\psi$  and  $\phi$ , Coiflets will work well.

Biorthogonal wavelets can be used for many of the same applications as orthogonal wavelets. In particular, orthogonality is not needed in the algorithm of Beylkin et al. The entire argument and the error estimates carry over to biorthogonal wavelets. For practical implementation, the only difference is that two sets of coefficients are needed in the biorthogonal case, one for decomposition and another for reconstruction.

Biorthogonal wavelets can be shorter than orthogonal wavelets for the same number of vanishing moments, so we expect faster algorithms, with comparable compression and accuracy. We will demonstrate in later sections that this is indeed the case. As an unexpected bonus, we also obtain better compression.

### 3. FORMULATION OF THE PROBLEM

As motivated in the preceding section, we pose the following questions:

- What are the biorthogonal wavelets of smallest support for which  $\psi$  has  $M$  vanishing moments?
- What are the biorthogonal wavelets of smallest support for which  $\phi$  and  $\psi$  both have  $M$  vanishing moments?
- How do these wavelets perform in practice, in terms of compression ratio, execution speed and accuracy?

To derive the coefficients of suitable wavelets, we introduce the auxiliary functions

$$\begin{aligned} h(z) &= \frac{1}{\sqrt{2}} \sum_k h_k z^k, \\ g(z) &= \frac{1}{\sqrt{2}} \sum_k g_k z^k, \\ \tilde{h}(z) &= \frac{1}{\sqrt{2}} \sum_k \tilde{h}_k z^k, \\ \tilde{g}(z) &= \frac{1}{\sqrt{2}} \sum_k \tilde{g}_k z^k, \end{aligned}$$

where  $z = e^{-i\xi}$ . These functions are also used in Daubechies [5], [7], Cohen et al. [2], and other papers.

In terms of these functions, the desired properties of coefficient sequences are stated as follows.

**Biorthogonality:**

$$\begin{pmatrix} h(z) & h(-z) \\ g(z) & g(-z) \end{pmatrix} \begin{pmatrix} \overline{\tilde{h}(z)} & \overline{\tilde{g}(z)} \\ \overline{\tilde{h}(-z)} & \overline{\tilde{g}(-z)} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \quad (3.1)$$

**Vanishing Moments for  $\{g_j\}$ :**

$$g(1) = g'(1) = \dots = g^{(M-1)}(1) = 0. \quad (3.2)$$

**Vanishing Moments for  $\{h_j\}$ :**

$$\begin{aligned} h(1) &= 1, \\ h'(1) &= \dots = h^{(M-1)}(1) = 0. \end{aligned} \quad (3.3)$$

**Minimal Smoothness:**

$$h(-1) = 0. \quad (3.4)$$

Condition (3.1) is necessary for the decomposition/reconstruction algorithm to work. Condition (3.2) is necessary to achieve compression. The condition  $h(1) = 1$  in (3.3) and condition (3.4) are only strictly necessary if we want functions  $\phi$ ,  $\psi$  to be associated with the coefficients; even without them, the algorithm would work algebraically. However, it has been observed in numerical experiments by the author and others (e.g. Daubechies [7]) that coefficient sequences which violate these conditions do not perform well. Finally, the remaining conditions in (3.3) are necessary for addressing our second goal, the biorthogonal counterparts to Coiflets.

For practical applications, we want coefficient sequences  $\{h_k\}$ ,  $\{g_k\}$ ,  $\{\tilde{h}_k\}$ ,  $\{\tilde{g}_k\}$  of finite length. The functions  $h(z)$ ,  $g(z)$ ,  $\tilde{h}(z)$ ,  $\tilde{g}(z)$  are then trigonometric polynomials.

Let

$$\Delta(z) = h(z)g(-z) - h(-z)g(z)$$

be the determinant of the first matrix in (3.1).  $\Delta(z)$  and  $1/\Delta(z)$  are both trigonometric polynomials, so we must have

$$\Delta(z) = \alpha z^m \quad \text{for some } m \in \mathbb{Z}, \alpha \neq 0.$$

Applying the inversion formula for  $2 \times 2$  matrices to (3.1) shows that  $m$  must be odd.

Furthermore, we can assume without loss of generality that  $\alpha = 1$ . The scaling of  $h$  is fixed, since  $h(1) = 1$ , but any value of  $\alpha$  different from 1 can be interpreted as a scaling of  $g$ , and inverse scaling of  $\tilde{g}$ . Choosing  $\alpha = 1$  leads to  $\{g_j\}$  of the same order of magnitude as  $\{h_j\}$ . It also implies  $g(-1) = 1$ .

Thus, we require

$$\Delta(z) = z^{2k+1} \quad \text{for some } k \in \mathbb{Z} \quad (3.5)$$

as a necessary condition for the existence of suitable biorthogonal wavelets.

Condition (3.5) is also sufficient. If  $g$ ,  $h$  satisfy (3.5), and the dual coefficients are calculated from

$$\begin{aligned} \tilde{h}(z) &= \Delta(z) \overline{g(-z)}, \\ \tilde{g}(z) &= -\Delta(z) \overline{h(-z)}, \end{aligned} \quad (3.6)$$

condition (3.1) can be verified directly.

For biorthogonal wavelets of compact support we can therefore replace (3.1) by (3.5). After suitable  $\{h_j\}$ ,  $\{g_j\}$  have been found, we can calculate  $\{\tilde{h}_j\}$ ,  $\{\tilde{g}_j\}$  from (3.6).

In practice, there are usually several wavelets satisfying these conditions for a given  $M$ , corresponding to different choices of  $k$ . Among all the possible solutions for fixed  $M$ , we select the one for which  $|G_M|$  is smallest. This leads to the highest possible compression. If there are several candidates, we select the numerically most stable one. This choice is supported by both theoretical considerations and practical experiments. Further details on numerical stability questions are given in section 6.

#### 4. DUAL B-SPLINE WAVELETS

Our first approach finds the biorthogonal wavelets with  $M$  vanishing moments that have the smallest possible support, as in the approach of Daubechies in [5]. From (3.2) and  $g(-1) = 1$ , it follows that  $g(z)$  must be of the form

$$g(z) = \left(\frac{1-z}{2}\right)^M. \quad (4.1)$$

Since

$$\tilde{h}(z) = \Delta(z)g(-1/z) = z^{2k+1-M} \left(\frac{1+z}{2}\right)^M,$$

the  $\tilde{h}$  coefficients are scaled and shifted binomial coefficients, and the corresponding  $\tilde{\phi}$  is a B-spline (see Chui [4]). For this reason, we call the wavelets associated with these coefficients *dual B-spline wavelets* in this paper.

We assume a general form for  $h(z)$

$$h(z) = (h_0 + h_1z + \cdots + h_Nz^N)z^p$$

for some integers  $p$ ,  $N$  and coefficients  $\{h_j\}$ .

With these assumptions,

$$\begin{aligned} \Delta(z) &= h(z)g(-z) - h(-z)g(z) \\ &= 2^{-M}z^p \left[ \sum_{j=0}^N \sum_{k=0}^M \binom{M}{k} h_j z^{j+k} - (-1)^p \sum_{j=0}^N \sum_{k=0}^M \binom{M}{k} (-1)^{k+j} h_j z^{j+k} \right] \\ &= 2^{-M}z^p \sum_{j=0}^N \sum_{k=0}^M \binom{M}{k} h_j z^{j+k} [1 - (-1)^p (-1)^{j+k}]. \end{aligned} \quad (4.2)$$

For even (resp. odd)  $p$ , all the terms in the sum with  $(j+k)$  even (resp. odd) vanish automatically.

We need to satisfy

$$\begin{aligned} \Delta(z) &= z^{2k+1} \quad \text{for some } k \in \mathbb{Z}, \\ h(-1) &= 0. \end{aligned}$$

The remaining conditions  $g(-1) = h(1) = 1$  are then automatically satisfied.

Counting equations, we find that  $(N+1)$  coefficients  $\{h_j\}$  need to satisfy between  $(M+N+2)/2$  and  $(M+N+4)/2$  equations, depending on whether  $p$  and  $(M+N)$  are even or odd. The shortest possible length for  $\{h_j\}$  is achieved for even  $p$ , with  $N = M$ .

It is sufficient to consider  $p = 0$ , since replacing any  $h(z)$  by  $h(z)z^{2p}$  results in a simple shift in the indices of all coefficients. In the decomposition algorithm, this results in corresponding shifts in the indices of matrix entries, at least at the first level of decomposition. At later

levels the effect is more complicated, but no noticeable effects on compression ratio or error were observed in numerical tests. Thus, we look for  $h(z)$  of the form

$$h(z) = h_0 + h_1z + \dots + h_Mz^M$$

From (4.2), a solution is only possible if  $0 \leq 2k + 1 \leq 2M$ , that is  $0 \leq k \leq M - 1$ . We can further restrict attention to

$$0 \leq k \leq \left\lceil \frac{M-1}{2} \right\rceil,$$

where  $\lceil x \rceil$  denotes the largest integer  $\leq x$ . In fact, it is easy to show that the  $\{h_j\}$  coefficients for  $\Delta = z^{2(M-k-1)+1}$  are simply the  $\{h_j\}$  coefficients for  $\Delta = z^{2k+1}$  in reverse. In numerical tests, the reversed and original sequences led to virtually the same compression factor and error.

The equations were constructed and solved using `mathematica`, a computer algebra system. For small values of  $M$  ( $M \leq 6$ ), all resulting systems of equations were found to have unique solutions  $h_j$ . All solutions for the same  $M$  share the same  $g$  coefficients and the same  $G_M$ , so selection was done on the basis of stability alone. In all cases, the most stable wavelets corresponded to the choice  $k = \lceil (M-1)/2 \rceil$ .

For  $M = 1$ , we obtain the Haar wavelets. For  $M = 2 \dots 6$ , the resulting  $\{h_j\}$  are shown in table 1. The  $\{g_j\}$  can be found from (4.1), the  $\{\tilde{h}_j\}$  and  $\{\tilde{g}_j\}$  from (3.6).

## 5. COMPLEMENTARY WAVELETS

Our second approach attempts to find biorthogonal wavelets with  $M$  vanishing moments for both  $g$  and  $h$ , similar to the derivation of Coiflets (Daubechies [6]).

There are many more solutions for each given  $M$  than in the previous case. Some experimentation with the equations yielded many solutions with the property

$$h(z) = 1 - g(z). \tag{5.1}$$

This property reduces the amount of computation during decomposition/reconstruction by almost 50%, so we limited our investigation to this case. The moment condition (3.3) for  $\{h_j\}$  is automatically satisfied if the  $\{g_j\}$  have  $M$  vanishing moments.

From

$$\Delta(z) = [1 - g(z)]g(-z) - [1 - g(-z)]g(z) = g(-z) - g(z) = z^{2k+1}$$

we see that  $g(z)$  must have the form

$$g(z) = (g_0 + g_2z^2 + \dots + g_{2N}z^{2N})z^{2p} - \frac{1}{2}z^{2k+1},$$

from which we verify that

$$g(-1) = 1, \quad h(-1) = 0.$$

Since  $g$  must satisfy the  $M$  moment conditions (3.2), we must have  $N \geq (M-1)$ . To get the shortest possible sequences, we choose  $N = (M-1)$  and  $p \leq k \leq M+p-2$ . This gives coefficient sequences of length  $(2M-1)$ , only  $(M+1)$  of which are nonzero.

Again, all choices of  $M, p, k$  we tried lead to unique solutions. Among the many possible solutions for fixed  $M$ , we select the unique solution with smallest condition number among the solutions with smallest  $|G_M|$ . For  $M = 1, 2$ , we obtain the same solutions as for the dual B-spline wavelets. The coefficients that give the best performance, for  $M = 3 \dots 6$ , are listed in table 2. We did not pursue higher values of  $M$  at this point.

## 6. COMPARISON OF WAVELETS

In this section, we list some of the properties of dual B-spline and complementary wavelets, and compare and contrast them with Daubechies wavelets and Coiflets.

The first three subsections are directly related to the numerical performance: size of first non-vanishing moment, computational effort, and numerical stability. The remaining subsection on smoothness is only included for general interest.

**6.1. First non-vanishing moment.** As outlined in section 2, the degree of compression possible depends on  $|G_M|$ , the first non-vanishing moment of  $\{g_j\}$ . For better compression,  $|G_M|$  should be as small as possible. Biorthogonal wavelets compare favorably to orthogonal wavelets by this measure, as shown in table 3. The exact effect of  $|G_M|$  on compression is detailed in section 8, for one example.

**6.2. Sequence Length and Computational Effort.** Wavelet coefficient sequences should be as short as possible for faster execution, and possibly better compression.

Table 4 compares the number of wavelet coefficients necessary to achieve  $M$  vanishing moments, and the number of multiplications necessary to multiply by  $\{h_j\}$  and  $\{g_j\}$  once. For the first three wavelets, the number of multiplications is simply twice the sequence length, but complementary wavelets are special in two respects: the sequences contain some zeros, and the  $\{h_j\}$  and  $\{g_j\}$  are highly correlated, which saves half the multiplications.

The execution times listed in example 1 in section 7 correspond well with the ratios listed in the last column.

**6.3. Numerical Stability.** Decomposition and reconstruction with orthogonal wavelets is a very stable numerical process. When using biorthogonal wavelets, we must consider stability more carefully. In fact, some of the coefficient sequences that satisfy the conditions in sections 4 and 5 turn out to be unstable in practice.

An illustration of this is given in table 5, which lists the worst  $\ell_2$  error observed in decomposing and reconstructing random vectors of length 16,384 through 12 levels. The coefficients for Daubechies wavelets and Coiflets were taken from Daubechies [7], where they are given to 12 decimals or less; this explains their lower accuracy. The calculations were done in double precision on a DEC 5000 workstation; anything of order  $10^{-14}$  represents machine accuracy.

We observe that all the complementary wavelets we considered, as well as the dual B-spline wavelets for  $M = 2, 3$  appear to be numerically stable. Higher dual B-spline wavelets show more and more pronounced instabilities, and should be used with caution.

Exact stability estimates for biorthogonal wavelets are derived in a forthcoming paper [8]. For the purposes of the present paper, a good indication of the stability behavior is the  $\|\cdot\|_2$ -condition number of a single level of decomposition or reconstruction, which can be explicitly calculated (see table 6). Small condition numbers correspond to wavelets that are stable in practice.

In the examples in section 7, the slightly larger errors observed for higher order B-spline wavelets in some cases may be due to beginning instability, but the effect is not very pronounced and can be removed by choosing smaller cutoffs. The compression ratio goes down a little in this case, but remains better than for orthogonal wavelets.

**6.4. Regularity.** Dual B-spline and complementary wavelets were derived for their vanishing moments. Regularity of  $\phi$ ,  $\psi$ , or even their existence as functions, is not required. We include only a very simple estimate.

There are various ways to estimate the regularity of the functions  $\phi$ ,  $\psi$  produced by a sequence of wavelet coefficients. Lemma (7.1.7) in Daubechies [7] applies to all biorthogonal wavelets considered in this paper and provides sharp estimates of the optimal decay of  $\hat{\phi}$ , i.e. the smallest  $\alpha$  such that

$$|\hat{\phi}(\xi)| \leq C(1 + |\xi|)^\alpha$$

for some constant  $C$  and all  $\xi$ . Smaller  $\alpha$  correspond to more regularity;  $\alpha < -1$  implies continuity. The exponents  $\alpha$  are listed table 7

We observe that we cannot infer continuity for any of these wavelets. The dual B-spline wavelets appear to become less regular with increasing  $M$ , while all complementary wavelets have about the same regularity. This coincides with the observations on numerical stability.

We plotted the graphs of some of these wavelets by a standard binary subdivision scheme, starting with the values of  $\phi$  at the integers. In all cases, the peaks tend to  $\infty$  during subdivision, which again indicates a lack of continuity. An approximate plot of the dual B-spline scaling function and wavelet for  $M = 2$ , using 256 points, is shown in figure 1.

## 7. NUMERICAL EXPERIMENTS

To test the performance of algorithms based on the wavelets derived above, we applied them to the six matrices from Beylkin et al. [1]. In all cases, we calculated the nonstandard decomposition of the matrix  $A$ , compressed it by setting values below the cutoff to zero, applied the compressed matrix to the decomposition of a random vector  $v$ , reconstructed the result and compared it to the usual matrix-vector product  $Av$ . The error incurred is measured as the relative  $\ell_2$  and relative  $\ell_\infty$  error in  $Av$ .

In each example, we chose the number of vanishing moments  $M$  and the cutoff as in [1], and used the Daubechies, dual B-spline and complementary wavelets for that  $M$ , as well as Coiflets in the case of even  $M$ .

Shown below are the results for  $256 \times 256$  matrices, for one choice of random vector. Results for other random vectors, other matrix sizes, other smooth matrices, different cutoffs or the standard instead of nonstandard decomposition were similar and are not shown. All calculations were done in double precision on a DEC 5000 workstation.

The numbers obtained for Daubechies wavelets closely match the results in [1], except for the compression factor for matrix 3. We have no explanation for this discrepancy.

In all examples, biorthogonal wavelets show similar or better compression with comparable errors, and are faster to evaluate. The higher compression rates obtained with biorthogonal wavelets are examined in more detail in section 8, for the matrix in example 1. Examples 2 through 6 could be treated similarly.

EXAMPLE 1: The matrix is given by

$$A_{ij} = \begin{cases} \frac{1}{i-j} & \text{if } i \neq j \\ 0 & \text{if } i = j. \end{cases}$$

We used wavelets with six vanishing moments and a cutoff of  $10^{-7}$ . The structure of the remaining non-zero entries for this example is shown in figure 2.

In this example, we show the computer time necessary for decomposition and compression. We have not implemented the faster algorithm possible with Coiflets and complementary wavelets yet, so these timings are nearly proportional to the number of multiplications per application of  $\{h_j\}$ ,  $\{g_j\}$  (last column in table 4).

EXAMPLE 2: The matrix is given by

$$A_{ij} = \begin{cases} \frac{\log|i - 2^{n-1}| - \log|j - 2^{n-1}|}{i - j} & \text{if } i \neq j \text{ and } i, j \neq 2^{n-1}, \\ 0 & \text{otherwise,} \end{cases}$$

where  $i, j = 1, \dots, N$  and  $N = 2^n$ . We used wavelets with six vanishing moments and a cutoff of  $10^{-7}$ .

EXAMPLE 3: The matrix is given by

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

where  $\Lambda(z) = \Gamma(z + 1/2)/\Gamma(z + 1)$  and  $\Gamma(z)$  is the gamma function.

We used wavelets with 5 vanishing moments and a cutoff of  $10^{-6}$ .

EXAMPLE 4: The matrix is given by

$$A_{ij} = \begin{cases} \log(i - j)^2 & \text{if } i \neq j, \\ 0 & \text{if } i = j. \end{cases}$$

We used wavelets with six vanishing moments and a cutoff of  $10^{-6}$ .

EXAMPLE 5: The matrix is given by

$$A_{ij} = \begin{cases} \frac{1}{i - j + (\cos ij)/2} & \text{if } i \neq j \\ 0 & \text{if } i = j. \end{cases}$$

We used wavelets with two vanishing moments and a cutoff of  $10^{-3}$ . The dual B-spline and complementary wavelets are identical in this case.

EXAMPLE 6: The matrix is given by

$$A_{ij} = \begin{cases} \frac{i \cos(\log i^2) - j \cos(\log j^2)}{(i - j)^2} & \text{if } i \neq j \\ 0 & \text{if } i = j. \end{cases}$$

We used wavelets with two vanishing moments and a cutoff of  $10^{-3}$ . The dual B-spline and complementary wavelets are identical in this case.

## 8. WHY DO BIORTHOGONAL WAVELETS LEAD TO HIGHER COMPRESSION?

The main reason for the derivation of dual B-spline and complementary wavelets was improved execution speed due to shorter length. The observed higher compression was an unexpected bonus.

In this section, we explain this result by estimating the number of remaining non-zero entries in the decomposed matrix in example 1. The estimates coincide closely with the actually observed values.

In example 1, the matrix  $A$  has values  $A_{ij} = k(ih, jh)$  with  $k(x, y) = 1/(x - y)$ , and  $h = 1$ . Let  $A^{(1)}$  be the matrix obtained from  $A$  after one level of decomposition. Using

estimates (2.3) for  $M = 6$ , we find for  $i, j = 0, \dots, n/2 - 1$

$$\begin{aligned} A_{ij}^{(1)} &\approx 2k(2ih, 2jh) = \frac{1}{i-j}, \\ A_{n/2+i,j}^{(1)} &\approx \sqrt{2} \frac{h^6}{6!} G_6 \frac{\partial^6}{\partial y^6} k(2ih, 2jh) = 2^{-13/2} G_6 \frac{1}{(i-j)^7}, \\ A_{i,n/2+j}^{(1)} &\approx \sqrt{2} \frac{h^6}{6!} G_6 \frac{\partial^6}{\partial x^6} k(2ih, 2jh) = 2^{-13/2} G_6 \frac{1}{(i-j)^7}, \\ A_{n/2+i,n/2+j}^{(1)} &\approx \frac{h^{12}}{(6!)^2} G_6^2 \frac{\partial^{12}}{\partial x^6 \partial y^6} k(2ih, 2jh) = 924 \cdot 2^{-13} G_6^2 \frac{1}{(i-j)^{13}} \end{aligned}$$

A cutoff of  $10^{-7}$  leads to an estimated bandwidth of  $10 \left( \sqrt{2} |G_6| \right)^{1/7}$  for the top right and bottom left quadrants (second and third estimates), and  $(924 \cdot G_6^2 \cdot 10^7)^{1/13}$  for the bottom right (fourth estimate). Estimated and observed bandwidth are listed in table 14. There is excellent agreement between predicted and observed values. The lower limit on the bandwidth due to sequence length is not reached in this example.

For the next two levels of decomposition, the estimates remain the same, and observed bandwidths do not deviate much from the estimates.

Since we have not yet implemented the end-point correction as described e.g. in [3], compression is influenced in a lesser way by bands along some edges. The bandwidth at the first level is  $(L - 2)/2$  for wavelets with  $L$  coefficients.

For Daubechies wavelets, assuming bands of width 24 or 14 along the diagonals, bands of width 5 along two edges, and a full remaining submatrix in the top left-hand corner, we estimate a compression ratio of 3.82, very close to the actually achieved 3.9.

The edge effects are small compared to the effects of the singularity, and could be removed by end-point corrections. The main reason for the better compression obtained with biorthogonal wavelets is the size of  $|G_M|$ .

## 9. SUMMARY

We observe that the matrix decomposition algorithm of Beylkin et al., which is based on orthogonal wavelets, can also be performed with biorthogonal wavelets. Error estimates indicate that the best possible compression can be expected by using wavelet coefficient sequences  $\{g_j\}$ ,  $\{h_j\}$  which are as short as possible and for which the  $\{g_j\}$  have several vanishing moments. The decomposition algorithm will work much faster, at the cost of reduced compression, if the  $\{h_j\}$  also have several vanishing moments. For orthogonal wavelets, this means one should use either Daubechies wavelets or Coiflets. For biorthogonal wavelets, we derived two corresponding classes of wavelets with fewer coefficients than their orthogonal counterparts.

Numerical experiments and theoretical estimates show that these biorthogonal wavelets give better compression with comparable errors, while simultaneously leading to faster algorithms.

**Acknowledgment.** The author would like to thank the referees of the original version for their suggestions which greatly improved the organization of this paper.

## REFERENCES

- [1] Gregory Beylkin, Ravi Coifman, and Vladimir Rokhlin, *Fast wavelet transforms and numerical algorithms: I*, Comm. Pure Appl. Math. **44** (1991), no. 2, 141–183.
- [2] A. Cohen, Ingrid Daubechies, and J.-C. Feauveau, *Biorthogonal bases of compactly supported wavelets*, Comm. Pure Appl. Math. **45** (1992), 485–560.
- [3] A. Cohen, I. Daubechies, and P. Vial, *Wavelets and fast wavelet transform on the interval*, preprint.
- [4] Charles K. Chui, *An introduction to wavelets*, Wavelet Analysis and Its Applications, vol. 1, Academic Press, San Diego, 1992.
- [5] Ingrid Daubechies, *Orthonormal bases of compactly supported wavelets*, Comm. Pure Appl. Math. **41** (1988), 909–996.
- [6] Ingrid Daubechies, *Orthonormal bases of compactly supported wavelets II: Variations on a theme*, SIAM J. Math. Anal. **24** (1993), no. 2, 499–519.
- [7] Ingrid Daubechies, *Ten lectures on wavelets*, CBMS-NSF Regional Conference Series in Applied Mathematics, vol. 61, SIAM, Philadelphia, 1992.
- [8] F. Keinert, *Numerical stability of biorthogonal wavelets*, in preparation.

TABLE 1. Coefficients for dual B-spline wavelets. Each coefficient should be multiplied by the factor below it.

	$M = 2$	3	4	5	6
$h_0$	3	-1	-5	3	7
$h_1$	2	3	20	-15	-42
$h_2$	-1	3	10	20	77
$h_3$		-1	-12	20	28
$h_4$			3	-15	-63
$h_5$				3	30
$h_6$					-5
factor	$\sqrt{2}/4$	$\sqrt{2}/4$	$\sqrt{2}/16$	$\sqrt{2}/16$	$\sqrt{2}/32$

TABLE 2. Coefficients for complementary wavelets. Each coefficient must be multiplied by the factor below it.

	$M = 3$	4	5	6
$h_{-4}$			-3	-3
$h_{-3}$			0	0
$h_{-2}$	1	1	20	25
$h_{-1}$	0	0	0	0
$h_0$	10	23	166	362
$h_1$	8	16	128	256
$h_2$	-3	-9	-60	-150
$h_3$		0	0	0
$h_4$		1	5	25
$h_5$				0
$h_6$				-3
factor	$\sqrt{2}/16$	$\sqrt{2}/32$	$\sqrt{2}/256$	$\sqrt{2}/512$

TABLE 3. First non-vanishing moments  $G_M$  for  $M = 2, \dots, 6$

	$M = 2$	3	4	5	6
Daubechies	-1.225	-3.354	-12.550	-59.529	-341.970
Coiflets	1.164		-11.185		386.540
dual B-spline	0.707	-1.061	2.121	-5.303	15.910
complementary	0.707	2.121	-6.364	-31.820	159.099

TABLE 4. Length of coefficient sequences and number of multiplications necessary for one discrete convolution with both  $\{g_j\}$  and  $\{h_j\}$ , for wavelets with  $M$  vanishing moments.

	length	multiplications
Daubechies	$2M$	$4M$
Coiflets	$3M$	$6M$
dual B-spline	$M + 1$	$2M + 1$
complementary	$2M - 1$	$M + 1$

TABLE 5. Worst relative  $\ell_2$  error due to roundoff observed in decomposing and reconstructing random vectors of length  $2^{14}$  (12 levels of decomposition and reconstruction).

	$M = 2$	3	4	5	6
Daubechies	$0.165 \cdot 10^{-14}$	$0.793 \cdot 10^{-10}$	$0.176 \cdot 10^{-10}$	$0.212 \cdot 10^{-10}$	$0.213 \cdot 10^{-10}$
Coiflet	$0.261 \cdot 10^{-11}$		$0.381 \cdot 10^{-10}$		$0.964 \cdot 10^{-12}$
dual B-spline	$0.245 \cdot 10^{-14}$	$0.519 \cdot 10^{-14}$	$0.137 \cdot 10^{-12}$	$0.299 \cdot 10^{-10}$	$0.100 \cdot 10^{-06}$
complementary	$0.245 \cdot 10^{-14}$	$0.206 \cdot 10^{-14}$	$0.153 \cdot 10^{-14}$	$0.205 \cdot 10^{-14}$	$0.226 \cdot 10^{-14}$

TABLE 6.  $\|\cdot\|_2$ -condition number of one level of decomposition

	$M = 2$	3	4	5	6
dual B-spline	2.618	4.000	9.141	16.000	38.765
complementary	2.618	1.768	2.618	2.071	2.618

TABLE 7. Optimal decay exponent  $\alpha$  in the estimate  $|\hat{\phi}(\xi)| \leq C(1 + |\xi|)^\alpha$ .

	$M = 2$	3	4	5	6
Dual B-spline	-0.14978	0.32193	1.14328	1.72792	2.53818
Complementary	-0.14978	-0.35269	-0.10198	-0.20049	-0.070254

TABLE 8. Numerical results for example 1

	compression factor	$\ell_2$ error	$\ell_\infty$ error	time (seconds)
Daubechies	3.9	$1.255 \cdot 10^{-7}$	$3.990 \cdot 10^{-7}$	6.66
Coiflet	3.7	$1.276 \cdot 10^{-7}$	$6.229 \cdot 10^{-7}$	9.75
dual B-spline	5.1	$6.219 \cdot 10^{-7}$	$1.745 \cdot 10^{-6}$	4.12
complementary	4.2	$7.925 \cdot 10^{-8}$	$2.172 \cdot 10^{-7}$	2.01

TABLE 9. Numerical results for example 2

	compression factor	$\ell_2$ error	$\ell_\infty$ error
Daubechies	3.8	$1.339 \cdot 10^{-7}$	$5.377 \cdot 10^{-7}$
Coiflet	3.6	$1.470 \cdot 10^{-7}$	$6.522 \cdot 10^{-7}$
dual B-spline	5.3	$5.482 \cdot 10^{-7}$	$1.837 \cdot 10^{-6}$
complementary	3.9	$1.430 \cdot 10^{-7}$	$4.839 \cdot 10^{-7}$

TABLE 10. Numerical results for example 3

	compression factor	$\ell_2$ error	$\ell_\infty$ error
Daubechies	6.5	$1.214 \cdot 10^{-6}$	$4.583 \cdot 10^{-6}$
dual B-spline	9.9	$1.547 \cdot 10^{-6}$	$5.661 \cdot 10^{-6}$
complementary	6.8	$1.486 \cdot 10^{-6}$	$1.046 \cdot 10^{-5}$

TABLE 11. Numerical results for example 4

	compression factor	$\ell_2$ error	$\ell_\infty$ error
Daubechies	3.9	$3.522 \cdot 10^{-6}$	$8.791 \cdot 10^{-6}$
Coiflet	3.6	$3.282 \cdot 10^{-6}$	$8.287 \cdot 10^{-6}$
dual B-spline	6.3	$3.208 \cdot 10^{-6}$	$9.795 \cdot 10^{-6}$
complementary	4.2	$2.407 \cdot 10^{-6}$	$6.163 \cdot 10^{-6}$

TABLE 12. Numerical results for example 5

	compression factor	$\ell_2$ error	$\ell_\infty$ error
Daubechies	6.6	$1.989 \cdot 10^{-3}$	$7.314 \cdot 10^{-3}$
Coiflet	6.4	$2.647 \cdot 10^{-3}$	$1.211 \cdot 10^{-2}$
dual B-spline	6.8	$2.027 \cdot 10^{-3}$	$5.795 \cdot 10^{-3}$

TABLE 13. Numerical results for example 6

	compression factor	$\ell_2$ error	$\ell_\infty$ error
Daubechies	8.2	$3.491 \cdot 10^{-3}$	$1.400 \cdot 10^{-2}$
Coiflet	8.2	$3.382 \cdot 10^{-3}$	$1.448 \cdot 10^{-2}$
dual B-spline	10.2	$2.450 \cdot 10^{-3}$	$1.112 \cdot 10^{-2}$

TABLE 14. Width of bands around singularity in example 1, for wavelets with 6 vanishing moments.

	$G_6$	top right/bottom left		bottom right	
		estimated	observed	estimated	observed
Daubechies	-341.98	24	24	14	15
Coiflet	292.41	24	24	14	17
dual B-spline	15.91	16	16	9	9
complementary	159.10	22	22	13	15

FIGURE 1. Dual B-spline scaling function and wavelet, 2 vanishing moments

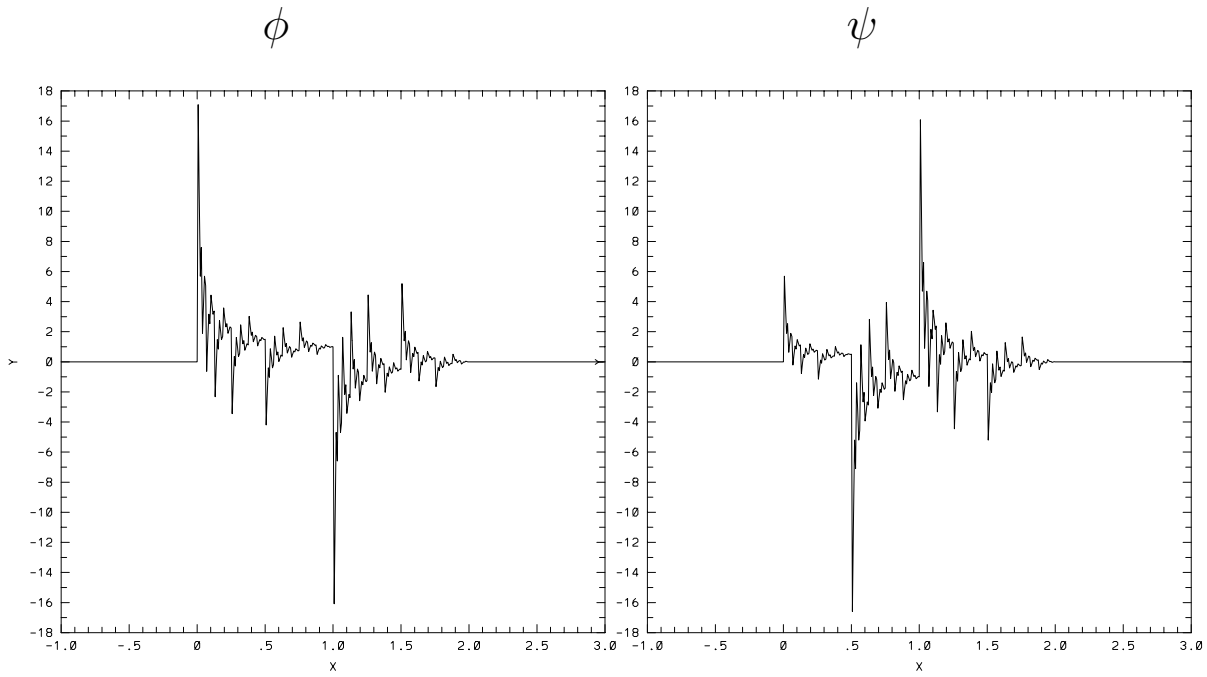


FIGURE 2. Sparse matrix structure for example 1

