Manish Okade (National Institute of Technology, Rourkela)
and Jayanta Mukherjee (Indian Institute of Technology, Kharagpur)

Discrete Cosine Transform: A Revolutionary Transform That Transformed Human Lives

The widely known Discrete Fourier Transform (DFT), the frequency-domain representation of a finite-length time-domain sequence is an orthogonal transform and has been known for a very long time and has found many application. The popularity of the DFT increased tremendously after the publication of the Fast Fourier Transform (FFT) algorithm by Cooley and Tukey in 1965 [1]. Orthogonal transforms offer many advantages, namely fast computational speeds, less storage space, less rounding off errors, etc. The benefits of the orthogonal transforms stem from the fact that they can be factored based on matrix computations. Data compression, digital filter design, speech processing, image and video processing are applications where the orthogonal transforms have made a significant impact.

Unfortunately, the FFT samples are complex-valued functions even though the original time-domain sequence has only real-valued samples. In 1974, Nasir Ahmed with T. Natarajan and K. Rao, proposed the Discrete Cosine Transform (DCT), also an orthogonal transform, which is a real-valued frequency-domain representation of a finite-length time-domain sequence with real-valued samples [2]. DCT became extremely popular due to its energy compaction property, data decorrelation property, and availability of a fast transform to speed up the computations. Moreover, DCT comes very close compared to the Karhunen-Loeve Transform (KLT), which to date is the optimal transform in the mean square sense yet is limited due to the lack of a fast transform to compute it as it is signal dependent. Although it had a humble beginning in the sense that it was being investigated with the motivation to achieve an approximation to the KLT along with being computationally faster, yet the reach and the scale that DCT has achieved in the 21st century is beyond words. The interesting story of how DCT was developed can be found in [3]. Nasir Ahmed's account on the genesis of the DCT can be seen in the video "The Algorithm That Transform The World: The Story Of Nasir Ahmed"' (https://www.youtube.com/watch?v=I9VXaVVs7WY).

Digital Object Identifier 10.1109/MCAS.2022.3214566
Date of current version: 13 January 2023

1. Transform Coding

In transform coding, N discrete-time data samples are firstly transformed, then the coefficients are individually quantized and entropy coded into binary bits serving two purposes: (i) to compact the energy of the original N samples into coefficients with increasingly smaller variances so that removing smaller coefficients have negligible reconstruction errors; and (ii) to decorrelate the original samples so that the coefficients can be quantized and entropy coded individually. Karhunen-Loeve transform (KLT) is an optimal transform for a signal with a stationary covariance matrix in the sense that it completely decorrelates the original samples and maximizes the energy compaction. Ahmed et al. [4] showed rigorously that DCT has an energy compaction performance almost as good as the KLT, and the DCT can be readily derived as the limiting case of the KLT of the first-order Markov processes, as the correlation coefficient approaches unity. Fortunately, most real-world 1-D or multi-dimensional signals can be well modeled by a first-order Markov process, making DCT superior to other orthogonal transforms for signal compression.

2. Discrete Cosine Transform

Orthogonal expansion for an even periodic function is possible with only cosine harmonics, while an odd periodic function requires only sine harmonics as basis-functions. These properties are suitably exploited for a finite discrete-time sequence to have orthogonal transforms with only discrete cosine or discrete sine basis vectors. These transforms are defined under the same framework of Generalized Discrete Fourier Transform (GDFT) by symmetrically or anti-symmetrically extending the sequence at both ends so that the extended sequence becomes either an even or an odd sequence about a point of symmetry. Thus, all these expansions can be expressed by a linear combination of either discrete cosine or discrete sine basis vectors. These transforms, in general, are referred to as discrete trigonometric transforms. Therefore, their expansions with either a set of cosine or sine basis vectors are possible. In this article, the development of the discrete cosine transform is considered. A finite discrete sequence of length N can be symmetrically extended at its endpoint in two ways. First, the extended sequence is symmetric about the end point itself, referred to as the whole-sample symmetric (WS) extension [5]. Second, the symmetric extension is centered about the midpoint between the end sample, referred to as the half-sample symmetric (HS) extension. Likewise, two types of antisymmetric extensions about a sample point of a sequence are possible, referred to as the whole-sample antisymmetric (WA) extension and the half-sample antisymmetric extension. This results in 16 distinct types of periodic extensions, of which 8 are symmetric-periodic, leading to 8 different types of DCTs, and 8 are antisymmetric-periodic, leading to 8 different types of DSTs. In this article, the focus is on DCTs, especially the Type-I and Type-II DCTs, due to the advantages their properties offer and their applicability in a wide range of applications. The governing equations for the Type-I DCT X_I and Type-II DCT X_{II} are given below [6]:

$$X_I(k) = \sqrt{\frac{2}{N}}\alpha^2(k)\sum_{n=0}^{N} x(n)\cos\frac{2\pi nk}{2N}, 0 \le k \le N \tag{1}$$

$$X_{II}(k) = \sqrt{\frac{2}{N}}\alpha(k)\sum_{n=0}^{N-1} x(n)\cos\frac{\pi k(2n+1)}{2N}, 0 \le k \le N-1 \tag{2}$$

where

$$\alpha(k) = \begin{cases} \sqrt{\frac{1}{2}}, & \text{if } k = 0 \text{ or } N \\ 1, & \text{otherwise.} \end{cases} \tag{3}$$

If the sequence $x(n)(n = 0, 1, \ldots N-1)$ is represented by N-dimensional column vector $\mathbf{x}$. The Type-II DCT of $\mathbf{x}$ can be expressed in matrix form as

$$\mathbf{X} = \mathbf{C_N} \cdot \mathbf{x} \tag{4}$$

where the $(k, n)^{th}$ element of $\mathbf{C_N}$ is given by

$$\mathbf{C_N}(k,n) = \sqrt{\frac{2}{N}}\alpha(k)\cos\frac{\pi k(2n+1)}{2N}, \quad 0 \le (k,n) \le N-1. \tag{5}$$

As in the conventional literature, by the word DCT, it implies Type-II DCT, we also follow the same convention. A Fast DCT (FDCT) algorithm [7] following an approach similar to Fast Fourier Transform (FFT) exists, i.e., by splitting the DCT sum into odd and even terms, making DCT popular in the research community.

3. DCT Properties

The popularity of the Type-II DCT is due to the number of interesting properties [5] as summarized below:

1) *Linearity and Orthogonality:* Since $\alpha x(n) + \beta y(n) \to \alpha X_{DCT}(k) + \beta Y_{DCT}(k)$ holds, DCT is said to be satisfy linearity. Linear DCT transform matrix $\mathbf{C}$ is real and orthogonal i.e. $\mathbf{C}^{-1} = \mathbf{C}^{\mathrm{T}}$ where

$$c(k,n) = \begin{cases} \alpha(0), & \text{for } k = 0 \\ \alpha(k)\cos\frac{\pi k(2n+1)}{2N}, & \text{otherwise.} \end{cases} \tag{6}$$

DCT is related to the DFT of a symmetrically extended signal, which gives less discontinuity at the boundaries and better energy compaction, unlike the DFT, which introduces discontinuities.

2) *Energy Conservation and Decorrelation:* Consider the transform as $\mathbf{y} = \mathbf{Cx}$ in vector form. Energy conservation property is derived as follows.

$$\begin{aligned} \|\mathbf{y}\|^2 = \|\mathbf{Cx}\|^2 = (\mathbf{Cx})^{\mathrm{T}}(\mathbf{Cx}) &= \mathbf{x}^{\mathrm{T}}\mathbf{C}^{\mathrm{T}}\mathbf{Cx} \\ &= \mathbf{x}^{\mathrm{T}}\mathbf{x} = \|\mathbf{x}\|^2 \end{aligned} \tag{7}$$

A large part of signal energy is packed in a few transform coefficients, typically in the low-frequency range, commonly known as the energy compaction property. The decorrelation property removes the high amount of correlation that exists in the spatial domain signal when DCT is applied to it. This can be validated by estimating the covariance matrix of the transformed signal, i.e., $E[(\mathbf{y} - E(\mathbf{y}))(\mathbf{y} - E(\mathbf{y}))^T]$where the off-diagonal elements in the covariance matrix tend to be small. Due to this property, the angles between vectors are preserved.

3) *Matrix Factorization:* $\mathbf{C_N}$, the N-point DCT matrix can be factored into a product of a few sparse matrices.

$$\mathbf{C_N} = \sqrt{\frac{1}{2}}\mathbf{P_N}\begin{bmatrix} \mathbf{C}_{\frac{N}{2}} & 0 \\ \mathbf{0}^2 & \mathbf{J}_{\frac{N}{2}}\mathbf{C}_{\frac{N}{2}}\mathbf{J}_{\frac{N}{2}} \end{bmatrix}\begin{bmatrix} \mathbf{I}_{\frac{N}{2}} & \mathbf{J}_{\frac{N}{2}} \\ \mathbf{J}_{\frac{N}{2}} & -\mathbf{I}_{\frac{N}{2}} \end{bmatrix} \tag{8}$$

where, $\mathbf{P_N}$, $\mathbf{I_N}$, $\mathbf{J_N}$, and $\mathbf{0_N}$ are the $N \times N$ permutation matrix, identity, reverse identity, and zero matrices, respectively. This property is useful in designing fast matrix multiplication operations for various algorithms.

4) *Subband Relationship and Approximate DCT Computation:* Let the sequence $x(n)$, $0 \le n \le (N-1)$ be an N-point sequence with even N. Let $x(n)$ be decomposed into two subbands $x_L(n)$ and $x_H(n)$ of length $\frac{N}{2}$ each as follows:

$$\begin{aligned} x_L(n) &= \frac{x(2n) + x(2n+1)}{2} \\ x_H(n) &= \frac{x(2n) - x(2n+1)}{2}, n = 0, 1, \ldots \frac{N}{2} - 1. \end{aligned} \tag{9}$$

The relationship between the DCT of subbands (of $\frac{N}{2}$ point), $X_L(k)$ and $S_H(k)$, with the original DCT (of N point) of the sequence, $X(k)$, is given by

$$X(k) = \sqrt{2}\cos\frac{\pi k}{2N}\overline{X_L}(k) + \sqrt{2}\sin\frac{\pi k}{2N}\overline{S_H}(k), 0 \le k \le N-1 \tag{10}$$

where,

$$\overline{X_L}(k) = \begin{cases} X_L(k), & 0 \le k \le \frac{N}{2}-1 \\ 0, & k = \frac{N}{2} \\ -X_L(N-k), & \frac{N}{2}+1 \le k \le N-1 \end{cases} \tag{11}$$

and

$$\overline{S_H}(k) = \begin{cases} S_H(k), & 0 \le k \le \frac{N}{2}-1 \\ \sqrt{2}\sum_{n=0}^{\frac{N}{2}-1}(-1)^n x_H(n), & k = \frac{N}{2} \\ S_H(N-k), & \frac{N}{2}+1 \le k \le N-1. \end{cases} \tag{12}$$

Since DCT packs most of the energy in the low-frequency coefficients, Eq. (10) can be approximated to contain only the cosine term referred as subband approximation of DCT given below.

$$X(k) \approx \begin{cases} \sqrt{2}\cos\frac{\pi k}{2N}\overline{X_L}(k), & k \in 0, 1, \ldots \frac{N}{2}-1 \\ 0, & \text{otherwise} . \end{cases} \tag{13}$$

5) *Decimation and Interpolation Property:* These properties are useful for decimating signals directly in the transform domain as well as carrying out the interpolation operations. Let us denote Type-I DCT and Type-II DCT of $x(n)$ as $X_1(k) = C_1(x(n))$, $k.n = 0, 1, 2, \ldots, N$ and $X_{II}(k) = C_2(x(n)), k,n = 0,1,2,\ldots,\frac{N}{2}$, respectively, where $C_1(.)$ **and** $C_2(.)$, refers to the corresponding transformation functions as given in Eqs. (1) and (2). Then, the decimation is expressed as

$$x_d(n) = x(2n+1) = C_2^{-1}\left\{\frac{X_I(k) - X_I(N-k)}{\sqrt{2}}\right\}, \quad k,n = 0, 1, 2, \ldots, \frac{N}{2}-1 \tag{14}$$

Similarly, the interpolation is expressed as

$$x_u(n) = C_1^{-1}\left\{\frac{X_{II}(k) - X_{II}(N-k)}{\sqrt{2}}\right\}, \quad k,n = 0, 1, 2, \ldots, N \tag{15}$$

where,

$$x_u(n) = \begin{cases} 0, & n \text{ even} \\ x\left(\frac{n-1}{2}\right), & n \text{ odd} \end{cases} \tag{16}$$

6) *Convolution-Multiplication Property:* Convolution between two sequences $x(n)$ and $y(n)$ is defined as the convolution between their symmetrically extended sequences of the same general period. When both the extended sequences are periodic, the convolution operation is defined as the circular convolution. If both are antiperiodic, the operation is defined as the skew-circular convolution. Let $u(n) = x(n) \circledast y(n), x(n)$ and $y(n)$ are such that their symmetric extensions produce a general period of 2 N. The convolution multiplication property for DCT-I and DCT-II are given as

$$C_1(u(n)) = \sqrt{2N}C_1(x(l))C_1(y(m)), 0 \le n, l, m \le N \tag{17}$$

$$C_2(u(n)) = \sqrt{2N}C_2(x(l))C_1(y(m)), \quad 0 \le n, l \le N-1, 0 \le m \le N \tag{18}$$

7) *Integer Cosine Transforms (ICT):* ICTs are derived from a DCT matrix as its integer approximations and preserves its properties of orthogonality, symmetry, relative order, and sign of the elements of the matrix. The motivation for utilization of ICTs is to achieve fast computations. For example, H.264 video codec utilizes a 4-point integer transform which is useful during matrix multiplication operations.

4. Applications

The DCT is a core technology behind almost all international standards for image and video compression, including the ISO JPEG image compression standard (1992) and a series of video coding standards from ITU and ISO: H.261 (1988), MPEG-1 (1993), MPEG-2 (1995), H.263 (1988), MPEG-4 (1998), H.264/AVC (2003), H.265/HEVC (2013). The impact of these standards in todays life is immense. Every digital camera and mobile phone camera supports the JPEG format, which is also used in essentially all websites. Other image formats using wavelets exist (JPEG 2000), but the low computation cost of JPEG, thanks to efficient DCT algorithms and its good compression performance, made JPEG much more popular. The impact of the ISO/ITU-T video formats based on the DCT is immense: early digital television and DVDs used the MPEG-2 format, and early videoconferencing systems used the H.263 format. Today, almost all digital video systems use either H.264

(for HD video) or H.265 (for Ultra HF/4k video). When we capture video on our phones, watch TV, Netflix, Amazon video, or make a Zoom or Teams call, we use these formats, which are all DCT-based. Today, almost 80% of all data traffic on the Internet is on video streams in those formats. Ahmed et al.'s [2] original DCT work was also an inspiration for designing DCT variants and extensions, such as the modified discrete cosine transform (MDCT) and lapped transforms (LTs), with overlapping basis functions that improve compression and reconstruction quality for images and audio. Essentially all modern digital audio formats, such as MP3, WMA, and AAC, are based on these digital variants. The JPEG XR image format also uses one of those variants and is used extensively by hundreds of millions of people, as it is integrated with all Microsoft Office applications. Moreover, compressed domain image and video processing which leverages the DCT coefficients, their residues along with other information like block coding modes, motion vectors have offered a completely new paradigm for algorithm development courtesy the unique DCT properties.

References

[1] J. W. Cooley and J. W. Tukey, "An algorithm for the machine calculation of complex Fourier series," *Math. Comput.*, vol. 19, no. 90, pp. 297–301, 1965.

[2] N. Ahmed, T. Natarajan, and K. Rao, "Discrete cosine transform," *IEEE Trans. Comput.*, vol. C-23, no. 1, pp. 90–93, Jan. 1974.

[3] N. Ahmed, "How I came up with the discrete cosine transform," *Digit. Signal Process.*, vol. 1, no. 1, pp. 4–5, Jan. 1991.

[4] M. D. Flickner and N. Ahmed, "A derivation for the discrete cosine transform," *Proc. IEEE*, vol. 70, no. 9, pp. 1132–1134, Sep. 1982.

[5] S. A. Martucci, "Symmetric convolution and the discrete sine and cosine transforms," *IEEE Trans. Signal Process.*, vol. 42, no. 5, pp. 1038–1051, May 1994.

[6] J. Mukhopadhyay, *Image and Video Processing in the Compressed Domain*. Boca Raton, FL, USA: CRC Press, 2011.

[7] J. Makhoul, "A fast cosine transform in one and two dimensions," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-28, no. 1, pp. 27–34, Feb. 1980.

Manish Okade (okadem@nitrkl.ac.in), (Senior Member, IEEE) is an Assistant Professor of electronics and communication engineering at the National Institute of Technology, Rourkela, India.

Jayanta Mukherjee (jay@cse.iitkgp.ac.in), (Senior Member, IEEE) is a Professor of computer science and engineering at the Indian Institute of Technology Kharagpur, India.

Interview *(continued from page 5)*

I do not agree when sometimes people say we have no prejudices, when you can cite example after example of things that wouldn't have happened if it wasn't for such a prejudice. So I hope this will change, and I think we have to work at it because unless we work at it, change will not come.

Question 8: You are on board of director of Enhancing Diversity in Graduate Education (EDGE). What role does this organization play?

Prof. Daubechies: I have worked with EDGE for many years. At first, I have been on the periphery, and now on the board. EDGE has helped young women, especially women from disadvantaged backgrounds who are interested in mathematics, to persist in mathematics. Even if their background may be less prepared than others, EDGE helps them with workshops, special activities, and a network that gives them both moral and mathematical support. The success rate of young people who came through EDGE and then went on to PhDs in mathematics is very high, which is very impressive. EDGE now also has a program that is not aimed only at women but at underrepresented mid-career fellows. It is a really remarkable program and I think it's the same idea again - but now for other groups in which we see underrepresentation in STEM, like people of color and African Americans, to an even more dramatic extent than for women. There are young people who have talent, and we should encourage them if they have the inclination to become STEM professionals. That's why I'm proud of my son, because at his school, 95% are black or Hispanic. Having a positive, fun math teacher can be life changing.

Question 9: What is your own definition of "success", or do you consider one of your past roles to be "success"?

Prof. Daubechies: Things that make me most proud is being a member of the community. If my students later say that they really appreciate my mentorship and they felt that I made a difference for them for the better, that is what I find most successful. I now have the next generation. My son has two small children and asked me to help. I am actually taking a week off next week to be a grandma, which also makes me feel very proud. I think the feeling that I have a meaningful role in people's lives gives me the greatest sense of accomplishment.