The Discrete Cosine Transform

Gilbert Strang
Massachusetts Institute of Technology
http://www-math.mit.edu/∼gs
gs@math.mit.edu
to appear in *SIAM Review*

Abstract

Each Discrete Cosine Transform uses N real basis vectors whose components are cosines. In the DCT-4, for example, the jth component of $\mathbf{v}_k$ is $\cos(j+\frac{1}{2})(k+\frac{1}{2})\frac{\pi}{N}$. These basis vectors are orthogonal and the transform is extremely useful in image processing. If the vector $\mathbf{x}$ gives the intensities along a row of pixels, its cosine series $\sum c_k\mathbf{v}_k$ has the coefficients $c_k = (\mathbf{x}, \mathbf{v}_k)/N$. They are quickly computed from an FFT. But a direct proof of orthogonality, by calculating inner products, does not reveal how natural these cosine vectors are.

We prove orthogonality in a different way. Each DCT basis contains the eigenvectors of a symmetric "second difference" matrix. By varying the boundary conditions we get the established transforms DCT-1 through DCT-4. Other combinations lead to four additional cosine transforms. The type of boundary condition (Dirichlet or Neumann, centered at a meshpoint or a midpoint) determines the applications that are appropriate for each transform. The centering also determines the period: $N-1$ or N in the established transforms, $N-\frac{1}{2}$ or $N+\frac{1}{2}$ in the other four. The key point is that all these "eigenvectors of cosines" come from simple and familiar matrices.

0. Introduction

Just as the Fourier series is the starting point in transforming and analyzing periodic functions, the basic step for vectors is the Discrete Fourier Transform (DFT). It maps the "time domain" to the "frequency domain." A vector with N components is written as a combination of N special basis vectors $\mathbf{v}_k$. Those are constructed from powers of the complex number $w = e^{2\pi i/N}$:

$$\mathbf{v}_k = \left(1, w^k, w^{2k}, \ldots, w^{(N-1)k}\right), \quad k = 0, 1, \ldots, N-1\,.$$

The vectors $\mathbf{v}_k$ are the columns of the Fourier matrix $F = F_N$. *Those columns are orthogonal.* So the inverse of F is its conjugate transpose, divided by $\| \mathbf{v}_k \|^2 = N$. The discrete Fourier series $\mathbf{x} = \sum c_k \mathbf{v}_k$ is $\mathbf{x} = F\mathbf{c}$. The inverse $\mathbf{c} = F^{-1}\mathbf{x}$ uses $c_k = (\mathbf{x}, \mathbf{v}_k)/N$ for the (complex) Fourier coefficients.

Two points to mention, about orthogonality and speed, before we come to the purpose of this note. For these DFT basis vectors, a direct proof of orthogonality is very efficient:

$$(\mathbf{v}_k, \mathbf{v}_\ell) = \sum_{j=0}^{N-1} (w^k)^j (\bar{w}^\ell)^j = \frac{(w^k \bar{w}^\ell)^N - 1}{w^k \bar{w}^\ell - 1}\,.$$

The numerator is zero because $w^N = 1$. The denominator is nonzero because $k \neq \ell$. This proof of $(\mathbf{v}_k, \mathbf{v}_\ell) = 0$ is short but not very revealing. I want to recommend a different proof, which recognizes the $\mathbf{v}_k$ as *eigenvectors.* We could work with any circulant matrix, and we will choose below a symmetric A_0. Then linear algebra guarantees that its eigenvectors $\mathbf{v}_k$ are orthogonal.

Actually this second proof, verifying that $A_0\mathbf{v}_k = \lambda_k \mathbf{v}_k$, brings out a central point of Fourier analysis. The Fourier basis diagonalizes every periodic constant coefficient operator.

Each frequency k (or $2\pi k/N$) has its own frequency response λ_k. The complex exponential vectors $\mathbf{v}_k$ are important in applied mathematics because they are eigenvectors!

The second key point is speed of calculation. The matrices F and F^{-1} are full, which normally means N^2 multiplications for the transform and the inverse transform: $\mathbf{y} = F\mathbf{x}$ and $\mathbf{x} = F^{-1}\mathbf{y}$. But the special form $F_{jk} = w^{jk}$ of the Fourier matrix allows a factorization into very sparse and simple matrices. This is the Fast Fourier Transform (FFT). It is easiest when N is a power 2^L. The operation count drops from N^2 to $\frac{1}{2}NL$, which is an enormous saving. But the matrix entries (powers of w) are complex.

The purpose of this note is to consider *real transforms that involve cosines.* Each matrix of cosines yields a Discrete Cosine Transform (DCT). There are four established types, DCT-1 through DCT-4, which differ in the boundary conditions at the ends of the interval. (This difference is crucial. The DCT-2 and DCT-4 are constantly applied in image processing; they have an FFT implementation and they are truly useful.) All four types of DCT are orthogonal transforms. The usual proof is a direct calculation of inner products of the N basis vectors, using trigonometric identities.

We want to prove this orthogonality in the second (indirect) *way.* The basis vectors of cosines are actually eigenvectors of symmetric second-difference matrices. This proof seems more attractive, and ultimately more useful. It also leads us, by selecting different boundary conditions, to four less familiar cosine transforms. The complete set of eight DCT's was found in 1985 by Wang [10], and we want to derive them in a simple way. We begin now with the DFT.

1. The Periodic Case and the DFT

The Fourier transform works perfectly for periodic boundary conditions (and constant coefficients). For a second difference matrix, the constant diagonals contain -1 and 2 and -1. The diagonals with -1 loop around to the upper right and lower left corners, by periodicity, to produce a circulant matrix:

$$A_0 = \begin{bmatrix} 2 & -1 & & & -1 \\ -1 & 2 & -1 & & \\ & & \ddots & & \\ & & -1 & 2 & -1 \\ -1 & & & -1 & 2 \end{bmatrix}.$$

For this matrix A_0, and every matrix throughout the paper, we look at three things:

1. The interior rows

2. The boundary rows (rows 0 and $N-1$)

3. The eigenvectors

The interior rows will be the same in every matrix! The jth entry of $A_0 u$ is $-u_{j-1} + 2u_j - u_{j+1}$, which corresponds to $-u''$. This choice of sign makes each matrix positive definite (or at least semidefinite). No eigenvalues are negative.

At the first and last rows ($j = 0$ and $j = N-1$), this second difference involves u_{-1} and u_N. It reaches beyond the boundary. Then the periodicity $u_N = u_0$ and $u_{N-1} = u_{-1}$ produces the -1 entries that appear in the corners of A_0.

Note: *The numbering throughout this paper goes from* 0 *to* $N-1$, since SIAM is glad to be on very friendly terms with the IEEE. But we still use i for $\sqrt{-1}$! No problem anyway, since the DCT is real.

We now verify that $\mathbf{v}_k = (1, w^k, w^{2k}, \ldots, w^{(N-1)k})$ is an eigenvector of A_0. It is periodic because $w^N = 1$. The jth component of $A_0\mathbf{v}_k = \lambda_k\mathbf{v}_k$ is the second difference

$$\begin{aligned}
-w^{(j-1)k} + 2w^{jk} - w^{(j+1)k} &= \left(-w^{-k} + 2 - w^k\right) w^{jk} \\
&= \left(-e^{-2\pi ik/N} + 2 - e^{2\pi ik/N}\right) w^{jk} \\
&= \left(2 - 2\cos\frac{2k\pi}{N}\right) w^{jk}\,.
\end{aligned}$$

A_0 is symmetric and those eigenvalues $\lambda_k = 2 - 2\cos\frac{2k\pi}{N}$ are real. The smallest is $\lambda_0 = 0$, corresponding to the eigenvector $\mathbf{v}_0 = (1, 1, \ldots, 1)$. In applications it is very useful to have this flat DC vector (direct current in circuit theory, constant gray level in image processing) as one of the basis vectors.

Since A_0 is a real symmetric matrix, its orthogonal eigenvectors can also be chosen real. In fact the real and imaginary parts of the $\mathbf{v}_k$ must be eigenvectors:

$$\begin{aligned}
\mathbf{c}_k &= \operatorname{Re}\mathbf{v}_k = \left(1, \cos\frac{2k\pi}{N}, \cos\frac{4k\pi}{N}, \ldots, \cos\frac{2(N-1)k\pi}{N}\right) \\
\mathbf{s}_k &= \operatorname{Im}\mathbf{v}_k = \left(0, \sin\frac{2k\pi}{N}, \sin\frac{4k\pi}{N}, \ldots, \sin\frac{2(N-1)k\pi}{N}\right).
\end{aligned}$$

The equal pair of eigenvalues $\lambda_k = \lambda_{N-k}$ gives the two eigenvectors $\mathbf{c}_k$ and $\mathbf{s}_k$. The exceptions are $\lambda_0 = 0$ with one eigenvector $\mathbf{c}_0 = (1, 1, \ldots, 1)$, and for even N also $\lambda_{N/2} = 4$ with $\mathbf{c}_{N/2} = (1, -1, \ldots, 1, -1)$. Those two eigenvectors have length $\sqrt{N}$, while the other $\mathbf{c}_k$ and $\mathbf{s}_k$ have length $\sqrt{N/2}$. It is these exceptions that make the real DFT (sines together with cosines) less attractive than the complex form. That factor $\sqrt{2}$ is familiar from ordinary Fourier series. It will appear in the $k = 0$ term for the DCT-1 and DCT-2, always with the flat basis vector $(1, 1, \ldots, 1)$.

We expect the cosines alone, without sines, to be complete over a half-period. In Fourier series this changes the interval from $[-\pi, \pi]$ to $[0, \pi]$. Periodicity is gone because $\cos 0 \neq \cos \pi$. The differential equation is still $-u'' = \lambda u$. The boundary condition that produces cosines is $u'(0) = 0$. Then there are two possibilities, Neumann and Dirichlet, at the other boundary:

Zero slope:	$u'(\pi) = 0$	gives eigenfunctions	$u_k(x) = \cos kx$
Zero value:	$u(\pi) = 0$	gives eigenfunctions	$u_k(x) = \cos\left(k + \frac{1}{2}\right) x\,.$

The two sets of cosines are orthogonal bases for $L^2[0, \pi]$. The eigenvalues from $-u_k'' = \lambda u_k$ are $\lambda = k^2$ and $\lambda = \left(k + \frac{1}{2}\right)^2$.

All our attention now goes to the discrete case. The key point is that every boundary condition has two fundamental approximations. At each boundary, the condition on u can be imposed *at a meshpoint or at a midpoint.* So each problem has four basic discrete approximations. (More than four, if we open up to further refinements in the boundary conditions — but four are basic.) Often the best choices use the same centering at the two ends — both meshpoint centered or both midpoint centered.

In our problem, $u'(0) = 0$ at one end and $u'(\pi) = 0$ or $u(\pi) = 0$ at the other end yield eight possibilities. Those eight combinations produce *eight cosine transforms.* Starting from $u(0) = 0$ instead of $u'(0) = 0$, there are also eight sine transforms. Our purpose is to organize this approach to the DCT (and DST), by describing the second difference matrices and identifying their eigenvectors.

Each of the eight (or sixteen) matrices has the tridiagonal form

$$A = \begin{bmatrix} \otimes & \otimes & & & & \\ -1 & 2 & -1 & & & \\ & -1 & 2 & -1 & & \\ & & \cdot & \cdot & \cdot & \\ & & & -1 & 2 & -1 \\ & & & & \boxtimes & \boxtimes \end{bmatrix}. \tag{1}$$

The boundary conditions decide the eigenvectors, with four possibilities at each end: Dirichlet or Neumann, centered at a meshpoint or a midpoint. The reader may object that symmetry requires off-diagonal -1's in the first and last rows. The meshpoint Neumann condition produces -2. So we admit that the eigenvectors in that case need a rescaling at the end (only involving $\sqrt{2}$) to be orthogonal. The result is a beautifully simple set of basis vectors. We will describe their applications in signal processing.

2. The Discrete Cosine Transform

The discrete problem is so natural, and almost inevitable, that it is really astonishing that the DCT was not discovered until 1974 [1]. Perhaps this time delay illustrates an underlying principle. Each continuous problem (differential equation) has many discrete approximations (difference equations). The discrete case has a new level of variety and complexity, often appearing in the boundary conditions.

In fact the original paper by Ahmed, Natarajan, and Rao derived the DCT-2 basis as approximations to the eigenvectors of an important matrix, with entries $\rho^{|j-k|}$. This is the covariance matrix for a useful class of signals. The number ρ (near 1) measures the correlation between nearest neighbors. The true eigenvectors would give an optimal "Karhunen-Loève basis" for compressing those signals. The simpler DCT vectors are close to optimal (and independent of ρ).

The four standard types of DCT are now studied directly from their basis vectors (recall that j and k go from 0 to $N-1$). The jth component of the kth basis vector is

DCT-1:	$\cos jk\frac{\pi}{N-1}$	(divide by $\sqrt{2}$ when j or k is 0 or $N-1$)
DCT-2:	$\cos\left(j+\frac{1}{2}\right)k\frac{\pi}{N}$	(divide by $\sqrt{2}$ when $k=0$)
DCT-3:	$\cos j\left(k+\frac{1}{2}\right)\frac{\pi}{N}$	(divide by $\sqrt{2}$ when $j=0$)
DCT-4:	$\cos\left(j+\frac{1}{2}\right)\left(k+\frac{1}{2}\right)\frac{\pi}{N}$	

Those are the orthogonal columns of the four DCT matrices C_1, C_2, C_3, C_4. The matrix C_3 with top row $\frac{1}{\sqrt{2}}(1,1,\ldots,1)$ is the transpose of C_2. All columns of C_2, C_3, C_4 have length $\sqrt{N/2}$. The immediate goal is to prove orthogonality.

Proof! These four bases (including the rescaling by $\sqrt{2}$) are eigenvectors of *symmetric* second difference matrices. Thus each basis is orthogonal. We start with matrices A_1, A_2, A_3, A_4 in the form (1), whose eigenvectors are pure (unscaled) cosines. Then symmetrizing these matrices introduces the $\sqrt{2}$ scaling; the eigenvectors become orthogonal. Three of the matrices were studied in an unpublished manuscript [12] by David Zachmann, who wrote down the explicit eigenvectors. His paper is very useful. He noted earlier references for the eigenvalues; a complete history would be virtually impossible.

We have seen that A_0, the periodic matrix with -1, 2, -1 in every row, shares the same cosine and sine eigenvectors as the second derivative. The cosines are picked out by a zero-slope boundary condition in the first row.

3. Boundary Conditions at Meshpoints and Midpoints

There are two natural choices for the discrete analogue of $u'(0)=0$:

Symmetry around the meshpoint $j = 0$: $\quad u_{-1} = u_1$
Symmetry around the midpoint $j = -\frac{1}{2}$: $\quad u_{-1} = u_0$

The first is called *whole*-sample symmetry in signal processing, the second is *half*-sample. Symmetry around 0 extends $(u_0, u_1, \ldots)$ evenly across the left boundary to $(\ldots, u_1, u_0, u_1, \ldots)$. Midpoint symmetry extends the signal to $(\ldots, u_1, u_0, u_0, u_1, \ldots)$ with u_0 repeated. Those are the simplest reflections of a discrete vector. We substitute the two options for u_{-1} in the second difference $-u_1 + 2u_0 - u_{-1}$ that straddles the boundary:

Symmetry at meshpoint: $\quad u_{-1} = u_1$ yields $2u_0 - 2u_1$
Symmetry at midpoint: $\quad u_{-1} = u_0$ yields $u_0 - u_1$.

Those are the two possible top rows for the matrix A:

$$\text{meshpoint:} \quad \otimes \quad \otimes = 2 \quad -2 \qquad \text{and} \qquad \text{midpoint:} \quad \otimes \quad \otimes = 1 \quad -1\,.$$

At the other boundary there are the same choices in replacing $u'(\pi) = 0$. Substituting $u_N = u_{N-2}$ or $u_N = u_{N-1}$ in the second difference $-u_{N-2} + 2u_{N-1} - u_N$ gives the two forms for the Neumann condition in the last row of A:

$$\text{meshpoint:} \quad \boxtimes \quad \boxtimes = -2 \quad 2 \qquad \text{and} \qquad \text{midpoint:} \quad \boxtimes \quad \boxtimes = -1 \quad 1\,.$$

The alternative at the right boundary is the Dirichlet condition $u(\pi) = 0$. The meshpoint condition $u_N = 0$ removes the last term of $-u_{N-2} + 2u_{N-1} - u_N$. The midpoint condition $u_N + u_{N-1} = 0$ is simple too, but the resulting matrix will be a little surprising. The 2 turns into 3:

$$\text{meshpoint:} \quad \boxtimes \quad \boxtimes = -1 \quad 2 \qquad \text{and} \qquad \text{midpoint:} \quad \boxtimes \quad \boxtimes = -1 \quad 3\,.$$

Now we have $2 \times 4 = 8$ combinations. Four of them give the standard basis functions of cosines, listed above. Those are the DCT-1 to DCT-4, and they come when the centering is the same at the two boundaries: both meshpoint centered or both midpoint centered. Zachmann [12] makes the important observation that *all those boundary conditions give second-order accuracy around their center points.* Finite differences are one-sided and less accurate only with respect to the wrong center! We can quickly write down the matrices A_1 to A_4 that have these cosines as eigenvectors.

4. The Standard Cosine Transforms

Notice especially that the denominator in the cosines (which is $N-1$ or N) agrees with the distance between "centers." This distance is an integer, measuring from meshpoint to meshpoint or from midpoint to midpoint. We also give the diagonal matrix D that makes $D^{-1}AD$ symmetric and makes the eigenvectors orthogonal:

DCT-1

Centers $j = 0$ and $N-1$
Components $\cos jk\frac{\pi}{N-1}$
$D_1 = \operatorname{diag}\left(\sqrt{2}, 1, \ldots, 1, \sqrt{2}\right)$

$$A_1 = \begin{bmatrix} 2 & -2 & & & \\ -1 & 2 & -1 & & \\ & \cdot & \cdot & \cdot & \\ \cdot & & -1 & 2 & -1 \\ & & & -2 & 2 \end{bmatrix}$$

DCT-2

Centers $j = -\frac{1}{2}$ and $N - \frac{1}{2}$
Components $\cos\left(j + \frac{1}{2}\right) k\frac{\pi}{N}$
$D_2 = I$

$$A_2 = \begin{bmatrix} 1 & -1 & & & \\ -1 & 2 & -1 & & \\ & \cdot & \cdot & \cdot & \\ & & -1 & 2 & -1 \\ & & & -1 & 1 \end{bmatrix}$$

DCT-3

Centers $j = 0$ and N
Components $\cos j\left(k + \frac{1}{2}\right)\frac{\pi}{N}$
$D_3 = \operatorname{diag}(\sqrt{2}, 1, \ldots, 1)$

$$A_3 = \begin{bmatrix} 2 & -2 & & & \\ -1 & 2 & -1 & & \\ & \cdot & \cdot & \cdot & \\ & & -1 & 2 & -1 \\ & & & -1 & 2 \end{bmatrix}$$

DCT-4

Centers $j = -\frac{1}{2}$ and $N - \frac{1}{2}$
Components $\cos\left(j + \frac{1}{2}\right)\left(k + \frac{1}{2}\right)\frac{\pi}{N}$
$D_4 = I$

$$A_4 = \begin{bmatrix} 1 & -1 & & & \\ -1 & 2 & -1 & & \\ & \cdot & \cdot & \cdot & \\ & & -1 & 2 & -1 \\ & & & -1 & 3 \end{bmatrix}$$

Recently Sanchez et al [7] provided parametric forms for *all* matrices that have the DCT bases as their eigenvectors. These are generally full matrices of the form "Toeplitz plus near-Hankel." Particular tridiagonal matrices (not centered differences) were noticed by Kitajima, Rao, Hou, and Jain. We hope that the pattern of *second differences with different centerings* will bring all eight matrices into a common structure. Perhaps each matrix deserves a quick comment.

DCT-1: The similarity transformation $D_1^{-1} A_1 D_1$ yields a symmetric matrix. This multiplies the eigenvector matrix for A_1 by D_1^{-1}. (Notice that $A\mathbf{x} = \lambda\mathbf{x}$ leads to $(D^{-1}AD)D^{-1}\mathbf{x} = \lambda D^{-1}\mathbf{x}$.) The eigenvectors become orthogonal for both odd N and even N, when D_1^{-1} divides the first and last components by $\sqrt{2}$:

$$\begin{array}{llll} N=3 & \left(\frac{1}{\sqrt{2}}, 1, \frac{1}{\sqrt{2}}\right) \quad \left(\frac{1}{\sqrt{2}}, 0, -\frac{1}{\sqrt{2}}\right) \quad \left(\frac{1}{\sqrt{2}}, -1, \frac{1}{\sqrt{2}}\right) & \text{for} & k = 0,1,2 \\ N=4 & \left(\frac{1}{\sqrt{2}}, 1, 1, \frac{1}{\sqrt{2}}\right) \quad \dots \quad \left(\frac{1}{\sqrt{2}}, -1, 1, -\frac{1}{\sqrt{2}}\right) & \text{for} & k = 0,1,2,3\,. \end{array}$$

The first and last eigenvectors have length $\sqrt{N-1}$, the others have length $\sqrt{(N-1)/2}$.

DCT-2: These basis vectors $\cos\left(j + \frac{1}{2}\right)k\frac{\pi}{N}$ are the most popular of all, because $k = 0$ gives the flat vector $(1, 1, \ldots, 1)$. Their first and last components are not exceptional. The boundary condition $u_{-1} = u_0$ is a zero derivative centered on a *midpoint*. Similarly the right end has $u_N = u_{N-1}$. When those outside values are eliminated, the boundary rows of A_2 have the neat 1 and -1.

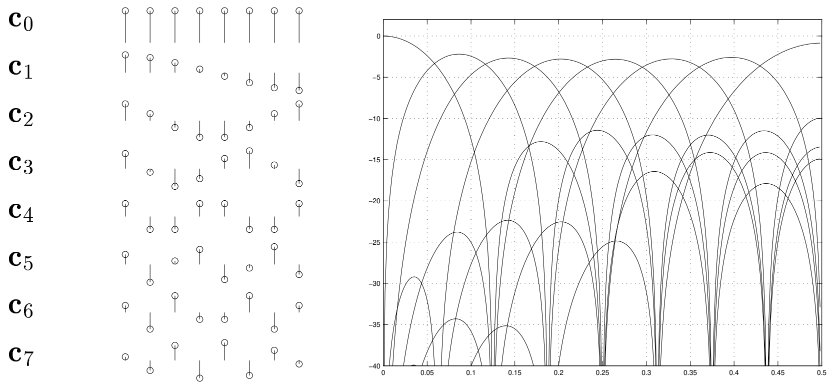

Figure 1: The eight DCT-2 vectors and their Fourier transforms (absolute values).

I believe that this DCT-2 (often just called the DCT) should be in applied mathematics courses along with the DFT. Figure 1 shows the eight basis vectors (when $N = 8$). On the right are the Fourier transforms of those vectors. Maybe you can see the first curve $|\Sigma e^{2\pi ij/8}|$ and especially its second lobe, rising to 13 decibels (which is $20 \log_{10} 13$) below the top. This is not a big dropoff! Like the closely connected Gibbs phenomenon, it does not improve as N increases. A good lowpass filter can drop by 40 or 50 db. The other seven transforms *vanish* at zero frequency (no leakage of the direct current DC term). Those seven vectors are orthogonal to $(1, 1, \ldots, 1)$.

This basis was chosen for the JPEG algorithm in image compression. Each 8×8 block in the image is transformed by a two-dimensional DCT. We comment below on the undesirable blocking artifacts that appear when the transform coefficients are compressed.

DCT-3: The vectors $\cos j\left(k + \frac{1}{2}\right)\frac{\pi}{N}$ are the discrete analogues of $\cos(k + \frac{1}{2})x$. The Neumann condition at the left and Dirichlet condition at the right are centered at meshpoints. For orthogonality we need the factor D_3^{-1} that divides the first components by $\sqrt{2}$. This basis loses to the DCT-4.

DCT-4: We had never seen the final entry "3" in the matrix A_4 but MATLAB insisted it was right. Now we realize that a zero boundary condition at a midpoint gives $u_N \approx -u_{N-1}$ (the extension is *antisymmetric*). Then $-1,\ 2,\ -1$ becomes $-1,\ 3$. The eigenvectors are *even* at the left end and *odd* at the right end. This attractive property leads to $j+\frac{1}{2}$ and $k+\frac{1}{2}$ and a symmetric eigenvector matrix C_4. Its applications to "lapped transforms" are described below.

Remember our proof of orthogonality! It is a verification that the cosine vectors are eigenvectors of $A_1,\ A_2,\ A_3,\ A_4$. For all the $-1,\ 2,\ -1$ rows this needs to be done only once (and it reveals the eigenvalues $\lambda = 2 - 2\cos\theta$). There is an irreducible minimum of trigonometry when the jth component of the kth vector $\mathbf{c}_k$ is $\cos j\theta$ in Types 1 and 3, and $\cos(j+\frac{1}{2})\theta$ in Types 2 and 4:

$$\begin{aligned} -\cos(j-1)\theta + 2\cos j\theta - \cos(j+1)\theta &= (2-2\cos\theta)\cos j\theta \\ -\cos\left(j-\frac{1}{2}\right)\theta + 2\cos\left(j+\frac{1}{2}\right)\theta - \cos\left(j+\frac{3}{2}\right)\theta &= (2-2\cos\theta)\cos\left(j+\frac{1}{2}\right)\theta\,. \end{aligned}$$

This is $A\mathbf{c}_k = \lambda_k \mathbf{c}_k$ on all interior rows. The angle is $\theta = k\frac{\pi}{N-1}$ for Type 1 and $\theta = k\frac{\pi}{N}$ for Type 2. It is $\theta = \left(k+\frac{1}{2}\right)\frac{\pi}{N}$ for A_3 and A_4. This leaves only the first and last components of $A\mathbf{c}_k = \lambda_k \mathbf{c}_k$ to be verified in each case.

Let us do only the fourth case, for the last row $-1,\ 3$ of the symmetric matrix A_4. A last row of $-1,\ 1$ would subtract the $j = N-2$ component from the $j = N-1$ component. Trigonometry gives those components as

$$\begin{aligned} j = N-1: &\qquad \cos\left(N-\frac{1}{2}\right)\left(k+\frac{1}{2}\right)\frac{\pi}{N} = \sin\frac{1}{2}\left(k+\frac{1}{2}\right)\frac{\pi}{N} \\ j = N-2: &\qquad \cos\left(N-\frac{3}{2}\right)\left(k+\frac{1}{2}\right)\frac{\pi}{N} = \sin\frac{3}{2}\left(k+\frac{1}{2}\right)\frac{\pi}{N}\,. \end{aligned}$$

We subtract using $\sin a - \sin b = -2\cos\left(\frac{b+a}{2}\right)\sin\left(\frac{b-a}{2}\right)$. The difference is

$$-2\cos\left(k+\frac{1}{2}\right)\frac{\pi}{N}\sin\frac{1}{2}\left(k+\frac{1}{2}\right)\frac{\pi}{N}. \tag{2}$$

The last row of A_4 actually ends with 3, so we still have 2 times the last component ($j = N-1$) to include with (2):

$$\left(2-2\cos\left(k+\frac{1}{2}\right)\frac{\pi}{N}\right)\sin\frac{1}{2}\left(k+\frac{1}{2}\right)\frac{\pi}{N}. \tag{3}$$

This is just λ_k times the last component of $\mathbf{c}_k$. The final row of $A_4\mathbf{c}_k = \lambda_k\mathbf{c}_k$ is verified.

There are also Discrete Sine Transforms DST-1 through DST-4. The entries of the basis vectors $\mathbf{s}_k$ are sines instead of cosines. These $\mathbf{s}_k$ are orthogonal because they are eigenvectors of symmetric second difference matrices, with a Dirichlet (instead of Neumann) condition at the left boundary. In writing about the applications to signal processing [9], we presented a third proof of orthogonality—which simultaneously covers the DCT and the DST, and shows their fast connection to the DFT matrix of order $2N$. This is achieved by a neat matrix factorization given by Wickerhauser [11]:

$$e^{-\pi i/4N} R^T F_{2N} R = \begin{bmatrix} C_4 & 0 \\ 0 & -iS_4 \end{bmatrix}.$$

The entries of S_4 are $\sin(j+\frac{1}{2})(k+\frac{1}{2})\frac{\pi}{N}$. The connection matrix R is very sparse, with $w = e^{\pi i/2N}$:

$$R = \frac{1}{\sqrt{2}}\begin{bmatrix} D & D \\ E & -E \end{bmatrix} \quad \text{with} \quad \begin{array}{rcl} D & = & \mathrm{diag}(1, \bar{w}, \ldots, \bar{w}^{N-1}) \\ E & = & \mathrm{antidiag}(w, w^2, \ldots, w^N). \end{array}$$

Since R^T and F_{2N} and R have orthogonal columns, so do C_4 and S_4.

5. Cosine Transforms with $N-\frac{1}{2}$ and $N+\frac{1}{2}$

There are four more combinations of the discrete boundary conditions. Every combination that produces a symmetric matrix will also produce (from the eigenvectors of that matrix) an orthogonal transform. But you will see $N-\frac{1}{2}$ and $N+\frac{1}{2}$ in the denominators of the cosines, because the distance between centers is no longer an integer. One center is a midpoint and the other is a meshpoint.

The transforms DCT-5 to DCT-8, when they are spoken of at all, are called "odd." They are denoted by DCT-IO to DCT-IVO in [5] and [7]. Three of the tridiagonal matrices (A_5, A_6, A_8) are quite familiar:

DCT-5

Centers $j=0$ and $N-\frac{1}{2}$
Components $\cos jk\frac{\pi}{N-\frac{1}{2}}$
$D_5=\operatorname{diag}(\sqrt{2},1,\ldots,1)$

$$A_5=\begin{bmatrix} 2 & -2 & & & \\ -1 & 2 & -1 & & \\ & \cdot & \cdot & \cdot & \\ & & -1 & 2 & -1 \\ & & & -1 & 1 \end{bmatrix}$$

DCT-6

Centers $j=-\frac{1}{2}$ and $N-1$
Components $\cos\left(j+\frac{1}{2}\right)k\frac{\pi}{N-\frac{1}{2}}$
$D_6=\operatorname{diag}(1,\ldots,1,\sqrt{2})$

$$A_6=\begin{bmatrix} 1 & -1 & & & \\ -1 & 2 & -1 & & \\ & \cdot & \cdot & \cdot & \\ & & -1 & 2 & -1 \\ & & & -2 & 2 \end{bmatrix}$$

DCT-7

Centers $j=0$ and $N-\frac{1}{2}$
Components $\cos j\left(k+\frac{1}{2}\right)\frac{\pi}{N-\frac{1}{2}}$
$D_7=\operatorname{diag}(\sqrt{2},1,\ldots,1)$

$$A_7=\begin{bmatrix} 2 & -2 & & & \\ -1 & 2 & -1 & & \\ & \cdot & \cdot & \cdot & \\ & & -1 & 2 & -1 \\ & & & -1 & 3 \end{bmatrix}$$

DCT-8

Centers $j=-\frac{1}{2}$ and N
Components $\cos\left(j+\frac{1}{2}\right)\left(k+\frac{1}{2}\right)\frac{\pi}{N+\frac{1}{2}}$
$D_8=I$

$$A_8=\begin{bmatrix} 1 & -1 & & & \\ -1 & 2 & -1 & & \\ & \cdot & \cdot & \cdot & \\ & & -1 & 2 & -1 \\ & & & -1 & 2 \end{bmatrix}.$$

We could study A_8 by reflection across the left boundary, to produce the pure Toeplitz -1, 2, -1 matrix (which is my favorite example in teaching). The eigenvectors become discrete sines on a double interval—*almost*. The length of the double interval is not $2N$, because the matrix from reflection has *odd order*. This leads to the new "period length" $N + \frac{1}{2}$ in the cosines.

Notice that A_5 has the boundary conditions (and eigenvector components) in reverse order from A_6. The first eigenvectors of A_5 and A_6 are $(1, 1, \ldots, 1)$, corresponding to $k = 0$ and $\lambda = 0$. This "flat vector" can represent a solid color or a fixed intensity by itself (this is terrific compression). The DCT-5 and DCT-6 have a coding gain that is completely comparable to the DCT-2.

So we think through the factors that come from $D_6 = \mathrm{diag}(1, \ldots, 1, \sqrt{2})$. The symmetrized $D_6^{-1} A_6 D_6$ has $-\sqrt{2}$ in the two lower right entries, where A_6 has -1 and -2. The last components of the eigenvectors are divided by $\sqrt{2}$; they are orthogonal but less beautiful. We implement the DCT-6 by keeping the matrix C_6 with pure cosine entries, and accounting for the correction factors by diagonal matrices:

$$\tfrac{4}{2N-1}\; C_6 \;\mathrm{diag}\left(\tfrac{1}{2}, 1, \ldots, 1\right)\; C_6^T \;\mathrm{diag}\left(1, \ldots, 1, \tfrac{1}{2}\right) = I. \tag{4}$$

The cosine vectors have squared length $\frac{2N-1}{4}$, except the all-ones vector that is adjusted by the first diagonal matrix. The last diagonal matrix corrects the Nth components as D_6 requires. The inverse of C_6 is not quite C_6^T (analysis is not quite the transpose of synthesis, as in an orthogonal transform) but the corrections have trivial cost. For $N = 2$ and $k = 1$, the matrix identity (4) involves $\cos \frac{1}{2} \frac{\pi}{3/2} = \frac{1}{2}$ and $\cos \frac{3}{2} \frac{\pi}{3/2} = -1$:

$$\frac{4}{3} \begin{bmatrix} 1 & \frac{1}{2} \\ 1 & -1 \end{bmatrix} \begin{bmatrix} \frac{1}{2} & \\ & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ \frac{1}{2} & -1 \end{bmatrix} \begin{bmatrix} 1 & \\ & \frac{1}{2} \end{bmatrix} = \begin{bmatrix} 1 & \\ & 1 \end{bmatrix}.$$

Malvar has added a further good suggestion: Orthogonalize the last $N-1$ basis vectors against the all-ones vector. Otherwise the DC component (which is usually largest) leaks into the other components. Thus we subtract from each $\mathbf{c}_k^6$ (with $k>0$) its projection onto the flat $\mathbf{c}_0^6$:

$$\tilde{\mathbf{c}}_k^6 = \mathbf{c}_k^6 - \frac{(-1)^k}{2N}(1,1,\ldots,1)\,. \tag{5}$$

The adjusted basis vectors are now the columns of $\widetilde{C}_6$, and (5) becomes

$$C_6 = \widetilde{C}_6 \begin{bmatrix} 1 & \frac{-1}{2N} & \frac{+1}{2N} & \cdots \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{bmatrix}.$$

This replacement in equation (4) also has trivial cost, and that identity becomes $\widetilde{C}_6\widetilde{C}_6^{-1} = I$. The coefficients in the cosine series for $\mathbf{x}$ are $\mathbf{y} = \widetilde{C}_6^{-1}\mathbf{x}$. Then $\mathbf{x}$ is reconstructed from $\widetilde{C}_6\mathbf{y}$ (possibly after compressing $\mathbf{y}$). You see how we search for a good basis...

Transforms 5 to 8 are not used in signal processing. The half-integer periods are a disadvantage, but reflection offers a possible way out. The reflected vectors have an integer "double period" and *they overlap.*

6. Convolution

The most important algebraic identity in signal processing is the *convolution rule.* A slightly awkward operation in the time domain (convolution, from a Toeplitz matrix or a circulant matrix) becomes beautifully simple in the frequency domain (just multiplication). This accounts for the absence of matrices in the leading textbooks on signal processing. The property of time invariance (delay of input simply delays the output) is always the starting point.

We can quickly describe the rules for doubly infinite convolution and cyclic convolution. A vector $\mathbf{h}$ of filter coefficients is convolved with a vector $\mathbf{x}$ of inputs. The output is $\mathbf{y} = \mathbf{h} * \mathbf{x}$ with no boundary and $\mathbf{y} = \mathbf{h} *_C \mathbf{x}$ in the cyclic (periodic) case:

$$y_n = \sum_{-\infty}^{\infty} h_k x_{n-k} \quad \text{or} \quad y_n = \sum_{k+\ell \equiv n (\operatorname{mod} N)} h_k x_\ell \,. \tag{6}$$

Those are matrix-vector multiplications $\mathbf{y} = H\mathbf{x}$. On the whole line ($n \in \mathbf{Z}$) the doubly infinite matrix H is Toeplitz; the number h_k goes down its kth diagonal. In the periodic case ($n \in \mathbf{Z}_N$) the matrix is a circulant; the kth diagonal continues with the same h_k onto the $(k-N)$th diagonal. The eigenvectors of these matrices are pure complex exponentials. So when we switch to the frequency domain, *the matrices are diagonalized.* The eigenvectors are the columns of a Fourier matrix, and $F^{-1}HF$ is diagonal. Convolution with $\mathbf{h}$ becomes multiplication by the eigenvalues $H(\omega)$ in the diagonal matrix:

$$\begin{array}{rclll} \left(\sum_{-\infty}^{\infty} h_k e^{-ik\omega}\right)\left(\sum_{-\infty}^{\infty} x_\ell e^{-i\ell\omega}\right) & = & \sum_{-\infty}^{\infty} y_n e^{-in\omega} & \text{is} & H(\omega)X(\omega) = Y(\omega) \qquad (7)_\infty \\ \left(\sum_0^{N-1} h_k w^k\right)\left(\sum_0^{N-1} x_\ell w^\ell\right) & = & \sum_0^{N-1} y_n w^n & \text{is} & H(w)X(w) = Y(w) \qquad (7)_N \end{array}$$

The infinite case (Discrete Time Fourier Transform) allows all frequencies $|\omega| \le \pi$. The cyclic case (DFT) allows the N roots of $w^N = 1$. The multiplications in (7) agree with the convolutions in (6) because $e^{-ikx}e^{-i\ell x} = e^{-i(k+\ell)x}$ and $w^k w^\ell = w^{k+\ell}$. The question is: *What convolution rule goes with the* DCT?

A complete answer was found by Martucci [5]. The finite vectors $\mathbf{h}$ and $\mathbf{x}$ are symmetrically extended to length $2N$ or $2N-1$, by reflection. Those are convolved in the ordinary cyclic way (so the double length DFT appears). Then the output is restricted to the original N components. This *symmetric convolution* $\mathbf{h} *_S \mathbf{x}$ corresponds in the transform domain to multiplication of the cosine series.

The awkward point, as the reader already knows, is that a symmetric reflection can match u_{-1} with u_0 or u_1. The centering can be whole sample or half sample at each boundary. The extension of $\mathbf{h}$ can be different from the extension of $\mathbf{x}$! This confirms again that discrete problems have an extra degree of complexity beyond continuous problems. (And we resist the temptation to compare combinatorics and linear algebra with calculus.)

In the continuous case, we are multiplying two cosine expansions. This corresponds to symmetric convolution of the coefficients in the expansions.

7. The DCT in Image Processing

Images are not infinite, and they are not periodic. The image has boundaries, and the left boundary seldom has anything to do with the right boundary. A periodic extension can be expected to have a discontinuity. That means a slow decay of Fourier coefficients and a Gibbs oscillation at the jump — the one place where Fourier has serious trouble! In the image domain this oscillation is seen as "ringing." The natural way to avoid this discontinuity is to *reflect* the image across the boundary. With cosine transforms, a double length periodic extension becomes continuous.

A two-dimensional image may have $(512)^2$ pixels. The gray level of the pixel at position (i, j) is given by an integer $x(i, j)$ (between 0 and 255, thus 8 bits per pixel). That long vector $\mathbf{x}$ can be filtered by $\mathbf{x} * \mathbf{h}$, first a row at a time (j fixed) and then by columns (using the 1D transforms of the rows). This is computationally and algebraically simplest: the 2D Toeplitz and circulant matrices are formed from 1D blocks.

Similarly the DCT-2 is applied to rows and then to columns; 2D is the tensor product of 1D with 1D. The JPEG compression algorithm (established by the Joint Photographic Experts Group) divides the image into 8×8 blocks of pixels. Each block produces 64 DCT-2 coefficients. Those 64-component vectors from the separate blocks are compressed by the *quantization* step that puts coefficients into a discrete set of bins. Only the bin numbers are transmitted. The receiver approximates the true cosine coefficient by the value at the middle of the bin (most numbers go into the zero bin). Figures 2–5 show the images that the receiver reconstructs at increasing compression ratios and decreasing bit rates:

1. The original image (1:1 compression, all 8 bits per pixel)

2. Medium compression (8:1, average 1 bit per pixel)

3. High compression (32:1, average $\frac{1}{4}$ bit per pixel)

4. Very high compression (128:1, average $\frac{1}{16}$ bit per pixel)

You see severe blocking of the image as the compression rate increases. In teleconferencing at a very low bit rate, you can scarcely recognize your friends. This JPEG standard for image processing is quick but certainly not great. The newer standards allow for other transforms, with overlapping between blocks. The improvement is greatest for high compression.

One personal comment on quantization: This more subtle and statistical form of round-off should have applications elsewhere in numerical analysis. Numbers are not simply rounded to fewer bits, regardless of size. Nor do we sort by size and keep only the largest (this is thresholding, when we *want* to lose part of the signal — it is the basic idea in denoising). The bit rate is controlled by the choice of bin sizes, and quantization is surprisingly cheap. Vector quantization, which puts vectors into multidimensional bins, is more expensive but in principle more efficient. This technology of coding is highly developed [3] and it must have more applications waiting to be discovered.

A major improvement for compression and image coding was Malvar's extension of the ordinary DCT to a *lapped transform*. Instead of dividing the image into completely separate blocks for compression, his basis vectors overlap two or more blocks. The overlapping has been easiest to develop for the DCT-4, using its even-odd boundary conditions—which the DCT-7 and DCT-8 share. Those conditions help to maintain orthogonality between the tail of one vector and the head of another. The basic construction starts with a symmetric lowpass filter of length $2N$. Its coefficients $p(0), \ldots p(2N-1)$ are modulated (shifted in frequency) by the DCT-4:

$$\textit{The } k\textit{th basis vector has } j\textit{th component } p(j)\cos\left[(k+\tfrac{1}{2})(j+\tfrac{N+1}{2})\tfrac{\pi}{N}\right].$$

There are N basis vectors of length $2N$, overlapping each block with the next block. The 1D transform matrix becomes block bidiagonal instead of block diagonal. It is still an orthogonal matrix [4, 9] provided $p^2(j) + p^2(j+N) = 1$ for each j. This is Malvar's *Modulated Lapped Transform*, which is heavily used by the Sony mini disc and Dolby AC-3. (It is included in the MPEG-4 standard for video.) We naturally wonder if this MLT basis is also the set of eigenvectors for an interesting symmetric matrix. Coifman and Meyer found the analogous construction [2] for continuous wavelets.

The success of any transform in image coding depends on a combination of properties—mathematical, computational, and *visual*. The relation to the human visual system is decided above all by experience. This article was devoted to the mathematical property of orthogonality (which helps the computations). There is no absolute restriction to second difference matrices, or to these very simple boundary conditions. We hope that the eigenvector approach will suggest more new transforms, and that one of them will be fast and visually attractive.

Web Links

JPEG	`http://www.jpeg.org/public/jpeglinks.htm`
DCT	`http://www.cis.ohio-state.edu/hypertext/faq/usenet/compression-faq/top.html` (includes source code)
Author	`http://www-math.mit.edu/~gs/`

References

[1] N. Ahmed, T. Natarajan, and K. R. Rao, *Discrete cosine transform*, IEEE Trans. on Computers, **C-23** (1974) 90–93.

[2] R. Coifman and Y. Meyer, *Remarques sur l'analyse de Fourier à fenêtre*, C. R. Acad. Sci. Paris **312** (1991) 259–261.

[3] N. J. Jayant and P. Noll, *Digital Coding of Waveforms*, Prentice-Hall (1984).

[4] H. S. Malvar, *Signal Processing with Lapped Transforms*, Artech House (1992).

[5] S. Martucci, *Symmetric convolution and the discrete sine and cosine transforms*, IEEE Trans. Signal Processing **42** (1994) 1038–1051.

[6] K. R. Rao and P. Yip, *Discrete Cosine Transforms*, Academic Press (1990).

[7] V. Sanchez, P. Garcia, A. Peinado, J. Segura, and A. Rubio, *Diagonalizing properties of the discrete cosine transforms*, IEEE Trans. Signal Processing **43** (1995) 2631–2641.

[8] G. Strang, *The search for a good basis*, in *Numerical Analysis 1997*, D. Griffiths, D. Higham, and A. Watson, eds., Pitman Lecture Notes, Addison Wesley Longman.

[9] G. Strang and T. Nguyen, *Wavelets and Filter Banks*, Wellesley-Cambridge Press (1996).

[10] Z. Wang and B. Hunt, *The discrete W-transform*, Appl. Math. Comput. **16** (1985) 19–48.

[11] M. V. Wickerhauser, *Adapted Wavelet Analysis from Theory to Software*, AK Peters (1994).

[12] D. Zachmann, *Eigenvalues and eigenvectors of finite difference matrices*, unpublished manuscript (1987), `http://` .