2016

Signal Processing based on Stable radix-2 DCT I-IV Algorithms having Orthogonal Factors

Sirani K. M. Perera
Embry-Riddle Aeronautical University - Daytona Beach, pereras2@erau.edu

Follow this and additional works at: https://repository.uwyo.edu/ela

Part of the Numerical Analysis and Computation Commons, Signal Processing Commons, and the Theory and Algorithms Commons

Recommended Citation
DOI: https://doi.org/10.13001/1081-3810.3207

This Article is brought to you for free and open access by Wyoming Scholars Repository. It has been accepted for inclusion in Electronic Journal of Linear Algebra by an authorized editor of Wyoming Scholars Repository. For more information, please contact scholcom@uwyo.edu.
SIGNAL PROCESSING BASED ON STABLE RADIX-2 DCT I-IV ALGORITHMS HAVING ORTHOGONAL FACTORS

SIRANI M. PERERA

Abstract. This paper presents stable, radix-2, completely recursive discrete cosine transform algorithms DCT-I and DCT-III solely based on DCT-I, DCT-II, DCT-III, and DCT-IV having sparse and orthogonal factors. Error bounds for computing the completely recursive DCT-I, DCT-II, DCT-III, and DCT-IV algorithms having sparse and orthogonal factors are addressed. Signal flow graphs are demonstrated based on the completely recursive DCT-I, DCT-II, DCT-III, and DCT-IV algorithms having orthogonal factors. Finally image compression results are presented based on the recursive 2D DCT-II and DCT-IV algorithms for image size $512 \times 512$ pixels with transfer block sizes $8 \times 8$, $16 \times 16$, and $32 \times 32$ with 93.75% absence of coefficients in each transfer block.

Key words. Discrete cosine transform, Sparse and orthogonal factors, Radix-2 algorithms, Recursive algorithms, Stable algorithms, Arithmetic cost, Error bounds, Image reconstruction, Signal flow graphs.

AMS subject classifications. 15A23, 15B10, 65F35, 65F50, 65T50, 65Y05, 65Y20, 94A08, 94A12.

1. Introduction. The Fast Fourier Transform is used to efficiently compute the Discrete Fourier Transform (DFT) and its inverse. The DFTs are widely used in numerous applications in applied mathematics and electrical engineering [3, 19, 23, 24, 27, 30], etc.

The DFT uses complex arithmetic. The DFT of an $n$-input sequence $\{x_k\}_{k=0}^{n-1}$ is the $n$-output sequence $\{y_k\}_{k=0}^{n-1}$ defined via

$$\begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_{n-1} \end{bmatrix} = \frac{1}{\sqrt{n}} \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & \omega_n & \cdots & \omega_n^{n-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \omega_n^{n-1} & \cdots & \omega_n^{(n-1)(n-1)} \end{bmatrix} \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_{n-1} \end{bmatrix},$$

where $\omega_n = e^{-\frac{2\pi i}{n}}$. There exist real analogues of the DFT, namely the Discrete Cosine Transforms and Discrete Sine Transforms, the main types are from I to IV. Similar to (1.1), the I-IV variants of cosine and sine matrices transform an $n$-input sequence into the $n$-output sequence via the transform matrices stated in Table 1.1, where for DCT-1 $j, k = 0, 1, \ldots, n$, DST-I $j, k = 0, 1, \ldots, n-2$, DCT and DST II-IV $j, k = 0, 1, \ldots, n-1$, and $n \geq 2$. Among DCT I-IV transforms, $C_{n+1}$ was introduced in [31], $C_{n}$ and its inverse $C_{n}^{-1}$ were introduced in [1], and $C_{n}$ was introduced into digital signal processing in [9]. Moreover, among DST I-IV transforms, $S_{n-1}^I$
Signal Processing Based on Stable Radix-2 DCT I-IV Algorithms Having Orthogonal Factors

<table>
<thead>
<tr>
<th>Cosine and Sine Transforms</th>
<th>Inverse Transforms</th>
</tr>
</thead>
<tbody>
<tr>
<td>( C_{n+1}^I = \sqrt{\frac{2}{n}} \epsilon_n(j) \frac{1}{n} \sum_{k=0}^{n-1} \epsilon_n(k) \cos \frac{k \pi}{n} ) ( (j+1) )</td>
<td>( (j+1) )</td>
</tr>
<tr>
<td>( C_{n+1}^II = \sqrt{\frac{2}{n}} \epsilon_n(j) \cos \frac{(2j+1) \pi}{2n} ) ( (j+1) )</td>
<td>( (j+1) )</td>
</tr>
<tr>
<td>( C_{n+1}^III = \sqrt{\frac{2}{n}} \epsilon_n(k) \cos \frac{(2j+1) \pi}{2n} ) ( (j+1) )</td>
<td>( (j+1) )</td>
</tr>
<tr>
<td>( C_{n+1}^IV = \sqrt{\frac{2}{n}} \cos \frac{(2j+1) \pi}{4n} ) ( (j+1) )</td>
<td>( (j+1) )</td>
</tr>
<tr>
<td>( S_{n+1}^I = \frac{1}{\sqrt{n}} \sin \frac{(j+1)(k+1) \pi}{n} )</td>
<td>( (j+1)(k+1) )</td>
</tr>
<tr>
<td>( S_{n+1}^II = \frac{1}{\sqrt{n}} \epsilon_n(j) \sin \frac{(j+1)(k+1) \pi}{2n} )</td>
<td>( (j+1)(k+1) )</td>
</tr>
<tr>
<td>( S_{n+1}^III = \frac{1}{\sqrt{n}} \epsilon_n(k) \sin \frac{(j+1)(k+1) \pi}{2n} )</td>
<td>( (j+1)(k+1) )</td>
</tr>
<tr>
<td>( S_{n+1}^IV = \frac{1}{\sqrt{n}} \sin \frac{(j+1)(k+1) \pi}{4n} )</td>
<td>( (j+1)(k+1) )</td>
</tr>
</tbody>
</table>

Table 1.1: Cosine and sine transform matrices.

and \( S_{n}^{IV} \) were introduced in [9, 10] and \( S_{n}^{II} \) and its inverse \( S_{n}^{III} \) were introduced in [14]. These classifications were also stated in [19, 30].

It has been stated, in e.g. [21, 22, 24], that these cosine and sine matrices of types I-IV are orthogonal. Strang, in [24], proved that the column vectors of each cosine matrix are eigenvectors of a symmetric second difference matrix under different boundary conditions, and are hence orthogonal. Later Britanak, Yip, and Rao in [3] followed very closely the presentation made by Strang’s [24] to point out that the column vectors of each cosine and sine matrix of types I-VIII are eigenvectors of a symmetric second difference matrix. Due to these DCT and DST properties, it was shown by many authors (see e.g. [2, 3, 4, 6, 7, 11, 12, 13, 14, 15, 16, 17, 24, 28, 29]) that these symmetric and asymmetric (rarely used) versions of DCT and DST can be widely used in image processing, signal processing, fingerprint enhancement, quick response code (QR code), etc.

To obtain real, fast DCT or DST algorithms one can mainly use a polynomial arithmetic technique or a matrix factorization technique. In the polynomial arithmetic technique (see e.g. [25]), components of \( C_n \) or \( S_n \) are interpreted as the nodes of a degree \( n \) polynomial, and then one applies the divide and conquer technique to reduce the degree of the polynomial. The matrix factorization technique is the direct factorization of the DCT or DST matrices into the product of sparse matrices (see e.g. [3, 18, 19, 30, 32]). Later it was found (see e.g. [26]) that if the factorization for DCT or DST does not preserve orthogonality the resulting DCT or DST algorithms lead to inferior numerical stability. The matrix factorization for DST-I in [32] used the results in [5] to decompose DST-I into DCT and DST. Also the decomposition for DCT-II in [30] is a slightly different version of the result in [5]. Though one can find orthogonal matrix factorizations for DCT and DST in [30], the resulting algorithms in [30] are not completely recursive, and hence do not lead to simple recursive algorithms. Moreover, [3] has used the same factorization for DST-II and DST-IV as in [30]. On the other hand, one can use these [3, 24, 30] results to derive recursive, stable...
algorithms as stated in [18, 19].

However, [19] has offered stable, recursive DCT-II and DCT-IV algorithms, based on DCT-II and DCT-IV. Thus, this paper completes the picture and provides completely recursive, stable, radix-2 DCT-I and DCT-III algorithms that are solely defined via DCT I-IV, having sparse and orthogonal factors. The paper also addresses the error bounds on computing completely recursive algorithms for DCT I-IV. Moreover, this paper elaborates signal transform designs and image compression results (absence of 93.75% coefficients in each transfer block) based on the completely recursive DCT I-IV algorithms.

In Section 2, we derive factorizations for DCT-I and DCT-III having orthogonal and sparse matrices, and state completely recursive DCT I-IV algorithms solely defined via DCT I-IV having sparse, orthogonal, and rotation/rotation-reflection matrices. Next, in Section 3, we present the arithmetic cost of computing these algorithms. In Section 4, we derive error bounds in computing these algorithms and discuss the stability. Finally, in Sections 5 and 6 respectively, we demonstrate signal flow graphs and image compression results based on these completely recursive DCT I-IV algorithms.

2. Completely recursive radix-2 DCT algorithms having orthogonal factors.
This section introduces sparse and orthogonal factorizations for DCT-I and DCT-III matrices. In the meantime, we present completely recursive, radix-2 DCT I-IV algorithms solely defined via DCT I-IV, having sparse, orthogonal, and butterfly matrices. One can observe a variant of the DCT-II and DCT-IV algorithms having almost orthogonal factors in [19].

The following notations and sparse matrices are used frequently in this paper. Denote an involution matrix \( \tilde{I}_n \) by \( \tilde{I}_n x = [x_{n-1}, x_{n-2}, \ldots, x_0]^T \), a diagonal matrix \( D_n \) by \( D_n x = \text{diag}((-1)^k)^{n-1}_{k=0} x \), an even-odd permutation matrix \( P_n (n \geq 3) \) by

\[
P_n x = \begin{cases} [x_0, x_2, \ldots, x_{n-2}, x_1, x_3, \ldots, x_{n-1}]^T & \text{even } n, \\ [x_0, x_2, \ldots, x_{n-1}, x_1, x_3, \ldots, x_{n-2}]^T & \text{odd } n, \end{cases}
\]

for any \( x = [x_j]_{j=0}^{n-1} \), and orthogonal matrices \( n \geq 4 \) by

\[
\hat{H}_{n+1} = \frac{1}{\sqrt{2}} \begin{bmatrix} I_{\tilde{\Phi}} & \tilde{I}_{\tilde{\Phi}} \\ I_{\tilde{\Phi}} & -\tilde{I}_{\tilde{\Phi}} \end{bmatrix}, \quad H_n = \frac{1}{\sqrt{2}} \begin{bmatrix} I_{\tilde{\Phi}} & \tilde{I}_{\tilde{\Phi}} \\ I_{\tilde{\Phi}} & -\tilde{I}_{\tilde{\Phi}} \end{bmatrix},
\]

\[
U_n = \frac{1}{\sqrt{2}} \begin{bmatrix} I_{\tilde{\Phi}}^{-1} & I_{\tilde{\Phi}}^{-1} \\ I_{\tilde{\Phi}}^{-1} & -I_{\tilde{\Phi}}^{-1} \end{bmatrix},
\]

\[
R_n = \begin{bmatrix} I_{\tilde{\Phi}} & \text{diag} C_{\tilde{\Phi}} \\ D_{\tilde{\Phi}} & \text{diag} (\text{diag} S_{\tilde{\Phi}}) \tilde{I}_{\tilde{\Phi}} \end{bmatrix} \begin{bmatrix} \text{diag} S_{\tilde{\Phi}} \\ \text{diag} \left( I_{\tilde{\Phi}} C_{\tilde{\Phi}} \right) \end{bmatrix},
\]
where for \( k = 0, 1, \ldots, \frac{2}{2} - 1 \),

\[
C_{\frac{\pi}{4}} = \begin{bmatrix} \cos \left( \frac{(2k+1)\pi}{4n} \right) \\
\sin \left( \frac{(2k+1)\pi}{4n} \right) \end{bmatrix} \quad \text{and} \quad S_{\frac{\pi}{4}} = \begin{bmatrix} \sin \left( \frac{(2k+1)\pi}{4n} \right) \\
\cos \left( \frac{(2k+1)\pi}{4n} \right) \end{bmatrix}.
\]

DCT-II and DCT-IV algorithms are the keys for the completely recursive procedure, so for a given vector \( x \in \mathbb{R}^n \), we present algorithms in order:

\[
y = C_{\frac{\pi}{4}}^n x, \quad y = C_{\frac{\pi}{8}}^n x, \quad y = C_{\frac{\pi}{16}}^n x, \quad y = C_{\frac{\pi}{32}}^n x.
\]

Following the matrix factorizations for DCT-II and DCT-IV in [19], let us first state recursive DCT-II and DCT-IV having orthogonal factors via algorithms (2.1) and (2.2), respectively.

**Algorithm 2.1.** (\( \cos 2(x, n) \))

*Input:* \( n = 2^t (t \geq 1), \ n_1 = \frac{n}{2}, \ x \in \mathbb{R}^n \).

1. If \( n = 2 \), then
   \[
y := \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\
1 & -1 \end{bmatrix} x.
\]
2. If \( n \geq 4 \), then
   \[
   [u_{j}]_{j=0}^{n-1} := H_n x, \\
z_1 := \cos 2 \left( [u_{j}]_{j=0}^{n-1}, n_1 \right), \\
z_2 := \cos 4 \left( [u_{j}]_{j=n_1}^{n-1}, n_1 \right), \\
y := P_n^T (z_1^T, z_2^T)^T.
\]

*Output:* \( y = C_{\frac{\pi}{4}}^n x \).

**Algorithm 2.2.** (\( \cos 4(x, n) \))

*Input:* \( n = 2^t (t \geq 1), \ n_1 = \frac{n}{4}, \ x \in \mathbb{R}^n \).

1. If \( n = 2 \), then
   \[
y := \begin{bmatrix} \cos \frac{\pi}{8} & \sin \frac{\pi}{8} \\
\sin \frac{\pi}{8} & -\cos \frac{\pi}{8} \end{bmatrix} x.
\]
2. If \( n \geq 4 \), then
   \[
   [u_{j}]_{j=0}^{n-1} := R_n x, \\
z_1 := \cos 2 \left( [u_{j}]_{j=0}^{n_1-1}, n_1 \right), \\
z_2 := \cos 2 \left( [u_{j}]_{j=n_1}^{n-1}, n_1 \right), \\
w := U_n (z_1^T, z_2^T)^T, \\
y := P_n^T w.
\]

*Output:* \( y = C_{\frac{\pi}{8}}^n x \).

By using the well known transpose property between DCT-II and DCT-III we can state an algorithm for DCT-III via (2.3). This algorithm executes recursively with the DCT-II and DCT-IV algorithms.
Algorithm 2.3. ($\cos^3(x, n)$)

Input: $n = 2^t (t \geq 1), n_1 = \frac{n}{2}, x \in \mathbb{R}^n$.

1. If $n = 2$, then
   \[ y := \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} x. \]

2. If $n \geq 4$, then
   \[ [u^*]_{j=0}^{n-1} := P_n x, \]
   \[ z_1 := \cos^3 \left( [u^*]_{j=0}^{n-1}, n_1 \right), \]
   \[ z_2 := \cos^4 \left( [u^*]_{j=0}^{n-1}, n_1 \right), \]
   \[ y := H_n^T (z_1^T, z_2^T)^T. \]

Output: $y = C^{III}_n x$.

Before stating the algorithm for DCT-I let us derive a sparse and orthogonal factorization for DCT-I.

Lemma 2.4. Let $n \geq 4$ be an even integer. The matrix $C^I_{n+1}$ can be factored in the form

\[
C^I_{n+1} = P^T_{n+1} \begin{bmatrix} C^I_{\frac{n+1}{2}} & 0 \\ 0 & C^{III}_{\frac{n+1}{2}} \end{bmatrix} H_{n+1}.
\]

Proof. Let’s apply $P_{n+1}$ to $C^I_{n+1}$ to permute rows and then partition the resultant matrix. So, (1,1) block becomes

\[
\sqrt{\frac{2}{n}} \left[ \epsilon_n(2j) \epsilon_n(k) \cos \frac{2jk\pi}{n} \right]_{j,k=0}^{\frac{n-1}{2}}.
\]

(1,2) block becomes

\[
\sqrt{\frac{2}{n}} \left[ \epsilon_n(2j) \epsilon_n \left( \frac{n}{2} + k + 1 \right) \cos \frac{j(n+2k+2)\pi}{n} \right]_{j,k=0}^{\frac{n-1}{2}}.
\]

(2,1) block becomes

\[
\sqrt{\frac{2}{n}} \left[ \epsilon_n(k) \cos \frac{(2j+1)k\pi}{n} \right]_{j,k=0}^{\frac{n-1}{2}}.
\]
and (2,2) block becomes

\[ \sqrt{\frac{2}{n}} \sum (n - 2k - 2) \pi j, k = 0 \]

\[ \sqrt{\frac{2}{n}} \sum (n - 2k - 2) \pi j, k = 0 \]

Hence,

\[ P_{n+1} C_{n+1} = \frac{1}{\sqrt{2}} \begin{bmatrix} C_{n+1}^1 & 0 & C_{n+1}^{III} & 0 \\ \frac{\sqrt{2}}{C_{n+1}^1} & 0 \end{bmatrix} \tilde{H}_{n+1}. \]

Thus, an algorithm for DCT-I can be stated via (2.5), which executes recursively with DCT II-IV algorithms.

**Algorithm 2.5.** \((\cos 1(x, n + 1))\)

*Input:* \(n = 2^t (t \geq 1), n_1 = \frac{n}{2}, x \in \mathbb{R}^{n+1}.*

1. If \(n = 2,\) then
   \[ y := \sqrt{\frac{2}{n}} \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 \sqrt{2} \\ 1 & -1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 1 \\ 0 & \sqrt{2} & 0 \\ 1 & 0 & -1 \end{bmatrix} x. \]

2. If \(n \geq 4,\) then
   \[ [u_{ij}]_{j=0}^n := \tilde{H}_{n+1} x, \]
   \[ z1 := \cos 1 \left( [u_{ij}]_{j=0}^{n_1}, n_1 + 1 \right), \]
   \[ z2 := \cos 3 \left( [u_{ij}]_{j=n_1+1}^n, n_1 \right), \]
   \[ y := P_{n+1}^T (z1^T, z2^T)^T. \]

*Output:* \(y = C_{n+1}^I x.*

3. Arithmetic cost of computing DCT algorithms. We first calculate the arithmetic cost of computing DCT I-IV algorithms. Let’s denote the number of additions and multiplications required to compute, say a length \(n\) DCT-II algorithm: \(y = C_n^II x\) by \(#a(DCT-II, n)\) and \(#m(DCT-II, n)\). Note that the multiplication of \(\pm 1\) and permutations are not counted. Once the cost is computed we show numerical results for the speed improvement factor of these algorithms.

### 3.1. Number of additions and multiplications in computing DCT I-IV algorithms.

Here we calculate the arithmetic cost of computing the DCT I-IV algorithms in order (2.1), (2.2), (2.3) and (2.5). The cost of addition in computing DCT-II and DCT-IV
algorithms is the same as in [19], but the cost of multiplication is different from [19]. The latter is because in this paper, not only DCT-I and DCT-III algorithms but also DCT-II and DCT-IV algorithms have orthogonal vectors instead of almost orthogonal vectors. Let us first derive explicitly the number of multiplications required to compute DCT-II and DCT-IV algorithms and then the arithmetic cost of DCT-III and DCT-I algorithms, respectively.

**Lemma 3.1.** Let \( n = 2^t \) \((t \geq 2)\) be given. Using algorithms (2.1) and (2.2), the arithmetic cost of computing length \( n \) DCT-II algorithm is given by

\[
\#a(DCT-II, n) = \frac{4}{3}n^t - \frac{8}{9}n - \frac{1}{9}(-1)^t + 1,
\]

\[
\#m(DCT-II, n) = \frac{5}{3}n^t - \frac{10}{9}n + \frac{1}{9}(-1)^t + 1.
\]

(3.1)

**Proof.** Following algorithms (2.1) and (2.2)

\[
\#m(DCT-II, n) = \#m(DCT-II, \frac{n}{2}) + \#m(DCT-IV, \frac{n}{2}) + \#m(H_n),
\]

(3.2)

\[
\#m(DCT-IV, n) = \#m(U_n) + 2 \cdot \#m(DCT-II, \frac{n}{2}) + \#m(R_n).
\]

By referring to the structures of \( H_n, U_n, \) and \( R_n \)

\[
\#a(H_n) = n, \quad \#m(H_n) = n,
\]

\[
\#a(U_n) = n - 2, \quad \#m(U_n) = n - 2,
\]

(3.3)

\[
\#a(R_n) = n, \quad \#m(R_n) = 2n.
\]

Thus,

\[
\#m(DCT-II, n) = \#m(DCT-II, \frac{n}{2}) + 2 \cdot \#m(DCT-II, \frac{n}{4}) + \frac{5}{2}n - 2.
\]

Since \( n = 2^t \) we can obtain the second order linear difference equation with respect to \( t \)(\( \geq 3 \))

\[
\#m(DCT-II, 2^t) - \#m(DCT-II, 2^{t-1}) - 2 \cdot \#m(DCT-II, 2^{t-2}) = 5 \cdot 2^{t-1} - 2.
\]

If \( \#m(DCT-II, 2^t) = \alpha^t \)(where \( \alpha \neq 0 \)) is a solution then the above follows

\[
\alpha^t - \alpha^{t-1} - 2 (\alpha^{t-2}) = 5 \cdot 2^{t-1} - 2.
\]

(3.4)

The homogeneous solution of the above is given by solving the characteristic equation

\[
\alpha^{t-2}(\alpha^2 - \alpha - 2) = 0.
\]

From which we get

\[
\#m(DCT-II, 2^t) = r_1 2^t + r_2 (-1)^t + \text{particular solution},
\]
where \( r_1 \) and \( r_2 \) are constants. Let \( \alpha = r_3 + r_4 t \cdot 2^t \) (where \( r_3 \) and \( r_4 \) are constants) be the particular solution. Substituting this potential equation into (3.4) and equating the coefficients we can find that

\[
#m(DCT-II, 2^t) = r_1 2^t + r_2 (-1)^t + \frac{5}{3} t \cdot 2^t + 1.
\]

Using the initial conditions \( #m(DCT-II, 2) = 2 \) and \( #m(DCT-II, 4) = 10 \), we can determine the general solution

\[
(3.5) #m(DCT-II, 2^t) = \frac{5}{3} t \cdot 2^t - \frac{10}{9} 2^t + \frac{1}{9} (-1)^t + 1.
\]

Thus, substituting \( n = 2^t \) we can obtain the number of multiplications required to compute DCT-II algorithm as stated in (3.1).

Again by algorithms (2.1) and (2.2) together with (3.3), we can state

\[
#a(DCT-II, n) = #a(DCT-II, \frac{n}{2}) + 2 \cdot #a(DCT-II, \frac{n}{4}) + 2n - 2.
\]

Since \( n = 2^t \) we can obtain the second order linear difference equation with respect to \( t(\geq 3) \)

\[
#a(DCT-II, 2^t) - #a(DCT-II, 2^{t-1}) - 2 \cdot #a(DCT-II, 2^{t-2}) = 2^{t+1} - 2.
\]

As derived analogously in the cost of multiplication, we can solve the above equation under the initial conditions \( #a(DCT-II, 2) = 2 \) and \( #a(DCT-II, 4) = 8 \) to obtain

\[
(3.6) #a(DCT-II, n) = \frac{4}{3} nt - \frac{8}{9} n - \frac{1}{9} (-1)^t + 1.
\]

**Corollary 3.2.** Let \( n = 2^t (t \geq 2) \) be given. Using algorithms (2.2) and (2.1), the arithmetic cost of computing length \( n \) DCT-IV algorithm is given by

\[
\begin{align*}
#a(DCT-IV, n) & = \frac{4}{3} nt - \frac{2}{9} n + \frac{2}{9} (-1)^t, \\
#m(DCT-IV, n) & = \frac{5}{3} nt + \frac{2}{9} n - \frac{2}{9} (-1)^t.
\end{align*}
\]

**Proof.** The number of multiplications required to compute DCT-IV algorithm can be found by substituting (3.5) at \( n = 2^t - 1 \) into the equation (3.2)

\[
#m(DCT-IV, n) = n - 2 + 2 \left( \frac{5}{3} \frac{n}{2} (t - 1) - \frac{10}{9} \frac{n}{2} + \frac{1}{9} (-1)^{t-1} + 1 \right) + 2n.
\]

Simplifying the above gives the cost of multiplication

\[
#m(DCT-IV, n) = \frac{5}{3} nt + \frac{2}{9} n - \frac{2}{9} (-1)^t.
\]
Similarly, the number of additions required to compute DCT-IV algorithm can be found by substituting (3.6) at $n = 2^t - 1$ to

$$
\#a(DCT-IV, n) = \#a(U_n) + 2 \cdot \#a(DCT-II, \frac{n}{2}) + \#a(R_n)
= 2 \cdot \#a(DCT-II, \frac{n}{2}) + 2n - 2.
$$

Simplifying the above yields

$$
\#a(DCT-IV, n) = \frac{4}{3}nt - \frac{2}{9}n + \frac{2}{9}(-1)^t. \quad \square
$$

The DCT-III algorithm (2.3) was stated using the transpose property of matrices so the following corollary is trivial.

**Corollary 3.3.** Let $n = 2^t$ $(t \geq 2)$ be given. If DCT-III could be computed by using algorithms (2.3), (2.2), and (2.1), then the arithmetic cost of computing a length $n$ DCT-III algorithm is given by

$$
\#a(DCT-III, n) = \frac{4}{3}nt - \frac{8}{9}n - \frac{1}{9}(-1)^t + 1,
\#m(DCT-III, n) = \frac{5}{3}nt - \frac{22}{9}n + \frac{1}{9}(-1)^t + 1.
$$

**Remark 3.4.** By using the DCT-III algorithm (2.3) and the arithmetic cost of computing the DCT-IV algorithm (in corollary (3.2)), one can obtain the same results as in corollary (3.3).

Let us state the arithmetic cost of computing the DCT-I algorithm (2.5).

**Lemma 3.5.** Let $n = 2^t$ $(t \geq 2)$ be given. Using algorithms (2.5), (2.3), (2.2) and (2.1), the arithmetic cost of a DCT-I algorithm of length $n + 1$ is given by

$$
\#a(DCT-I, n + 1) = \frac{4}{3}nt - \frac{14}{9}n + \frac{1}{18}(-1)^t + t + \frac{7}{2},
\#m(DCT-I, n + 1) = \frac{5}{3}nt - \frac{22}{9}n - \frac{1}{18}(-1)^t + t + \frac{11}{2}.
$$

**Proof.** Referring to the DCT-I algorithm (2.5)

$$
\#a(DCT-I, n + 1) = \#a(DCT-I, \frac{n}{2} + 1) + \#a(DCT-III, \frac{n}{2}) + \#a(\tilde{H}_{n+1}).
$$

The structure of $\tilde{H}_{n+1}$ leads to $\#a(\tilde{H}_{n+1}) = n$. This together with the arithmetic cost of computing DCT-III (3.8) algorithm, we can rewrite (3.10)

$$
\#a(DCT-I, n + 1) = \#a(DCT-I, \frac{n}{2} + 1) + \frac{2}{3}nt - \frac{1}{9}n + \frac{1}{9}(-1)^t + 1.
$$
Since $n = 2^t$ we can obtain the first order linear difference equation with respect to $t(\geq 2)$

\[(3.11) \quad \#a (\text{DCT-I}, 2^t + 1) - \#a (\text{DCT-I}, 2^{t-1} + 1) = \frac{2}{3}t \cdot 2^t - \frac{1}{9}2^t + \frac{1}{9}(-1)^t + 1.\]

We can obtain the number of additions required to compute the DCT-I algorithm by solving (3.11) with initial condition $\#a (\text{DCT-I}, 3) = 4$. Analogously, one can solve the first order linear difference equation with initial condition $\#m (\text{DCT-I}, 3) = 5$ to obtain the number of multiplications.

### 3.2. Speed improvement factor of DCT I-IV algorithms.

Based on the results in lemmas 3.1, 3.5 and corollaries 3.2, 3.3, we graph the speed improvement factor of DCT I-IV algorithms having orthogonal factors. It is known that the speed improvement factor plays a critical role in the DFT algorithms as it gives us an idea about the processing speed of the algorithms. We should recall here that this factor increases with the size of matrix.

In our case, the speed improvement factor says the ratio between the number of additions and multiplications required to compute the DCT I-IV algorithms, and the direct computation cost of computing these algorithms which is $2n^2 - n$ for DCT II-IV, and $2n^2 + 3n + 1$ for DCT-I. Figure 3.1 shows the speed improvement factor corresponding to the DCT I-IV algorithms with respect to the size of matrix. These numerical data correspond to MATLAB (R2014a version) with machine precision $2 \times 10^{-16}$.

![Fig. 3.1: Speed improvement factor of DCT I-IV algorithms.](image)

### 4. Error bounds and stability of DCT algorithms.

Error bounds and stability of computing the DCT I-IV algorithms are the main concern in this section. Here, to verify the stability, we will use error bounds (using perturbation of the product of matrices stated in [8]) in computing these algorithms. Let us assume that the computed trigonometry functions ($d_r := \sin \frac{r\pi}{4n}$ or $\cos \frac{r\pi}{4n}$ are the entries of the butterfly matrix) $\hat{d}_r$ are used and satisfy

\[(4.1) \quad \hat{d}_r = d_r + \epsilon_r, \quad |\epsilon_r| \leq \mu\]
for all \( r = 1, 3, 5, \ldots, n - 1 \), where \( \mu := O(u) \) and \( u \) is the unit roundoff.

Let’s recall the perturbation of the product of matrices stated in [8, Lemma 3.7], i.e., if \( A_k + \Delta A_k \in \mathbb{R}^{n \times n} \) satisfies \( \| \Delta A_k \| \leq \delta_k \| A_k \| \) for all \( k \), then

\[
\left| \prod_{k=0}^{m} (A_k + \Delta A_k) - \prod_{k=0}^{m} A_k \right| \leq \left( m \prod_{k=0}^{m} (1 + \delta_k) - 1 \right) \prod_{k=0}^{m} |A_k|,
\]

where \( |\delta_k| < u \). Moreover, recall \( \prod_{k=1}^{n} (1 + \delta_k)^{\pm 1} = 1 + \theta_n \) where \( |\theta_n| \leq \frac{nu}{1-nu} =: \gamma_n \) and \( \gamma_k + u \leq \gamma_{k+1}, \gamma_k + \gamma_j + \gamma_k \gamma_j \leq \gamma_{k+j} \) from [8, Lemmas 3.1 and 3.3].

We explicitly derive an error bound for computing the DCT-I algorithm (as the rest follow analogously).

**Theorem 4.1.** Let \( \tilde{y} = fl(C_n^T x) \), where \( n = 2^t (t \geq 2) \), be computed using the algorithms (2.5), (2.3), (2.2), (2.1), and assume that (4.1) holds. Then

\[
\frac{\| y - \tilde{y} \|_2}{\| y \|_2} \leq \frac{\gamma t}{1 - \gamma t}.
\]

**Proof.** Using the algorithms (2.5), (2.3), (2.2), (2.1), and the computed matrices \( \tilde{B}_k \) (in terms of the computed \( \tilde{d}_k \)) for \( k = 2, 3, \ldots, t - 2 \):

\[
\tilde{y} = fl \left( A_0 A_1 \cdots A_{t-2} C_{t-1} \tilde{B}_{t-2} \cdots \tilde{B}_2 B_1 B_0 x \right)
\]

\[
= (A_0 + \Delta A_0) \cdots (A_{t-2} + \Delta A_{t-2}) (C_{t-1} + \Delta C_{t-1})
\]

\[
(\tilde{B}_{t-2} + \Delta \tilde{B}_{t-2}) \cdots (\tilde{B}_2 + \Delta \tilde{B}_2)(B_1 + \Delta B_1)(B_0 + \Delta B_0)x.
\]

Each \( A_k \) is formed containing a combination of matrices \( P_\frac{t}{2k+1}, P_\frac{t}{2k}, H_\frac{t}{2k} \), and \( U_\frac{t}{2k} \) except \( A_0 = P_\frac{t}{2+k} \) and \( A_1 = \text{blkdiag} \left( P_\frac{t}{2+k}, H_\frac{t}{2k} \right) \). Using the fact that each row in \( A_k \) has at most two non-zero entries with mostly ones per row:

\[
|\Delta A_0| = 0, \quad |\Delta A_k| \leq \gamma_2 |A_k| \quad \text{for} \quad k = 1, 2, \ldots, t - 2.
\]

Also each \( B_k \) is formed containing a combination of matrices \( \tilde{H}_{\frac{t}{2k+1}}, P_\frac{t}{2k}, R_\frac{t}{2k} \) except \( B_0 = \tilde{H}_{\frac{t}{2+k}} \) and \( B_1 = \text{blkdiag} \left( \tilde{H}_{\frac{t}{2+k}}, P_\frac{t}{2k} \right) \). Using the fact that each row in \( B_k \) has at most two non-zero entries per row:

\[
|\Delta B_0| \leq \gamma_2 |B_0|, \quad |\Delta B_1| \leq \gamma_2 |B_1|, \quad |\Delta B_k| \leq \gamma_3 |B_k| \quad \text{for} \quad k = 2, 3, \ldots, t - 2.
\]

\( C_{t-1} \) is a block diagonal matrix containing \( C_1^T, C_2^T, C_3^T \) and \( C_t^T \) hence

\[
|\Delta C_{t-1}| \leq \gamma_3 |C_{t-1}|.
\]

Using direct call of computing trigonometric functions, i.e., the view of (4.1),

\[
\tilde{B}_k = B_k + \Delta B_k, \quad |\Delta B_k| \leq \mu |B_k|.
\]
Thus, overall
\[
\hat{y} = (A_0 + \Delta A_0) \cdots (A_{t-2} + \Delta A_{t-2})(C_{t-1} + \Delta C_{t-1})
(B_{t-2} + E_{t-2}) \cdots (B_2 + E_2)(B_1 + \Delta B_1)(B_0 + \Delta B_0)x,
\]
\[
|E_k| \leq (\mu + \gamma t(1 + \mu))|B_k| \leq \gamma t|B_k|.
\]
Hence,
\[
|y - \hat{y}| \leq \left((1 + \gamma t)^t(1 + \gamma t^3) - 1\right)|A_0||A_1|\cdots|A_{t-2}|
|C_{t-1}||B_{t-2}||B_{t-3}|\cdots|B_0||x|
\]
where
\[
(1 + \gamma t)^t(1 + \gamma t^3) - 1 \leq (1 + \gamma t)^t(1 + \gamma t^3) - 2 \leq (1 + \gamma t)^t - 1 \leq \frac{\gamma t}{1 - \gamma t}.
\]
Since \(A_k, C_{t-1}, B_k\) are orthogonal matrices, \(\|A_i\|_2 = \|C_{t-1}\|_2 = \|B_k\|_2 = 1\). By orthogonality of \(C_{t+1}^t, \|y\|_2 = \|x\|_2\). Hence
\[
\|y - \hat{y}\|_2 \leq \frac{\gamma t}{1 - \gamma t} \|y\|_2.
\]

**Corollary 4.2.** \(y = C_{t+1}^t x\) is forward and backward stable.

**Proof.** The above theorem says that radix 2 DCT-I yields a tiny forward error provided that \(\sin \frac{\pi}{2^n}\) and \(\cos \frac{\pi}{2^n}\) are computed stably. It immediately follows that the computation is backward stable because \(\hat{y} = y + \Delta y = C_{t+1}^t x + \Delta y\) implies \(\hat{y} = C_{t+1}^t (x + \Delta x)\) with \(\frac{\Delta y}{y} = \frac{\Delta x}{x}\). If we form \(y = C_{t+1}^t x\) by using exact \(C_{t+1}^t\), then \(|y - \hat{y}| \leq \gamma_{t+1} \|C_{t+1}^t\|_1 \|x\|_2\) so \(\|y - \hat{y}\|_2 \leq \gamma_{t+1} \|y\|_2\).

As \(\mu\) is of order \(u\), the \(C_{t+1}^t\) has an error bound smaller than that for usual multiplication by the same factor as the reduction in complexity of the method, so DCT-I is perfectly stable.

A similar analogue as in theorem 4.1 results in the error bound for computing the DCT-II algorithm.

**Theorem 4.3.** Let \(\hat{y} = fl(C_{t}^t x)\), where \(n = 2^t (t \geq 2)\), be computed using the algorithms (2.1), (2.2), and assume that (4.1) holds, then
\[
\frac{\|y - \hat{y}\|_2}{\|y\|_2} \leq \frac{\gamma(t - 1)}{1 - \gamma(t - 1)}.
\]

Following the equivalent lines as in corollary 4.2, one can confirm the stability of the DCT-II algorithm.

**Corollary 4.4.** \(y = C_{n}^{II} x\) is forward and backward stable.

Due to the relationship between the DCT-II and DCT-III algorithms, the following results are trivial.

**Corollary 4.5.** Let \(\hat{y} = fl(C_n^{III} x)\), where \(n = 2^t (t \geq 2)\), be computed using the algorithms (2.3), (2.2), (2.1), and assume that (4.1) holds, then
\[
\frac{\|y - \hat{y}\|_2}{\|y\|_2} \leq \frac{\gamma(t - 1)}{1 - \gamma(t - 1)}.
\]

**Corollary 4.6.** \(y = C_n^{III} x\) is forward and backward stable.
Analogue of the results of theorem 4.1 and corollary 4.2 lead to:

**Theorem 4.7.** Let $\hat{y} = f(C_n^IV x)$, where $n = 2^t (t \geq 2)$, be computed using the algorithms (2.2), (2.1), and assume that (4.1) holds, then

\begin{equation}
\frac{\|y - \hat{y}\|_2}{\|y\|_2} \leq \frac{\gamma t}{1 - \gamma t}.
\end{equation}

**Corollary 4.8.** $y = C_n^IV x$ is forward and backward stable.

5. **Signal flow graphs for DCT algorithms.** Signal flow graphs commonly represent the realization of systems such as electronic devices in electrical engineering, control theory, system engineering, theoretical computer science, etc. Simply put, the objective is to build a device to implement or realize an algorithm, using devices that implement the algebraic operations used in these recursive algorithms. These building blocks are shown next in Figure 5.1.

![Fig. 5.1: Signal flow graphs building blocks.](image)

This section presents signal flow graphs for 9-point DCT-I and 8-point DCT II-IV algorithms via Figures 5.2, 5.3, 5.4, and 5.5. As shown in the flow graphs, in each graph signal flows from the left to the right. These signal flow graphs are corresponding to the decimation-in-frequency algorithms. However, one can convert these decimation-in-frequency DCT algorithms into decimation-in-time DCT algorithms. In each Figure (5.2, 5.3, 5.4, and 5.5), $\epsilon := \frac{1}{\sqrt{2}}$, $C_{i,j} := \cos \frac{i\pi}{2^t}$, and $S_{i,j} = \sin \frac{i\pi}{2^t}$.

![Fig. 5.2: Flow graph for 9-point DCT-I algorithm.](image)
As shown in Figures 5.3, 5.4, and 5.5, the input signals are in order: \( x = \{x(0), x(1), \ldots, x(7)\} \) and output signals are in bit-reversed order: \( y = \{y(0), y(4), y(2), y(6), y(1), y(5), y(3), y(7)\} \). In bit-reversed order, each output index is represented as a binary number and the indices’ bits are reversed. Say for 8-point DCT II, the sequential order of the input indices’ bits is \{000, 001, 010, 011, 100, 101, 110, 111\}. Then reversing these input signal bits yields \{000, 100, 010, 110, 001, 101, 011, 111\} which is the output signal.
6. Image compression results based on DCT algorithms. Discretized images can be considered as matrices. To compress such images one can apply different quantization techniques. In this section, we analyze quantization techniques using these recursive DCT-II and DCT-IV algorithms to compress the Polygon image of size 512 × 512 pixels. At first, we convert the RGB Polygon image into a grayscale scale image. Next, the image is discretized into 8 × 8, 16 × 16, and 32 × 32 blocks. After that, using these recursive DCT-II and DCT-IV algorithms, two-dimensional DCTs of each block in the image are computed. Then the DCT-II and DCT-IV coefficients are quantized by transforming absence of 93.75% of the DCT coefficients in each block. e.g. for 8 × 8 transfer block, after recursively computing two-dimensional DCT-II and DCT-IV of each 8 × 8 block in the image, we discard all but 4 of the 64 DCT coefficients in each block, i.e., applying the matrix
\[
\begin{bmatrix}
1 & 1 \\
1 & 1 \\
0_{6\times6}
\end{bmatrix},
\]
where \(0_{6\times6}\) is the zero matrix of order 6 × 6 to each transform block, and analogously for 16 × 16 and 32 × 32. Next, we reconstruct the image using the inverse two-dimensional DCT-II and DCT-IV of each blocks. Finally, putting each block back together into a single image leads to Figures 6.1 and 6.2. Simply put, we analyzed image compression results via the RGB Polygon image with 93.75% of discarded coefficients in each transfer block using these recursive DCT-II and DCT-IV algorithms.

Figure 6.1 shows images with discarded coefficients (except the top left 6.25%) in each transfer block, after applying DCT-II algorithm, and then running recursively with the DCT-IV algorithm.

Figure 6.2 shows images with discarded coefficients (except the top left 6.25%) in each transfer block after applying DCT-IV algorithm and then running recursively with the DCT-II algorithm.

Comparing to Figures 6.1 and 6.2, the image reconstruction results corresponding to DCT-II algorithm are better than that of the DCT-IV algorithm. Though the quality of reconstructed images in Figures 6.1 and 6.2 are somewhat lost, those images are clearly recognizable even though
Fig. 6.1: (6.1a) Original polygon image. (6.1b) Reconstructed image with 93.75% discarded DCT-II coefficients in each 8 \times 8 transfer block. (6.1c) Reconstructed image with 93.75% discarded DCT-II coefficients in each 16 \times 16 transfer block. (6.1d) Reconstructed image with 93.75% discarded DCT-II coefficients in each 32 \times 32 transfer block.

93.75% of the DCT-II and DCT-IV coefficients are discarded in each transfer block.

7. Conclusion. This paper provided stable, completely recursive, radix-2 DCT-I and DCT-III algorithms having sparse, orthogonal and rotation/rotation-reflection matrices, defined solely via DCT I-IV algorithms. The arithmetic cost and error bounds of computing DCT I-IV algorithms are addressed. Signal flow graphs are presented for these solely based on orthogonal factorization of
Fig. 6.2: (6.2a) Original polygon image. (6.2b) Reconstructed image with 93.75% discarded DCT-IV coefficients in each $8 \times 8$ transfer block. (6.2c) Reconstructed image with 93.75% discarded DCT-IV coefficients in each $16 \times 16$ transfer block. (6.2d) Reconstructed image with 93.75% discarded DCT-IV coefficients in each $32 \times 32$ transfer block.

DCT I-IV in decimation-of-frequency. Using the recursive DCT-II and DCT-IV algorithms with the absence of 93.75% coefficients in each transfer block in 2D DCT-II and DCT-IV, one can reconstruct $512 \times 512$ images without seriously affecting the quality.

Acknowledgment. The author would like to thank James G. Nagy and the anonymous referee for their helpful suggestions that allowed to improve the exposition of the manuscript.
REFERENCES


