Discrete wavelet transforms for Toeplitz matrices

Fu-Rong Lin a,b,1, Wai-Ki Ching b,2, Michael K. Ng b,*,3

aDepartment of Mathematics, Shantou University, Shantou, Guangdong 515063, PR China
bDepartment of Mathematics, The University of Hong Kong, Pokfulam road, Hong Kong

Received 18 November 2002; accepted 18 January 2003

Submitted by R.J. Plemmons

Abstract

In this paper, we discuss discrete wavelet transforms for Toeplitz matrices and block–Toeplitz–Toeplitz–block matrices. The main contribution of this paper is to give the Toeplitz-like structure of the wavelet transformed Toeplitz matrices, and show that the computational cost for such structure is $O(k^3\ln n)$ where $n$ is the size of the Toeplitz matrix, $k$ is the order of the wavelet and $l$ is the level used in the wavelet transform. The comparison between the wavelet transformed Toeplitz matrices and the Fourier transformed Toeplitz matrices is also given.

Keywords: Toeplitz matrix; Circulant matrix; Discrete wavelet transforms; Discrete Fourier transform; Toeplitz-like structure

1. Introduction

Wavelets can be viewed as the synthesis of various ideas originating from different disciplines including mathematics, physics, and engineering, see for instance [9]. Recently, wavelet transforms, particularly those for orthonormal wavelets with compact support, have emerged as a new mathematical tool in signal processing, image
processing, approximation theory, numerical analysis, etc. [9]. The most remarkable property about wavelets is that the whole wavelet basis is determined by a single “element”: in continuous wavelets, the basis functions are determined by “mother wavelet”; in discrete wavelets, the wavelet coefficients matrix is defined by wavelet filter coefficients, cf. (3) and (5) in Section 2.

The discrete wavelet transform (DWT) algorithm was first introduced by Mallat [15]. The DWT consists of applying a wavelet coefficients matrix hierarchically, first to the full data vector of length \( n \), then to the “smooth” vector \( \frac{n}{2} \), then to the “smooth–smooth” vector of length \( \frac{n}{4} \) until only trivial number of very smooth components remain. The procedure is called a pyramidal algorithm, see for instance [15,16]. The computational cost of the DWT to an \( n \)-vector \( x \) is \( O(n) \) operations which is less than \( O(n \log n) \) operations, the computational cost required by the fast Fourier transform (FFT) to an \( n \)-vector. It is interesting to note that the DWT is invertible (i.e., \( W_n^{-1} \) exists) and can also be orthogonal for orthonormal wavelets (i.e., \( W_n^T W_n = W_n W_n^T = I_n \)) [7]. Here \( I_n \) and \( W_n \) are identity matrix and wavelet transform matrix respectively. In the following discussion, we are interested in orthogonal DWTs.

For an \( n \)-by-\( n \) matrix \( A_n \), the matrix \( W_n A_n W_n^T \), which is the representation of two-dimensional data under discrete wavelet basis, is called the wavelet transformed matrix of \( A_n \). One of the most interesting and promising wavelet applications is to compress the matrix data, i.e., the wavelet transformed matrix \( W_n A_n W_n^T \) can be very sparse even though \( A_n \) is dense. For example, let \( A_n \) be a Calderón–Zygmund matrix, then the wavelet transformed matrix \( W_n A_n W_n^T \) is sparse in the sense that a large fraction of entries are very small and can be regarded as negligible, see for instance [3].

In this paper, we study DWTs for Toeplitz matrices. An \( n \)-by-\( n \) matrix \( T_n \) is said to be Toeplitz if

\[
T_n = \begin{bmatrix}
    t_0 & t_{-1} & \cdots & t_{2-n} & t_{1-n} \\
    t_1 & t_0 & t_{-1} & \cdots & t_{2-n} \\
    \vdots & t_1 & t_0 & \ddots & \vdots \\
    t_{n-2} & \cdots & t_{-1} & \cdots & t_1 \\
    t_{n-1} & t_{n-2} & \cdots & t_1 & t_0
\end{bmatrix},
\]

where \( T_n \) is constant along its diagonals. The name Toeplitz originates from the work of Otto Toeplitz in the early 1900’s on bilinear forms related to Laurent series, see Grenander and Szegö [10] for instance. If we replace each \( t_i \) entry in (1) by an \( m \times m \) matrix, then we call the corresponding matrix a block–Toeplitz matrix. An \( mn \)-by-\( mn \) matrix \( T_{n,m} \) is called a block–Toeplitz–Toeplitz–block (BTTB) matrix if it is an \( n \)-by-\( n \) block–Toeplitz matrix with each block being an \( m \)-by-\( m \) Toeplitz matrix. Toeplitz matrices arise in a variety of applications in mathematics and engineering, see for example [3].

\[4\] The variance of the entries of the vector is small.
instance Bunch [4], Chan and Ng [5], Kailath and Sayed [13] and the references therein.

We also discuss wavelet transforms for circulant matrices. An $n$-by-$n$ matrix $C_n$ is said to be circulant if

$$C_n = \begin{bmatrix}
c_0 & c_{-1} & \cdots & c_{2-n} & c_{1-n} \\
c_1 & c_0 & \cdots & c_{1-n} & c_{2-n} \\
\vdots & c_1 & c_0 & \cdots & \vdots \\
c_{n-2} & \cdots & \cdots & c_1 & c_0 \\
c_{n-1} & c_{n-2} & \cdots & c_1 & c_0
\end{bmatrix},$$

where $c_{-k} = c_{n-k}$ for $1 \leq k \leq n - 1$. Circulant matrices can be diagonalized by the Fourier matrix $F_n$, i.e.,

$$C_n = F_n^* A_n F_n. \quad (2)$$

Here $A_n$ is a diagonal matrix holding the eigenvalues of $C_n$, see for instance Davis [8]. We remark that circulant approximations to Toeplitz matrices have been considered and used for engineering applications [5,14]. In these applications, the circulant approximations thus obtained were used to replace the given Toeplitz matrices in subsequent computations. The Fourier transformed Toeplitz matrix $F_n T_n F_n^*$ is dense in general. However, because of the Toeplitz structure, Huckle [12] has shown that the entries of $F_n T_n F_n^*$ can be computed in $O(n \log n)$ operations.

The main contribution of this paper is twofolds: we first give the Toeplitz-like structure of the wavelet transformed Toeplitz matrices $W_n T_n W_n^T$, we then show that the computational cost for such structure is $O(k^3 l n)$, where $n$ is the size of the Toeplitz matrix, $k$ is the order of the wavelet and $l$ is the level used in the wavelet transform.

Such sparse and structured properties of the wavelet transformed Toeplitz matrices would be useful in some applications. For instance, the inverse of a Toeplitz matrix is required. One can apply the Newton method [17] to compute the inverse of a Toeplitz matrix. The direct application of the Newton method to $T_n$, the iterates are not Toeplitz matrices and therefore the storage would increase significantly. However, the application of the Newton method to a sparse matrix $W_n T_n W_n^T$, the iterates are still sparse in the computation process [1]. Because of the Toeplitz-like structure of $W_n T_n W_n^T$, there is no need to compute all the entries of the transformed matrices or to estimate the magnitude of these entries in advance for the compression. We remark that in many applications, e.g., in the compression of the discretization matrix of linear integral equations, we must “predict” the sparsity pattern of the wavelet transformed matrix and therefore we can avoid computing small entries which are not in the sparsity pattern, see for instance [1]. Another application is the preconditioning. Since the wavelet transformed Toeplitz matrices is much more sparser than the Fourier transformed Toeplitz matrices (see Section 3), we expect that the performance of the wavelet preconditioner is better than circulant preconditioners.
The rest of the paper is organized as follows. In Section 2, we discuss wavelet transforms for Toeplitz matrices. The structure of the wavelet transformed Toeplitz matrix is given in detail. In Section 3, a comparison between the wavelet transformed Toeplitz matrices and the Fourier transformed Toeplitz matrices is presented. In Section 4, we extend our results to BTTB matrices. Finally concluding remarks are given in Section 5.

2. Wavelet transforms for Toeplitz matrices

In this section, we consider the use of DWTs to Toeplitz matrices $T_n$ defined in (1), where the matrix size $n$ is a power of 2. We will consider orthogonal wavelets. Let the filter coefficients be given by $h_0, h_1, \ldots, h_{k-1}$, where $k$ is the order of the orthogonal wavelet. That is, the wavelet $\psi(t)$ has vanishing moments of the $(k/2)$th order

\[
\int_{-\infty}^{\infty} t^j \psi(t) \, dt = 0, \quad j = 0, 1, \ldots, \frac{k}{2} - 1,
\]

see [7]. Then the level-1 wavelet matrix is an $n$-by-$n$ orthogonal matrix given as follows:

\[
W_n^{(1)} = \begin{pmatrix}
  h_0 & h_1 & \ldots & h_{k-2} & h_{k-1} \\
  h_0 & h_1 & \ldots & h_{k-2} & h_{k-1} \\
  \vdots & \vdots & \ddots & \vdots & \vdots \\
  g_0 & g_1 & \ldots & g_{k-2} & g_{k-1} \\
  g_0 & g_1 & \ldots & g_{k-2} & g_{k-1} \\
  \vdots & \vdots & \ddots & \vdots & \vdots \\
  g_0 & g_1 & \ldots & g_{k-2} & g_{k-1} \\
  \vdots & \vdots & \ddots & \vdots & \vdots \\
  \vdots & \vdots & \ddots & \vdots & \vdots \\
  \vdots & \vdots & \ddots & \vdots & \vdots \\
  \vdots & \vdots & \ddots & \vdots & \vdots \\
\end{pmatrix}
\]

where $g_i = (-1)^i h_{k-1-i}$, $i = 0, 1, \ldots, k - 1$. The coefficients $g_i$, $i = 0, 1, \ldots, k - 1$ can be obtained by using $k/2$ vanishing moment conditions and $k/2$ orthonormal conditions. For instance, for Daubechies’s wavelets, the coefficients satisfy:

\[
\begin{align*}
\sum_{j=0}^{k-1} j^p g_j &= 0, \quad p = 0, 1, \ldots, \frac{k}{2} - 1 \text{ (vanishing moments)}, \\
\sum_{j=2p}^{k-1} g_j g_{j-2p} &= \delta_{0,p}, \quad p = 0, \ldots, \frac{k}{2} - 1 \text{ (orthonormal)}. 
\end{align*}
\]
In many applications, \( k = 4 \) is a good choice. In general, the level-\( j \) wavelet matrix is given by

\[
W_{n}^{(j)} = \begin{pmatrix}
W_{n/2}^{(1)} & \cdots & W_{n}^{(1)} \\
I_{n/2} & \cdots & I_{n} 
\end{pmatrix}.
\] (4)

The wavelet matrix is sparse since there are at most \( k \) nonzero entries in each row.

The DWT consists of applying a wavelet matrix like (3) hierarchically. Firstly, the level-\( -1 \) wavelet matrix applies to a full data vector, then the level-\( -2 \) wavelet matrix applies to the resulting vector again. By noting (4), the level-\( -2 \) wavelet matrix only operates on the first half of the entries of the vector. The process continues until the coarse level reaches. The procedure is called a pyramidal algorithm [7]. Let \( l \) be the number of levels used in the DWT, a DWT to \( T_{n} \) will be equal to

\[
T_{n}^{(l)} = \tilde{W}_{n}^{(l)} T_{n} \tilde{W}_{n}^{(l)} T,
\]

where

\[
\tilde{W}_{n}^{(l)} = W_{n}^{(l)} W_{n}^{(l-1)} \cdots W_{n}^{(1)}.
\] (5)

We will show that the cost of the DWT to \( T_{n} \) is \( O(lk^{3}n) \) and the wavelet trans-
formed matrix \( T_{n}^{(l)} \) has a Toeplitz-like structure (cf. Theorem 7). Let us first introduce the structured matrices which will be used in the representation of the Toeplitz-like structure of \( T_{n}^{(l)} \).

**Definition 1** (\( p \)-edge). An \( m \)-by-\( n \) matrix \( A_{m \times n} \) is called a \( p \)-edge matrix if it is of the form

\[
A_{m \times n} = \begin{bmatrix}
O_{(m-p) \times (n-p)} & * \\
* & * 
\end{bmatrix},
\]

where \( O_{(m-p) \times (n-p)} \) is the \( (m-p) \times (n-p) \) matrix with all zero entries.

**Definition 2** (\( s \)-Toeplitz and \( s \)-circulant). Let \( s \) be a positive integer. An \( m \)-by-\( sm \) matrix \( A_{m \times sm} \) is called an \( s \)-Toeplitz matrix if

\[
a_{i+1,j+s} = a_{i,j}, \quad 1 \leq i \leq m-1, \quad 1 \leq j \leq s(m-1).
\]

In other words, the submatrices \( A_{m \times sm}(i : j : s : (m-1)s + j) \) \((j = 1, 2, \ldots, s)\) are Toeplitz matrices. Here we refer to MATLAB notation. In particular, 1-Toeplitz matrices are Toeplitz matrices. Similarly, an \( m \)-by-\( sm \) matrix \( A_{m \times sm} \) is called an \( s \)-circulant matrix if the submatrices \( A_{m \times sm}(i : j : s : (m-1)s + j) \) \((j = 1, 2, \ldots, s)\) are circulant matrices.

It is obvious that the number of nonzero entries in an \( m \)-by-\( np \)-edge matrix is equal to or less than

\[
mp + np - p^{2} < (m + n)p.
\] (6)
An s-Toeplitz matrix is determined by its first row and its first s columns, i.e., an m-by-sm s-Toeplitz matrix is determined by
\[ sm + sm - s < 2sm \] (7) entries. Similarly, an m-by-sm s-circulant matrix is determined by its first row, i.e., sm entries. It is interesting to note that the level-1 wavelet matrix consists of two 2-circulant matrices. Let the wavelet matrix defined in (3) be partitioned as
\[ W_n^{(1)} = \begin{bmatrix} W_{11} \\ W_{21} \end{bmatrix}, \]
where \( W_{11} \) and \( W_{21} \) are \( n/2 \times n \) matrices. It is easy to see that both \( W_{11} \) and \( W_{21} \) are 2-circulant matrices.

The following lemmas and corollaries give the structures of the wavelet transformed s-Toeplitz matrices, s-circulant matrices and p-edge matrices.

**Lemma 3.** Let \( T_n \) be an n-by-n Toeplitz matrix, k be the order of the orthogonal wavelet, and \( W_n^{(1)} \) be the level-1 wavelet matrix. Then
\[ W_n^{(1)} T_n [W_n^{(1)}]^T = \begin{bmatrix} T_{11}^{(1)} & T_{12}^{(1)} \\ T_{21}^{(1)} & T_{22}^{(1)} \end{bmatrix}, \]
where the n/2-by-n/2 matrices \( T_{ij}^{(1)} \), \( i, j = 1, 2 \) are of the form
\[ T_{ij}^{(1)} = \tilde{T}_{ij}^{(1)} + E_{ij}^{(1)}, \]
\[ i, j = 1, 2. \]
Here \( \tilde{T}_{ij}^{(1)} \) are Toeplitz matrices and \( E_{ij}^{(1)} \) are p-edge matrices with \( p = (k - 2)/2 \).

**Proof.** Denote \( T_n \) by \( [t_{i-j}]_{i,j=1}^n \). Define an \((n + k - 2) \times (n + k - 2)\) matrix \( \tilde{T}_{n+k-2} \) by
\[ \tilde{t}_{i,j} = t_{i,j}, \quad 1 \leq i, j \leq n, \]
\[ \tilde{t}_{i,n+j} = t_{i,j}, \quad 1 \leq i \leq n, 1 \leq j \leq k - 2, \]
\[ \tilde{t}_{n+i,j} = t_{i,j}, \quad 1 \leq i \leq k - 2, 1 \leq j \leq n, \]
\[ \tilde{t}_{n+i,n+j} = t_{i,j}, \quad 1 \leq i, j \leq k - 2. \]
By direct computation, we have for \( 1 \leq i, j \leq n/2 \)
\[ [T_{11}^{(1)}]_{i,j} = \sum_{i',j'=0}^{k-1} h_i h_j^* \tilde{t}_{2i-1+i',2j-1+j'}, \]
\[ [T_{12}^{(1)}]_{i,j} = \sum_{i',j'=0}^{k-1} h_i g_j^* \tilde{t}_{2i-1+i',2j-1+j'}. \]
\[ [T_{21}^{(1)}]_{i,j} = \sum_{i',j'=0}^{k-1} g_{i'} h_{j'} \tilde{t}_{2i-1+i',2j-1+j'}, \]

\[ [T_{22}^{(1)}]_{i,j} = \sum_{i',j'=0}^{k-1} g_{i'} h_{j'} \tilde{t}_{2i-1+i',2j-1+j'}. \]

Note that if \( 1 \leq i, j \leq (n - k + 2)/2 \) and \( 0 \leq i', j' \leq k - 1 \), then \( 2i - 1 + i' \leq n \) and \( 2j - 1 + j' \leq n \). Therefore, for \( 1 \leq i, j \leq (n - k + 2)/2 \),

\[ [T_{11}^{(1)}]_{i,j} = \sum_{i',j'=0}^{k-1} h_{i'} h_{j'} \tilde{t}_{2i-1+i',2j-1+j'} = \sum_{i',j'=0}^{k-1} h_{i'} h_{j'} t_{2i-1+i',2j-1+j'} \]

\[ = \sum_{i',j'=0}^{k-1} h_{i'} h_{j'} t_{2(i-j)+(i'-j')}. \]

That is, the \((n - k + 2)/2\) leading principal submatrix of \(T_{11}^{(1)}\) is a Toeplitz matrix. Thus, we can write \(T_{11}^{(1)} = \tilde{T}_{11}^{(1)} + E_{11}^{(1)}\), where \(\tilde{T}_{11}^{(1)}\) is a Toeplitz matrix and \(E_{11}^{(1)}\) is a \(p\)-edge matrix with \(p = (k - 2)/2\). Obviously, the matrices \(T_{12}^{(1)}, T_{21}^{(1)}, \) and \(T_{22}^{(1)}\) are of the same form. \(\square\)

**Corollary 4.** Let \(C_n\) be an \(n\)-by-\(n\) circulant matrix, \(k\) be the order of the orthogonal wavelet, and \(W_n^{(1)}\) be the level-1 wavelet matrix. Then

\[ W_n^{(1)} C_n [W_n^{(1)}]^T = \begin{bmatrix} C^{(1)}_{11} & C^{(1)}_{12} \\ C^{(1)}_{21} & C^{(1)}_{22} \end{bmatrix}, \]

where the \(n/2\)-by-\(n/2\) matrices \(C^{(1)}_{ij}\), \(i, j = 1, 2\) are circulant matrices.

**Proof.** Let \(C_n = [c_{i-j}]_{i,j=1}^n\), where \(c_i = c_{i+n}\) for \(i = -(n-1), \ldots, -1\). For the matrix \(C_n\), we define an \((n+k-2)\times(n+k-2)\) matrix \(\tilde{C}_{n+k-2}\) by

\[ \tilde{c}_{i,j} = c_{i,j}, \quad 1 \leq i, j \leq n, \]

\[ \tilde{c}_{i,n+j} = c_{i,j}, \quad 1 \leq i \leq n, 1 \leq j \leq k-2, \]

\[ \tilde{c}_{n+i,j} = c_{i,j}, \quad 1 \leq i \leq k-2, 1 \leq j \leq n, \]

\[ \tilde{c}_{n+i,n+j} = c_{i,j}, \quad 1 \leq i, j \leq k-2. \]

It is easy to check that \(\tilde{C}_{n+k-2}\) is a Toeplitz matrix. Let \(\tilde{C}_{n+k-2} = [\tilde{c}_{i,j}]_{i,j=1}^{n+k-2}\), it follows that \(\tilde{c}_{j} = \tilde{c}_{j+n}\) for \(-(n + k - 3) \leq j \leq k - 3\). Thus, for \(1 \leq i, j \leq n/2\)

\[ [C_{11}^{(1)}]_{i,j} = \sum_{i',j'=0}^{k-1} h_{i'} h_{j'} \tilde{c}_{2i-1+i',2j-1+j'} = \sum_{i',j'=0}^{k-1} h_{i'} h_{j'} \tilde{c}_{2(i-j)+(i'-j')}, \]
it follows that $C_{11}^{(1)}$ is a Toeplitz matrix. We will prove that it is a circulant matrix, i.e., let $C_{11}^{(1)} = [c_{i-j}^{(1)}]_{i,j=1}^{n/2}$, we prove that $c_{j}^{(1)} = c_{j+n/2}^{(1)}$ for $j = -(n/2 - 1), \ldots, -1$. Note that for $-(n/2 - 1) \leq j \leq -1$ and $0 \leq i', j' \leq k - 1$, we have $- (n - k + 3) \leq 2j + (i' - j') \leq k - 3$. By using $\tilde{c}_j = \tilde{c}_{j+n}$ we get that for $j = -(n/2 - 1), \ldots, -1$, 

$$
c_{j}^{(1)} = \sum_{i',j'=0}^{k-1} h_{i'} h_{j'} \tilde{c}_{2j+(i'-j')} = \sum_{i',j'=0}^{k-1} h_{i'} h_{j'} \tilde{c}_{2j+(i'-j')+n} 
$$

Thus, $C_{11}^{(1)}$ is a circulant matrix. Similarly, $C_{12}^{(1)}$, $C_{21}^{(1)}$, and $C_{22}^{(1)}$ are circulant matrices.

**Lemma 5.** Let $A_n$ be an $n$-by-$n$ p-edge matrix, $k$ be the order of the orthogonal wavelet, and $W_n^{(1)}$ be the level-1 wavelet matrix. Then

$$
W_n^{(1)}A_n[W_n^{(1)}]^{T} = \begin{bmatrix}
A_{11}^{(1)} & A_{12}^{(1)} \\
A_{21}^{(1)} & A_{22}^{(1)}
\end{bmatrix},
$$

where

$$
A_{ij}^{(1)} = \begin{bmatrix}
O_{n-q} & * \\
* & * 
\end{bmatrix}, \quad i, j = 1, 2
$$

and $q = \lfloor (k + p - 1)/2 \rfloor$. Here $\lfloor x \rfloor$ is the largest integer that does not greater than $x$. In particular, if $p \leq k - 2$, then $q \leq k - 2$.

**Proof.** Define the $(n + k - 2) \times (n + k - 2)$ matrix $\tilde{A}_{n+k-2}$ by

$$
\tilde{a}_{i,j} = a_{i,j}, \quad 1 \leq i, j \leq n, \\
\tilde{a}_{i,n+j} = a_{i,j}, \quad 1 \leq i \leq n, \quad 1 \leq j \leq k - 2, \\
\tilde{a}_{n+i,j} = a_{i,j}, \quad 1 \leq i \leq k - 2, \quad 1 \leq j \leq n, \\
\tilde{a}_{n+i,n+j} = a_{i,j}, \quad 1 \leq i, j \leq k - 2.
$$

We have

$$
[A_{11}^{(1)}]_{i,j} = \sum_{i',j'=0}^{k-1} h_{i'} h_{j'} \tilde{a}_{2i-1+i',2j-1+j'}, \quad 1 \leq i, j \leq n/2.
$$

Using $q = \lfloor (k + p - 1)/2 \rfloor$ we get $2q \geq k + p - 2$. It follows that for $1 \leq i, j \leq n/2 - q$ and $0 \leq i', j' \leq k - 1$, we have $2i - 1 + i' \leq 2(n/2 - q) - 1 + (k - 1) \leq n - (k + p - 2) - 1 + (k - 1) = n - p$ and $2j - 1 + j' \leq n - p$. Note that $a_{ij} = 0$ for $1 \leq i, j \leq n - p$, we have
Lemma 6. Let $q$ be the order of the orthogonal wavelet and $W_n^{(1)}$ be the level-1 wavelet matrix. Let $A_{m \times (sm)}$ be the sum of an $s$-Toeplitz matrix and a $p$-edge matrix with $p \leq 2$. Then

$$W_m^{(1)} A_{m \times (sm)} = \begin{bmatrix} T_{11} \\ T_{21} \end{bmatrix},$$

where the $(m/2) \times (sm)$ matrices $T_{11}$ and $T_{21}$ are sums of $2s$-Toeplitz matrices and $q$-edge matrices with $q = \lfloor (k + p - 1)/2 \rfloor \leq k - 2$. In particular, if $A_{m \times (sm)}$ is an $s$-circulant matrix, then $T_{11}$ and $T_{21}$ are $2s$-circulant matrices.

Proof. Let $\tilde{A}_{(m+k-2) \times (sm)}$ be an $(m+k-2) \times (sm)$ matrix defined as

$$\tilde{a}_{i,j} = a_{i,j}, \quad 1 \leq i \leq m, \quad j \leq sm,$$

$$\tilde{a}_{m+i,j} = a_{i,j}, \quad 1 \leq i \leq k-2, \quad 1 \leq j \leq sm.$$

It is easy to see that for $1 \leq i \leq m/2$, $1 \leq j \leq sm$,

$$[T_{11}]_{ij} = \sum_{i'=0}^{k-1} h_i \tilde{a}_{2i-1+i',j}, \quad [T_{21}]_{ij} = \sum_{i'=0}^{k-1} g_i \tilde{a}_{2i-1+i',j}.$$

We note that

$$a_{i+1,j+s} = a_{i,j}, \quad 1 \leq i \leq m - p - 1, \quad 1 \leq j \leq s(m - 1) - p$$

and

$$[T_{11}]_{ij} = \sum_{i'=0}^{k-1} h_i a_{2(i+1)-1+i',j}, \quad 1 \leq i \leq m/2 - q, \quad 1 \leq j \leq sm - p.$$

Therefore for $i = 1, \ldots, m/2 - q - 1$ and $j = 1, \ldots, (s-2)m - p$, we have

$$[T_{11}]_{i+1,j+2s} = \sum_{i'=0}^{k-1} h_i a_{2(i+1)-1+i',j+2s} = \sum_{i'=0}^{k-1} h_i a_{2(i+1)-1+i',j+2s}$$

$$= \sum_{i'=0}^{k-1} h_i a_{2i-1+i',j} = [T_{11}]_{i,j}.$$

Thus, $T_{11}$ is the sum of a $2s$-Toeplitz matrix and a $q$-edge matrix. Similarly, $T_{21}$ is the sum of a $2s$-Toeplitz matrix and a $q$-edge matrix. If $A_{m \times (sm)}$ is an $s$-circulant matrix, we can prove that $T_{11}$ and $T_{21}$ are $2s$-circulant matrices by proving that $T_{11}$ and $T_{21}$ are $2s$-Toeplitz matrices first. □
Based on the results in Lemmas 3, 5 and 6, and Corollary 4 we show that the wavelet transformed Toeplitz matrices and circulant matrices have Toeplitz-like structure and circulant-like structure respectively.

**Theorem 7.** Let $T_n$ be an $n$-by-$n$ Toeplitz matrix, $k$ be the order of the orthogonal wavelet, and $W_{n}^{(j)}$ be the level-$j$ wavelet matrix. Let

$$T_n^{(l)} = (W_n^{d}) \cdots W_n^{(1)} T_n \left( [W_n^{(1)}]^T \cdots [W_n^{(l)}]^T \right).$$

Let the matrix $T_n^{(l)}$ be partitioned as

$$T_n^{(l)} = [T_{ij}^{(l)}]_{i,j=1}^{l+1},$$

where $T_{ij}^{(l)}$ are $n_i \times n_j$ matrices. Here $n_1 = n_2 = n/2^l$, $n_i = n/2^{l-i+2}$ for $i = 3, \ldots, l+1$. Then,

$$T_{ij}^{(l)} = \hat{T}_{ij}^{(l)} + E_{ij}^{(l)}, \quad i, j = 1, 2, \ldots, l+1,$$

where $E_{ij}^{(l)}$ is a $p$-edge matrix with $p \leq k-2$, $\hat{T}_{ij}^{(l)}$ and $[\hat{T}_{ji}^{(l)}]^T$ are $(n_j/n_i)$-Toeplitz matrices for $i \leq j$.

**Proof.** Using the results of Lemmas 3, 5 and 6, one can prove the theorem by induction.

As an example, we show the structure of the wavelet transformed matrix $T_n^{(3)}$ in Fig. 1.

<table>
<thead>
<tr>
<th></th>
<th>1 + $E$</th>
<th>2 + $E$</th>
<th>4 + $E$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 + $E$</td>
<td>2 + $E$</td>
<td>4 + $E$</td>
<td></td>
</tr>
<tr>
<td>1 + $E$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1 + $E$</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1 + $E$</td>
<td></td>
</tr>
</tbody>
</table>

**Fig. 1.** Structure of the wavelet transformed Toeplitz matrix: level = 3. In the figure, the symbol $s + E$ denotes that the matrix in the area is the sum of an $s$-Toeplitz matrix and a edge matrix.
Corollary 8. Let $C_n$ be an $n$-by-$n$ circulant matrix. Let the sequence of matrices $C_n^{(j)}$, $j = 1, 2, \ldots, l$ be defined by $C_n^{(j)} = W^{(j)}C_n^{(j-1)}[W^{(j)}]^T$, where $C_n^{(0)} = C_n$ and $W^{(j)}$ is the level-$j$ wavelet matrix. Let the matrix $C_n^{(l)}$ be partitioned as $C_n^{(l)} = [C_{ij}^{(l)}]_{i,j=1}^{n}$, where $C_{ij}^{(l)}$ are $n_i \times n_j$ matrices. Here $n_1 = n_2 = n/2^l$, $n_i = n/2^{l+i} - 1$ for $i = 3, \ldots, l + 1$. Then $C_{ij}^{(l)}$ and $[C_{ij}^{(l)}]^T$ are $(n_j/n_i)$-circulant matrices for $i \leq j$.

2.1. The computational cost and the storage requirement

In this subsection, we discuss the computational cost and the memory requirement of a DWT applied to Toeplitz matrices. We will show that by exploiting the Toeplitz-like structure of the wavelet transformed matrix $T_n^{(l)}$, we can compute all different entries of $T_n^{(l)}$ in $O(k \ln n)$ operations and the storage requirement is less than $2k \ln n$.

Let us first discuss the computational cost in each level of the DWT step. In each level of the wavelet transform, each entry requires $k^2$ multiplications and $k^2 - 1$ additions to compute. Therefore, one requires only to know how many entries should be computed. To compute $T_n^{(l)} = W_n^{(l)}T_n^{(l-1)}[W_n^{(l)}]^T$, one needs to compute $T_{ij}^{(1)}$ and $E_{ij}^{(1)}$, $i, j = 1, 2$, where $T_{ij}^{(1)}$ and $E_{ij}^{(1)}$ are $n/2$-by-$n/2$ Toeplitz matrices and $p$-edge matrices ($p \leq k - 2$) respectively. Using (6) and (7), we see that the number of different entries in $T_n^{(1)}$ is less than

$$4(2 \cdot n/2 + (n/2 + n/2) \cdot (k - 2)) = 4(k - 1)n.$$

To compute $T_n^{(2)} = W_n^{(2)}T_n^{(1)}[W_n^{(2)}]^T$, one needs to compute $T_{ij}^{(2)}$, $i, j = 1, 2$ and $T_{13}^{(2)} \cdot T_{23}^{(2)} \cdot T_{31}^{(2)}$ and $T_{32}^{(2)}$. From the discussion on the computational cost in level-1 wavelet transform, we see that the number of different entries of the submatrices $T_{ij}^{(2)}$, $(i, j = 1, 2)$ is less than

$$4(k - 1)(n/2) = 2(k - 1)n.$$

Note that the $(n/4)$-by-$(n/4)$ matrices $T_{13}^{(2)} \cdot T_{23}^{(2)} \cdot [T_{31}^{(2)}]^T$ and $[T_{32}^{(2)}]^T$ are sums of $2$-Toeplitz matrices and $p$-edge matrices with $p \leq k - 2$, it follows that the number of different entries in each of the above matrices is less than

$$2 \cdot n/2 + (n/2 + n/4) \cdot (k - 2) = n + 3n(k - 2)/4.$$

Therefore, the number of entries one requires to compute in level-2 wavelet transform is less than

$$2(k - 1)n + 4 \cdot (n + 3n(k - 2)/4) = (5k - 4)n.$$

In general, to compute $T_n^{(l)} = W_n^{(l)}T_n^{(l-1)}[W_n^{(l)}]^T$, we need to compute the submatrices $T_{ij}^{(l)}$ and $T_{ji}^{(l)}$ for $i = 1, 2, j = 1, \ldots, l + 1$. It is not difficult to check that the number of entries to compute is less than
Based on the above analysis, we see that to obtain $T^{(l)}_{n}$ from $T_{n}$, the number of all entries one needs to compute in the full DWT step is less than

$$4(k-1)n + \sum_{j=2}^{l} \left[ 4k + \frac{(k-2)j-k}{2^{j-2}} \right] n.$$

Next we discuss the storage requirement for the wavelet transformed matrix $T^{(l)}_{n}$. Note that $T^{(l)}_{ij}$ or $[T^{(l)}]^{T}$ is the sum of an $(n_{j}/n_{i})$-Toeplitz matrix and a $p$-edge matrix with $p \leq k-2$ (cf. Theorem 7). Therefore the number of different entries in $T^{(l)}_{ij}$ is less than

$$2 \max(n_{i}, n_{j}) + (k-2)(n_{i} + n_{j}).$$

It follows that the number of entries to be stored for $T^{(l)}_{n}$ is less than

$$\sum_{i,j=1}^{l+1} \left[ 2 \max(n_{i}, n_{j}) + (k-2)(n_{i} + n_{j}) \right]$$

$$= 2(2l - 1)n + n2^{2-l} + 2(k-2)(l+1)n$$

$$\leq 4n + 2(k-2)(l+1)n = 2(kl + k - 2)n < 2k(l + 1)n.$$  \hfill (9)

**Remark.** For circulant matrices, by using the circulant-like structure of the wavelet transformed circulant matrix, it can be proved that in order to obtain $C^{(l)}_{n}$ for $C_{n}$, we require to compute

$$\sum_{j=1}^{l} 4n(1 - 2^{-j}) = 4nl - 4n(1 - 2^{-l}) \leq 4ln$$

different entries. According to Corollary 8, to store $C^{(l)}_{n}$, we require to store

$$\sum_{i,j=1}^{l+1} \max(n_{i}, n_{j}) = (2l - 1)n + n2^{1-l} \leq 2ln$$

different entries.

By (8) we see that in order to compute $T^{(l)}_{n}$ from $T_{n}$, one needs to compute about $4kln$ entries. Since each entry needs $O(k^2)$ operations, the cost of a full DWT step for Toeplitz matrices is of order $O(lk^3n)$. From (9) we have derived that the storage requirement for the final wavelet transformed matrix $T^{(l)}_{n}$ is $2k(l + 1)n$. For a circu-
Fig. 2. Sparsity patterns of $T^{(l)}_{n,\epsilon}$ (left) and $C^{(l)}_{n,\epsilon}$ (right). Here $T_n = [t_{i,j}]_{i,j=1}^n$ with $n = 256$, $t_j = 1/(j+1)$ for $j = 1, \ldots, n-1$ and $C_n$ is the optimal circulant preconditioner for $T_n$ [6]. The threshold and the level are set to $\epsilon = 10^{-4}$ and $l = 4$ respectively.

lant matrix, the cost of a full DWT step and the storage requirement for the wavelet
transformed matrix are about $1/k$ of those for a Toeplitz matrix.

The most interesting and promising wavelet applications is to compress the matrix
data. After throwing off entries of $T^{(l)}_{n,\epsilon}$ and $C^{(l)}_{n,\epsilon}$ which are below some threshold $\epsilon$ we get sparse matrices $T^{(l)}_{n,\epsilon}$ and $C^{(l)}_{n,\epsilon}$. Examples of the sparse patterns of $T^{(l)}_{n,\epsilon}$ and $C^{(l)}_{n,\epsilon}$ are illustrated in Fig. 2. Here we use the Daubechies wavelet of order 4 in the examples.

3. Fourier transforms for Toeplitz matrices

Let $A_n$ be an $n$-by-$n$ matrix and $F_n$ be the Fourier matrix of order $n$, we call the matrix

$$B_n = F_n A_n F_n^*,$$

the Fourier transformed matrix of $A_n$. In general, the computation of $B_n = [F_n(F_n A_n F_n^*)]^*$ requires $O(n^2 \log n)$ operations by applying FFT to the matrices $A_n$ and $F_n A_n$. If $A_n = T_n$ is a Toeplitz matrix, however, the matrix $B_n$ can be obtained
in $O(n \log n)$ operations by using FFT [12]. In fact, for $i \neq j$,

$$b_{i,j} = \frac{1}{n(1 - \exp(2\pi(j-i)\sqrt{-1}/n))} (g_i - g_j + \bar{g}_j - \bar{g}_i),$$

where

$$g_i = \sum_{r=1}^{n-1} \tilde{t}_r \exp(-2\pi(i-1)\sqrt{-1}/n), \quad i = 1, 2, \ldots, n,$$

which can be computed by FFT applied to $(0, \tilde{t}_1, \ldots, \tilde{t}_{n-1})^T$ in $O(n \log n)$ operations. Moreover, the diagonals $b_{i,i}, i = 1, \ldots, n$ also can be computed by FFT in $O(n \log n)$ operations since they are the eigenvalues of the optimal circulant
preconditioner of $T_n$ [6]. We note that the complexity of the application of the matrix $B_n$ to a vector is also $O(n \log n)$. For more details, we refer readers to [12].

3.1. Comparison

For a large class of matrices, the wavelet transformed matrices are sparse after throwing off small entries. On the other hand, the Fourier transformed matrix is not sparse in general. Here we give two examples: symmetric Toeplitz matrices with first columns given by

$$t_j = 1/(1 + j) \text{ and } t_j = \exp(-0.5j^2),$$

$j = 0, 1, \ldots, n - 1$, respectively. The first sequence decays slowly and the second sequence decays quickly. In order to compare the sparse patterns of the wavelet

![Fig. 3. Absolute values of wavelet transformed matrix $T_{256}^{(4)}$ (left) and Fourier transformed matrix $B_{256}$ (right) of $T_{256}$. Here $T_{256} = [1/(1 + |i - j|)]_{i,j=1}^{256}$.](image1)

![Fig. 4. Absolute values of wavelet transformed matrix $T_{256}^{(4)}$ (left) and Fourier transformed matrix $B_{256}$ (right) of $T_{256}$. Here $T_{256} = [\exp(-0.5(j - i)^2)]_{i,j=1}^{256}$.](image2)
transformed Toeplitz matrices $T_n^{(l)}$ and the Fourier transformed Toeplitz matrices $B_n$, we first rearrange the absolute values of all entries of each transformed matrix as a vector in increasing manner and then plot the resulted vectors. From Figs. 3 and 4 it is clearly seen that the wavelet transformed matrix $T_{256}^{(4)}$ has much more small entries than the Fourier transformed matrix $B_{256}$ has. In other words, by throwing off entries of $T_n^{(l)}$ with absolute values under certain threshold, say $10^{-4}$, we get a very spare matrix $T_n^{(l),\epsilon}$ while $B_n,\epsilon$ is still a dense matrix, where $B_n,\epsilon$ is the matrix obtaining from $B_n$ by throwing off small entries. Here we use the Daubechies wavelet of order 4 in the examples.

4. Wavelet transforms for BTTB matrices

Let $T_{m,n}$ be an $m$-by-$m$ block Toeplitz matrix with $n$-by-$n$ Toeplitz block (BTTB). We discuss the computation of the wavelet transform for $T_{m,n}$:

$$T_{m,n}^{(l_1,l_2)} = (\tilde{W}_m^{(l_1)} \otimes \tilde{W}_n^{(l_2)}) T_{m,n} (\tilde{W}_m^{(l_1)} \otimes [\tilde{W}_n^{(l_2)}]^T)$$

$$= (\tilde{W}_m^{(l_1)} \otimes I_n) (I_m \otimes \tilde{W}_n^{(l_2)}) T_{m,n} (I_m \otimes [\tilde{W}_n^{(l_2)}]^T) ([\tilde{W}_m^{(l_1)}]^T \otimes I_n),$$

where $l_1$ and $l_2$ are the levels of wavelet transforms. The transformation can be divided into two steps as follows.

Step 1. Compute $T_{m,n}^{(0,l_2)} = (I_m \otimes \tilde{W}_n^{(l_2)}) T_{m,n} (I_m \otimes [\tilde{W}_n^{(l_2)}]^T)$.

Step 2. Compute $T_{m,n}^{(l_1,0)} = (\tilde{W}_m^{(l_1)} \otimes I_n) T_{m,n}^{(0,l_2)} ([\tilde{W}_m^{(l_1)}]^T \otimes I_n)$.

Step 1 is equivalent to applying the wavelet transform to the Toeplitz blocks of the first column and the first row of the BTTB matrix. Obviously, $T_{m,n}^{(0,l_2)}$ is a block Toeplitz matrix. It is known that there exists a permutation matrix $P$ such that $P^T(\tilde{W}_m^{(l_1)} \otimes I_n) P = I_n \otimes \tilde{W}_m^{(l_1)}$ and $P^T T_{m,n}^{(0,l_2)} P$ is an $n \times n$ block matrix with $m \times m$ Toeplitz blocks. Therefore, the wavelet transformation of Step 2:

$$= P \cdot P^T(\tilde{W}_m^{(l_1)} \otimes I_n) P \cdot P^T T_{m,n}^{(0,l_2)} P \cdot (I_n \otimes [\tilde{W}_m^{(l_1)}]^T) \cdot P^T$$

$$= P \cdot (I_n \otimes \tilde{W}_m^{(l_1)}) \cdot P^T T_{m,n}^{(0,l_2)} P \cdot (I_n \otimes [\tilde{W}_m^{(l_1)}]^T) \cdot P^T$$

(11)

can be performed efficiently by using Toeplitz structure.

According to the discussions in the previous section, we see that Step 1 requires to compute about $(2m - 1) \cdot (4nk^{l_2}) \approx 8mnkl^2$ entries. Using (8), (9), and (11), we see that Step 2 requires to compute about

$$[2nk(l_2 + 1)] \cdot (4nkl_1) = 8mnk^2l_1(l_2 + 1)$$
entires. Thus, the full DWT step to a BTTB matrix requires the computation of about \(8mnk[l_1(l_2 + 1) + l_2]\) entries. For a block–circulant–circulant–block (BCCB) matrix \(C_{m,n}\), Steps 1 and 2 requires to compute about \(4mnl_2\) and \(8mnl_1l_2\) entries respectively. Therefore the full DWT step for a BCCB matrix requires the computation of about \(4mnl_2(2l_1 + 1)\) entries.

The main cost is in Step 2 and the cost of the DWT to a BCCB matrix is about \(1/(2k^2)\) that for a BTTB matrix. After throwing off entries of \(T(l_1,l_2)\) and \(C(l_1,l_2)\) which are below the threshold \(\epsilon\) we get sparse matrices \(T(l_1,l_2,\epsilon)\) and \(C(l_1,l_2,\epsilon)\). To end this section, we note that to make the DWT less expensive, one may throw away entries with absolute value less than the given threshold in each level.

5. Concluding remarks

In this paper, we have showed that the wavelet transformed Toeplitz matrix has Toeplitz-like structure. We can compute the whole wavelet transformed matrix in \(O(nk^3l)\) operations, where \(n, k,\) and \(l\) are the matrix size, the order of wavelet transform, and the level of wavelet transform respectively. The wavelet transformed matrix can be used as a preconditioner. One can apply sparse solvers, see for instance [1], or incomplete factorization [2] to solve the preconditioned system. In our future research, we will study how to construct sparse preconditioners for Toeplitz-related systems based on wavelet transforms as proposed in [11]. Obviously, the convergence rate of the preconditioned conjugate gradient method depends on the threshold \(\epsilon\) chosen: the smaller the threshold, the faster the convergence. On the other hand, a small threshold will give rise to a dense wavelet transformed matrix. In general, the difference between \(T^{(l)}_{n,\epsilon}\) and \(T^{(l)}_{n,0}\) is a small norm matrix (depends on \(\epsilon\)) while the difference between the circulant approximation and the Toeplitz matrix is the sum of a small norm matrix and a finite rank matrix. That is, the approximation based on the wavelet transform can be better. Therefore, with the proper choice of threshold, we can derive more efficient wavelet transform based preconditioners than circulant preconditioners.

Acknowledgments

The authors would like to thank the anonymous referees and Prof. R. Plemmons for their helpful comments and suggestions in revising the paper.

References