Finite Fourier Transform, Circulant Matrices,
and the Fast Fourier Transform

Roe Goodman

Supplementary Class Notes for Math 642:550 Linear Algebra and Applications
Rutgers, The State University of New Jersey
Revised September 11, 2015

1 Sampling and Aliasing

Suppose we have a function \( s(t) \) that measures the sound level at time \( t \) of an analog audio signal. We assume that \( s(t) \) is piecewise-continuous and of finite duration: \( s(t) = 0 \) when \( t \) is outside some interval \( a \leq t \leq b \). Make a change of variable \( x = (t-a)/(b-a) \) and set \( f(x) = s(t) \). Then \( 0 \leq x \leq 1 \) when \( a \leq t \leq b \), and \( f(x) \) is a piecewise continuous function of \( x \).

We convert \( f(x) \) into a digital signal \( y \in \mathbb{C}^N \) by sampling at the \( N \) equal-spaced values \( x_j = j/N \) for \( j = 0, 1, \ldots, N-1 \):

\[
y = \begin{bmatrix} y[0] \\ y[1] \\ \vdots \\ y[N-1] \end{bmatrix} \quad \text{where } y[j] = f(x_j) \text{ for } j = 0, 1, \ldots, N-1.
\] (1)

Here we are using the notation \( y[k] \) for the value of the digital signal \( y \) at discrete time \( k \) (note that the indexing of the components in \( y \) is different than the usual MATLAB indexing, which would go from 1 to \( N \)). We call \( N \) the sampling rate.

**Example 1.1.** Suppose the signal consists of a single oscillating wave:

\[
f(x) = e^{2\pi ikx} = \cos(2\pi kx) + i \sin(2\pi kx) \quad \text{for } 0 \leq x \leq 1,
\] (2)

where \( k \) is an integer and \( i = \sqrt{-1} \) (engineering notation uses \( j \) instead of \( i \) for \( \sqrt{-1} \)). If we define \( f(x) \) for all \( x \in \mathbb{R} \) by letting \( f(x+m) = f(x) \) for every integer \( m \), then the extended function satisfies \( f(x+T) = f(x) \) for all \( x \), with the period \( T = 1/k \) (so the frequency of the wave is \( \omega = 1/T = k \)). The sample values are

\[
f(x_j) = e^{2\pi ikj/N} = w^{kj},
\]

where \( w = e^{2\pi i/N} \) is a primitive \( N \)th root of unity. In this case the vector \( y \) of sampled values in (1) is

\[
E_k = \begin{bmatrix} 1 \\ w^k \\ w^{2k} \\ \vdots \\ w^{(N-1)k} \end{bmatrix}.
\] (3)
Notice that \( \mathbf{E}_{k+N} = \mathbf{E}_k \), since \( w^{j(k+N)} = w^{jk} w^j = w^{jk} (w^N)^j = w^{jk} \). Thus the functions \( e^{2\pi ikx} \) and \( e^{2\pi (k+N)x} \) yield the same vector in \( \mathbb{C}^N \) when sampled at rate \( N \). This phenomenon is called aliasing. On the other hand, if \( 0 < |k - p| < N \), then \( w^{k-p} \neq 1 \) and hence \( w^k \neq w^p \). Thus \( \mathbf{E}_k \neq \mathbf{E}_p \) in this case and the functions \( e^{2\pi ikx} \) and \( e^{2\pi ipx} \) give different vectors in \( \mathbb{C}^N \) when sampled at rate \( N \) (no aliasing).

**Proposition 1.2.** Let the vectors \( \mathbf{E}_k \in \mathbb{C}^N \) be defined by (3) with \( w = e^{2\pi j/N} \). These vectors have the following properties:

1. \( \langle \mathbf{E}_k, \mathbf{E}_k \rangle = N \).
2. \( \langle \mathbf{E}_p, \mathbf{E}_k \rangle = 0 \) if \( k \) and \( p \) are integers with \( 0 < |k-p| < N \).
3. For any integer \( p \) the set \( \{ \mathbf{E}_p, \mathbf{E}_{p+1}, \ldots, \mathbf{E}_{p+N-1} \} \) is an orthogonal basis for \( \mathbb{C}^N \).
4. If \( \mathbf{y} \in \mathbb{C}^N \) then for any integer \( p \)
   \[
   \mathbf{y} = d_p \mathbf{E}_p + d_{p+1} \mathbf{E}_{p+1} + \cdots + d_{p+N-1} \mathbf{E}_{p+N-1}, \quad \text{where} \quad d_k = \frac{1}{N} \langle \mathbf{E}_k, \mathbf{y} \rangle. \tag{4}
   \]

Here \( \langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{u}^\dagger \mathbf{v} \) is the standard inner product on \( \mathbb{C}^N \).

**Proof.** Since \( \overline{w} = w^{-1} \), the inner product between \( \mathbf{E}_p \) and \( \mathbf{E}_k \) is
   \[
   (\mathbf{E}_p)^\dagger \mathbf{E}_k = 1 + w^{k-p} + (w^{k-p})^2 + \cdots + (w^{k-p})^{N-1}. \tag{5}
   \]
Set \( W = w^{k-p} = \cos(2\pi (k-p)/N) + i \sin(2\pi (k-p)/N) \). Then the right side of (5) is
   \[
   U = 1 + W + W^2 + \cdots + W^{N-1}.
   \]
If \( k = p \), then \( W = 1 \) and \( U = N \), proving (1). If \( 0 < |k-p| < N \), then \( 0 < |2\pi (k-p)/N| < 2\pi \), so \( W \neq 1 \). But
   \[
   WU = W + W^2 + \cdots + W^{N-1} + W^N = U,
   \]
since \( W^N = (w^N)^k - p = 1 \). Hence \( (W - 1)U = 0 \), which forces \( U = 0 \). This proves (2). Now (3) follows from (1) and (2), since a set of nonzero mutually orthogonal vectors is always linearly independent. Formula (4) follows from (1) and (3).

**Example 1.3.** Suppose that that time \( t \) is measured in seconds and the signal is a superposition of a fundamental low-frequency oscillation \( 2 \sin(6\pi t) \) with frequency 3 Hertz (cycles per second) and maximum amplitude 2, together with a weaker high-frequency oscillation \( 0.5 \sin(18\pi t) \) with frequency 9 Hertz (which is three times the fundamental frequency) and maximum amplitude 0.5:
   \[
   s(t) = 2 \sin(6\pi t) + 0.5 \sin(18\pi t).
   \]
If we want to record the signal over the time interval \( 0 \leq t \leq 2 \), then we can rescale the interval to \( 0 \leq x \leq 1 \) by setting \( x = t/2 \) and
   \[
   f(x) = 2 \sin(12\pi x) + 0.5 \sin(36\pi x) = -i \left( e^{12\pi ix} - e^{-12\pi ix} + 0.25 e^{36\pi ix} - 0.25 e^{-36\pi ix} \right).
   \]
In the second equation we have used the formula \( \sin(z) = (e^{iz} - e^{-iz})/2i \). On the left side of Figure 1 we have plotted \( f(x) \) for \( 0 \leq x \leq 1 \): notice the modulation of the low-frequency wave by the high-frequency wave.

If we sample \( f(x) \) on the interval \( 0 \leq x \leq 1 \) at rate \( N \), we get the vector
   \[
   -i \left( \mathbf{E}_6 - \mathbf{E}_{-6} + 0.25 \mathbf{E}_{18} - 0.25 \mathbf{E}_{-18} \right) \in \mathbb{C}^N.
   \]
If $N > 18 - (-18) = 36$, then by Proposition 1.2 all the terms in this sum are mutually orthogonal vectors, and hence there is no aliasing. For example, if we sample at rate $N = 48$, we get the vector on the right side of Figure 1 (plotted as a stem graph). Notice that the modulation of the low-frequency wave by the high-frequency is evident.

If we use a smaller value of $N$, say $N = 24$, then $E_{18} = E_{18-24} = E_{-6}$ and $E_{-18} = E_{-18+24} = E_{6}$. Thus the the sampled vector becomes

$$\frac{3}{41} (E_{6} - E_{-6}) \in \mathbb{C}^{24}.$$  

This is the same vector that we would get by sampling the function $g(x) = 1.5 \sin(12\pi x)$ at this rate (see Figure 2, where we have plotted $g(x)$ as a dashed line). Thus the high frequency component in the signal disappears, while the amplitude of the low frequency component is increased.

In Example 1.3 the highest frequency was 18 cycles (when the time interval was rescaled to be $0 \leq x \leq 1$). Sampling without aliasing required a sample rate $N > 36$, larger than twice the highest frequency present in the signal. This illustrates the
Nyquist sampling criterion: If a digital sample is to reproduce frequencies up to some level \( \omega \) without aliasing, then the sampling rate \( N \) must be greater than \( 2\omega \).

For example, in audio recording the frequency of 20,000 Hertz (cycles per second) is taken as an upper limit of human hearing. The commercial compact disk (CD) digital recording standard with a sampling rate \( N = 44,100 \) Hertz, which replaced the long-playing (LP) analog recording technology in the 1980’s, satisfies the Nyquist criterion.

## 2 Fourier Matrix and Discrete Fourier Transform

Consider the expansion of \( y \in \mathbb{C}^N \) in (4) in terms of the orthogonal basis \( E_k \). If we take \( p = 0 \), then this formula can be written as

\[
y = \frac{1}{N} \left( d_0 E_0 + d_1 E_1 + \cdots + d_{N-1} E_{N-1} \right) \quad \text{where} \quad d_k = E_k^H y. \quad (6)
\]

This motivates the following definition (see Strang, §3.5).

**Definition 2.1.** The *Fourier matrix* \( F_N \) is the \( N \times N \) matrix columns are \( E_0, \ldots, E_{N-1} \). Thus the \((j,k)\) entry of \( F_N \) is \( \omega^{jk} \) for \( j,k = 1, \ldots, N \), where \( \omega = e^{2\pi i/N} \).

The matrix \( F_N \) is symmetric and the entries in the first column (or row) are all 1. The second column (or row) consists of the powers of \( \omega \) from 0 to \( N-1 \), the third column (or row) consists of the powers of \( \omega^2 \) from 0 to \( N-1 \), and so on.

For \( N = 2 \) the Fourier matrix is

\[
F_2 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}.
\]

since \( e^{2\pi i/2} = -1 \). For \( N = 4 \) we have \( \omega = e^{2\pi i/4} = i \) and \( \omega^{-1} = -i \). Hence the \( 4 \times 4 \) Fourier matrix is

\[
F_4 = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & i & i^2 & i^3 \\ 1 & i^2 & i^4 & i^6 \\ 1 & i^3 & i^6 & i^9 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & i & -1 & -i \\ 1 & -1 & 1 & -1 \\ 1 & -i & -1 & i \end{bmatrix}.
\]

**Definition 2.2.** Given \( y \in \mathbb{C}^N \), we set \( Y = F_N^H y \), and call \( Y \) the *discrete Fourier transform* (DFT) of \( y \).

Since the rows of \( F_N \) are \( E_0^H, E_1^H, \ldots, E_{N-1}^H \), the entries of \( Y \) are

\[
Y = \begin{bmatrix} E_0^H y \\ E_1^H y \\ \vdots \\ E_{N-1}^H y \end{bmatrix} = \begin{bmatrix} d_0 \\ d_1 \\ \vdots \\ d_{N-1} \end{bmatrix},
\]

where \( d_k = Y[k] \) is the same as in (6). Thus

\[
Y[k] = \sum_{j=0}^{N-1} \omega^{-jk} y[j]. \quad (10)
\]

Here we are using the indexing convention for the components of a vector as in formula (1).
In particular,
\[ F_N^H e_k = N e_{k+1} \quad \text{for } k = 0, 1, \ldots, N - 1, \]  
(11)
where \( e_1, \ldots, e_N \) are the standard basis vectors for \( \mathbb{C}^N \).
From (9) we also see that
\[ F_N y = d_0 e_0 + d_1 e_1 + \cdots + d_{N-1} e_{N-1} = N y \]  
(12)
for all vectors \( y \in \mathbb{C}^N \), where we have used (6) for the last equality.

**Theorem 2.3.** The inverse of the Fourier matrix is \( (1/N) F_N^H \). Furthermore, the normalized matrix \( (1/\sqrt{N}) F_N \) is unitary.

**Proof.** The first assertion follows from (12). By Proposition 1.2 the columns of \( (1/\sqrt{N}) F_N \) are an orthonormal set. This proves the second assertion. \( \blacksquare \)

**Corollary 2.4.**

(a) Let \( \{e_1, \ldots, e_N\} \) be the standard basis for \( \mathbb{C}^N \). Set \( u_j = (1/\sqrt{N}) F_N e_j \). Then \( \{u_1, \ldots, u_N\} \) is an orthonormal basis for \( \mathbb{C}^N \), called the Fourier basis.

(b) Let \( y \in \mathbb{C}^N \) and set \( Y = F_N y \). Then \( \|y\|^2 = \frac{1}{N} \|Y\|^2 \).

**Proof.** (a): Note that \( u_j \) is the \( j \)th column of the unitary matrix \( (1/\sqrt{N}) F_N \).
(b): Write \( y = (1/\sqrt{N}) F_N u \), where \( u = (1/\sqrt{N}) Y \). Since \( (1/\sqrt{N}) F_N \) is a unitary matrix, the vectors \( u \) and \( y \) have the same norm. But \( \|u\| = (1/\sqrt{N}) \|Y\| \). \( \blacksquare \)

**Remark 2.5.** If we think of the standard basis \( e_j \) as a sampled version of a signal, then this signal is localized in time, since only one component of \( e_j \) is nonzero. The discrete Fourier transform of \( e_j \) is the vector \( E_j \) with all entries nonzero. Thus the discrete Fourier transform removes the time localization. In the opposite direction, the vector \( E_k \) is the sample of a wave having only one frequency. This digital signal is completely spread out in time, since all the entries have absolute value one. The discrete Fourier transform of \( E_k \) is the standard basis vector \( e_{k+1} \), which has only one nonzero component and hence is localized in frequency.

**Example 2.6.** Suppose \( N = 4 \) and \( y = \begin{bmatrix} 1 \\ 2 \\ -1 \\ 0 \end{bmatrix} \). Then
\[ Y = F_4 y = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & i & -1 & -i \\ 1 & -1 & 1 & -1 \\ 1 & -i & -1 & i \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ -1 \\ 0 \end{bmatrix} = \begin{bmatrix} 2 \\ 2 + 2i \\ -2 \\ 2 - 2i \end{bmatrix}. \]
In this case \( \|y\|^2 = 1 + 2^2 + (-1)^2 = 6 \), while we have
\[ \frac{1}{4} \|Y\|^2 = \frac{1}{4}[2^2 + (2 - 2i)(2 + 2i) + (-2)^2 + (2 + 2i)(2 - 2i)] = \frac{1}{4}[4 + 8 + 4 + 8] = 6, \]
as predicted by Corollary 2.4. \( \blacksquare \)
Example 2.7. Suppose we sample the signal \( f(x) = 2\sin(12\pi x) + 0.5\sin(36\pi x) \) in Example 1.3 at a rate satisfying the Nyquist criterion, say \( N = 48 \). Since \(-6 \equiv 42 \mod (48)\) and \(-18 \equiv 30 \mod (48)\), the sampled vector is

\[
y = -i\left( E_6 + 0.25E_{18} - 0.25E_{30} - E_{42} \right) \in \mathbb{C}^{48}.
\]

By (11) we see that

\[
Y = -48i\left( e_7 + 0.25e_{19} - 0.25e_{31} - e_{43} \right) \in \mathbb{C}^{48}.
\] (13)

In Figure 3 we have plotted the signal vector \( y \) (which has real components \( y[k] \) for \( k = 0, \ldots, 47 \)) and the imaginary part of its discrete Fourier transform vector \( Y = F_{48}y \) (which has purely imaginary components \( Y[k] \) for \( k = 0, \ldots, 47 \)). The horizontal axis in the left graph represents time, whereas the horizontal axis in the right graph represents frequency. The vector \( y \) is spread out in time. By contrast, the only nonzero entries in \( Y \) are \( Y[6], Y[18], Y[30] \) and \( Y[42] \). Figure 3 shows the two frequencies present in the original signal, with the values \( Y[6], Y[42] \) corresponding to the low frequency sine wave and the values \( Y[18], Y[30] \) corresponding to the high frequency sine wave. ■

Remark 2.8. The graph in Figure 3 is skew-symmetric around the midpoint \( N/2 = 24 \) of the frequency range (this is called the Nyquist point). To understand this symmetry, assume that \( N \) is even. Then \( w^{N/2} = e^{i\pi} = -1 \) and \( w^{jN/2} = (-1)^j \). Hence by (10)

\[
Y[k + N/2] = \sum_{j=0}^{N-1} (-1)^j w^{-jk} y[j],
\]

\[
Y[-k + N/2] = \sum_{j=0}^{N-1} (-1)^j w^{jk} y[j],
\]

for any vector \( y \in \mathbb{C}^N \). If \( y \in \mathbb{R}^N \) is a vector with real components, then these formulas show that

\[
\overline{Y[k + N/2]} = Y[-k + N/2],
\]

where the bar denotes complex conjugation (since \( w^{-jk} = w^{jk} \)). Hence the graph of the real part of \( Y \) is symmetric about the Nyquist point \( k = N/2 \), whereas the graph of the imaginary part of \( Y \) is skew symmetric about the Nyquist point. For the real signal in Example 2.7, the real part of \( Y \) is zero, and we have explained the skew-symmetry of the graph in Figure 3.
3 Shift-Invariant Transformations and Circulant Matrices

Consider a finite digital signal $y$ with $N$ values, say $y[0], y[1], \ldots, y[N-1]$. As in Section 1, we can view $y$ either as a column vector
\[
\begin{bmatrix}
y[0] \\
y[1] \\
\vdots \\
y[N-1]
\end{bmatrix} \in \mathbb{C}^N
\]
(14)
or as a function of a discrete time variable. A basic operation in signal processing is to take a moving average of the signal. For example, we can replace each value $y[j]$ by the average of the values $y[j-1]$ and $y[j+1]$. This gives a new signal $z$ with
\[
z[j] = (y[j-1] + y[j+1])/2.
\]
(15)
There is a bug in formula (15), however. To calculate $z[0]$ or $z[N-1]$ we need the values $y[-1]$ and $y[N]$, which aren’t available. We will solve this problem by using the periodic extension of $y$:
\[
y[j + kN] = y[j] \quad \text{for } j = 0, 1, \ldots, N-1 \text{ and all integers } k.
\]
(16)
Thus we set $y[-1] = y[N-1]$ and $y[N] = y[0]$, since $-1 = N - 1 + N$ and $N = 0 + N$. In terms of modular arithmetic, we have $y[m] = y[j]$ when $m = j \pmod{N}$. With this adjustment formula (15) makes sense. It can be written in a case-by-case way as
\[
z[j] = \begin{cases} 
(y[N-1] + y[1])/2 & \text{for } j = 0 , \\
(y[j-1] + y[j+1])/2 & \text{for } j = 1, \ldots, N-2 , \\
(y[N-2] + y[0])/2 & \text{for } j = N-1 . 
\end{cases}
\]
For example, if $y = [1, 2, -1, 0]^T$ as in Example 2.6, then
\[
z[0] = (0 + 2)/2 , \quad z[1] = (1 - 1)/2 , \quad z[2] = (2 + 0)/2 , \quad z[3] = (-1 + 1)/2 .
\]
Define the shift operator $S$ on periodic signals $y$ of period $N$ by
\[
Sy[j] = y[j-1] \quad \text{for } j = 0, 1, \ldots, N-1 .
\]
Here $Sy[0] = y[N-1]$, since $y$ is periodic. It is clear from the definition that $S$ is linear and invertible:
\[
S^{-1}y[j] = y[j+1] .
\]
We can write formula (15) as
\[
z = \frac{1}{2}(S + S^{-1})y .
\]
(17)
It follows that formula (15) has the following properties:

(linearity) The output signal $z$ depends linearly on the input signal $y$.

(shift invariance) If the input signal $y$ is replaced by $Sy$, then the output signal $z$ is also replaced by $Sz$.

We now show that every shift-invariant linear transformation $C$ can be expressed as a linear combination of powers of the shift operator $S$. We first observe that the property of shift-invariance for $C$ is the same as the commutativity relation
\[
CS = SC .
\]
(Shift Invariance)
In particular, any linear combination of powers of $S$ is shift invariant, since $CS^k = S^kC$ for all integers $k \geq 0$. To prove the converse, we identify the periodic signals of period $N$ with $\mathbb{C}^N$ by (14). Then $S$ becomes
a linear transformation of $\mathbb{C}^N$. We calculate its matrix relative to the standard basis of $\mathbb{C}^N$ as follows: Suppose the signal $y$ corresponds to the standard basis vector $e_k$. Then $y[j] = 1$ if $j + 1 = k$, and otherwise $y[j] = 0$ (note the index shift by one). Since $Sy[j] = y[j-1]$, we see that $Sy[j] = 1$ if $j = k$ and $Sy[j] = 0$ if $j \neq k$. This shows that

$$Se_k = e_{k+1} \quad \text{for} \quad k = 1, 2, \ldots, N$$

(for this formula to be valid we must label the basis vectors circularly modulo $N$: $e_{N+1} = e_1$, $e_{N+2} = e_2$ and so on). We see that $S$ acts as a circular permutation of the standard basis vectors.

**Example 3.1.** Suppose $N = 3$. Then $Se_1 = e_2$, $Se_2 = e_3$, and $Se_3 = e_1$, so the matrix of the shift operator $S$ relative to the standard basis for $\mathbb{C}^3$ is

$$S = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}.$$  

Notice that $S^2e_1 = e_3$, $S^2e_2 = e_1$, and $S^2e_3 = e_2$. Also $S^3 = I$. Thus

$$S^{-1} = S^2 = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} = S^T.$$  

We have $S^{-1} = S^T$ since $\{Se_1, Se_2, Se_3\}$ is an orthonormal basis for $\mathbb{C}^3$.  

The general features of Example 3.1 are valid for the shift operator for any value of $N$. Namely, $S^N = I_N$ and $S^{-1} = S^{N-1}$. The matrix of $S$ relative to the standard basis for $\mathbb{C}^N$ is real and orthogonal, so in matrix form $S^{-1} = S^T$.

**Theorem 3.2.** Let $S$ be the shift operator, viewed as an $N \times N$ matrix relative to the standard basis for $\mathbb{C}^N$. Suppose $C$ is any shift-invariant linear transformation of $N$-periodic signals. View $C$ as an $N \times N$ matrix relative to the standard basis for $\mathbb{C}^N$ and let the first column of $C$ be $[c_0, c_1, \ldots, c_{N-1}]^T$. Then

$$C = c_0I + c_1S + c_2S^2 + \cdots + c_{N-1}S^{N-1},$$  

where $I$ denotes the $N \times N$ identity matrix.

**Proof.** The first column of $C$ is the vector $Ce_1$, so this vector can be written in terms of the standard basis as

$$Ce_1 = c_0e_1 + c_1e_2 + \cdots + c_{N-1}e_N.$$  

Now we calculate the columns $Ce_k$ of $C$ for $k = 2, \ldots, N$. Since $C$ is shift-invariant we have $S^{k-1}C = CS^{k-1}$. Thus if we multiply both sides of (19) by $S^{k-1}$ and use the property $S^{k-1}e_1 = e_k$, we obtain

$$Ce_k = C^{k-1}e_1 = S^{k-1}Ce_1 = c_0S^{k-1}e_1 + c_1S^{k-1}e_2 + c_2S^{k-1}e_3 + \cdots + c_{N-1}S^{k-1}e_N = c_0e_k + c_1Se_k + c_2S^2e_k + \cdots + c_{N-1}S^{N-1}e_k.$$  

This calculation shows that the $k$th column of the matrix $C$ is the same as the $k$th column of the matrix $c_0I + c_1S + c_2S^2 + \cdots + c_{N-1}S^{N-1}$ for $k = 1, \ldots, N$. Hence the matrices are the same.

**Example 3.3.** Suppose $N = 3$ and $C = c_0I + c_1S + c_2S^2$ is a $3 \times 3$ shift-invariant matrix. From Example 3.1 we have

$$C = c_0 \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} + c_1 \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} + c_2 \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} = \begin{bmatrix} c_0 & c_2 & c_1 \\ c_1 & c_0 & c_2 \\ c_2 & c_1 & c_0 \end{bmatrix}. $$
Hence the successive columns of $C$ are obtained by circular permutation of the first column. Matrices of this form are called *circulant matrices*. For example, when $N = 4$ the averaging operation from (15) is given by the circulant matrix

$$C = \frac{1}{2}(S + S^{-1}) = \frac{1}{2} \begin{bmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix}.$$  

We now obtain the connection between shift-invariant linear transformations and the Fourier matrix. Recall the vector

$$E_k = \begin{bmatrix} 1 \\ w^k \\ w^{2k} \\ \vdots \\ w^{(N-1)k} \end{bmatrix}$$

(where $w = e^{2\pi i/N}$ and $k = 0, 1, \ldots, N-1$) obtained in Section 1 by sampling the frequency-$k$ analogue signal $f(x) = e^{2\pi ikx}$ at $N$ equally-spaced times $x = 0, 1/N, \ldots, (N-1)/N$ in the interval $0 \leq x \leq 1$. Since $S$ shifts the entries in $E_k$ down one place, with the last entry moved to the top, we have

$$SE_k = \begin{bmatrix} w^{(N-1)k} \\ 1 \\ w^k \\ \vdots \\ w^{(N-2)k} \end{bmatrix} = w^{-k} \begin{bmatrix} w^{Nk} \\ w^k \\ \vdots \\ w^{(N-1)k} \end{bmatrix} = w^{-k} E_k. \quad \text{(20)}$$

This equation says that $E_k$ is an *eigenvector* for the matrix $S$ with *eigenvalue* $w^{-k}$.

Define a diagonal matrix with the $N$th roots of 1 on the diagonal, enumerated in *clockwise order* around the unit circle starting at 1:

$$D_N = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & w^{-1} & 0 & \cdots & 0 \\ 0 & 0 & w^{-2} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & w^{-N+1} \end{bmatrix} \quad \text{(21)}$$

(note that every $N$th root of 1 occurs exactly once, and the last diagonal entry is just $w$). The Fourier matrix has columns $E_0, E_1, \ldots, E_{N-1}$. Using equation (20) we find that

$$SF_N = [ SE_0 \ SE_1 \ \cdots \ SE_{N-1} ] = [ E_0 \ w^{-1}E_1 \ \cdots \ w^{-N+1}E_{N-1} ] = F_N D_N. \quad \text{(22)}$$

By Theorem 2.3 we know that $F_N F_N^H = NI_N$. Hence multiplying both sides of equation (22) on the right by $(1/N) F_N^H$, we obtain

$$S = (1/N)F_N D_N F_N^H = F_N D_N F_N^{-1}. \quad \text{(23)}$$

The last equation can also be written as $F_N^{-1}SF_N = D_N$. We can summarize these calculations as follows:
**Theorem 3.4.** The $N \times N$ shift matrix $S$ is diagonalized by the Fourier matrix $F_N$. The columns $E_0, E_1, \ldots, E_{N-1}$ of $F_N$ are eigenvectors of $S$. The corresponding eigenvalues are the $N$ complex numbers $w^{-j}$ for $j = 0, 1, \ldots, N - 1$ (the $N$th roots of unity), enumerated clockwise around the unit circle starting at 1.

Combining the last two theorems gives us the main result of this section:

**Theorem 3.5 (Diagonalization of Circulant Matrices).** Suppose that $C$ is an $N \times N$ shift-invariant (circulant) matrix. Write $C = c_0I + c_1S + c_2S^2 + \cdots + c_{N-1}S^{N-1}$ and define the polynomial $p(z) = c_0 + c_1z + c_2z^2 + \cdots + c_{N-1}z^{N-1}$. Then $E_k$ is an eigenvector for $C$ with eigenvalue $p(w^{-k})$ for $k = 0, 1, \ldots, N - 1$. Hence $C$ is diagonalized by the Fourier matrix:

$$F_N^{-1}CF_N = p(D_N) = \begin{bmatrix} p(1) & 0 & 0 & \cdots & 0 \\ 0 & p(w^{-1}) & 0 & \cdots & 0 \\ 0 & 0 & p(w^{-2}) & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & p(w^{-N+1}) \end{bmatrix}.$$  

**Proof.** Using equation (20) repeatedly, we get $S^jE_k = w^{-jk}E_k$ for all integers $j$. Hence

$$CE_k = c_0E_kI + c_1SE_k + c_2S^2E_k + \cdots + c_{N-1}S^{N-1}E_k$$

$$= (c_0 + c_1w^{-k} + c_2w^{-2k} + \cdots + c_{N-1}w^{-(N-1)k})E_k$$

$$= p(w^{-k})E_k.$$  

This equation says that $E_k$ is an eigenvector for the matrix $C$ with eigenvalue $p(w^{-k})$.

Equation (23) implies that $F_NSF_N^{-1} = D_N^k$ for all integers $k$. Since $C$ is a linear combination of the powers of $S$, it follows that $C$ satisfies the corresponding equation:

$$F_NCF_N^{-1} = c_0I + c_1D_N + c_2D_N^2 + \cdots + c_{N-1}D_N^{N-1}.$$  

The right side of this equation is $p(D_N)$. $lacksquare$

**Example 3.6.** Consider the $4 \times 4$ circulant matrix $C = (S + S^{-1})/2 = (S + S^3)/2$ from Example 3.3 (note that $S^{-1} = S^3$ since $S^4 = I$). Then $p(z) = (z + z^3)/2$. Since the fourth roots of 1 are $1, i, -1, -i$, the eigenvalues of $C$ are

$$p(1) = 1,$$

$$p(-1) = (-1 + (-1)^3)/2 = -1,$$

$$p(i) = (1 + (i)^3)/2 = 0,$$

and $C$ acts on the eigenvectors by $CE_0 = E_0$, $CE_1 = 0$, $CE_2 = -E_2$, $CE_3 = 0$. Using the Fourier matrix, we can write any vector as $y = d_0E_0 + d_1E_1 + d_2E_2 + d_3E_3$. Then $Cy = d_0E_0 - d_2E_2$. $lacksquare$

## 4 Downsampling and the Fast Fourier Transform

The effectiveness of the discrete Fourier transform (DFT) as a computational tool depends on a remarkable fast algorithm for calculating the matrix-vector product $F_Nv$ when $N = 2^k$ is a power of 2 (similar fast algorithms exist for every highly composite number $N$, such as $N = 2^k3^m$). The Fast Fourier Transform (FFT) algorithm is based on the observation that the conjugated Fourier matrix $F_{2n}^H$ that is used for the DFT can be written as product of a permutation matrix (which has no arithmetic computational cost) and a $2 \times 2$ block matrix, where the blocks are $F_n^H$ or a diagonal matrix multiplying $F_n^H$. 


Example 4.1. Consider \( n = 2 \). Recall that

\[
F_2^H = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}, \quad F_4^H = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -i & -1 & i \\ 1 & -1 & 1 & -1 \\ 1 & i & -1 & -i \end{bmatrix} = \begin{bmatrix} h_0 & h_1 & h_2 & h_3 \end{bmatrix}.
\]

Let \( y \in \mathbb{C}^4 \). By the definition of matrix-vector multiplication we can write

\[
\]
as a linear combination of the columns \( h_j \) of \( F_4^H \). Rearrange this sum according to the even and odd column indices:

\[
y[0] h_0 + y[2] h_2 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} y[0] \\ y[2] \end{bmatrix}, \quad y[1] h_1 + y[3] h_3 = \begin{bmatrix} 1 & 1 \\ -i & i \\ 1 & -1 \\ -i & -1 \end{bmatrix} \begin{bmatrix} y[1] \\ y[3] \end{bmatrix}. \tag{25}
\]

Define

\[
y_{\text{even}} = \begin{bmatrix} y[0] \\ y[2] \end{bmatrix}, \quad y_{\text{odd}} = \begin{bmatrix} y[1] \\ y[3] \end{bmatrix}, \quad \bar{D}_2 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.
\]

Then, using block multiplication of matrices, we can write the formulas (25) as

\[
y[0] h_0 + y[2] h_2 = F_2^H \begin{bmatrix} F_2^H y_{\text{even}} + \bar{D}_2 F_2^H y_{\text{odd}} \\ \bar{D}_2 y_{\text{even}} - D_2 F_2^H y_{\text{odd}} \end{bmatrix} \begin{bmatrix} I_2 & D_2 \\ I_2 & -D_2 \end{bmatrix} \begin{bmatrix} F_2^H \\ 0 \end{bmatrix} = F_4^H P_4 y.
\]

The splitting of \( y \) into even/odd vectors of half length can be accomplished by the permutation matrix

\[
P_4 = \begin{bmatrix} e_1 & e_3 & e_2 & e_4 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad P_4 y = \begin{bmatrix} y_{\text{even}} \\ y_{\text{odd}} \end{bmatrix}.
\]

We can write the calculations above in block form as

\[
F_4^H y = \begin{bmatrix} F_2^H y_{\text{even}} + \bar{D}_2 F_2^H y_{\text{odd}} \\ \bar{D}_2 y_{\text{even}} - D_2 F_2^H y_{\text{odd}} \end{bmatrix} \begin{bmatrix} I_2 & D_2 \\ I_2 & -D_2 \end{bmatrix} \begin{bmatrix} F_2^H \\ 0 \end{bmatrix} P_4 y.
\]

This shows that \( F_4^H \) has the factorization

\[
\begin{bmatrix} I_2 & D_2 \\ I_2 & -D_2 \end{bmatrix} \begin{bmatrix} F_2^H \\ 0 \end{bmatrix} \begin{bmatrix} F_2^H \\ 0 \end{bmatrix} P_4 \]

\( P_4 \) in terms of \( F_2^H \) and the diagonal matrix \( D_2 \).

The same splitting into even and odd components works for the DFT of any signal

\[
y = \begin{bmatrix} y[0] & y[1] & \ldots & y[2n-2] & y[2n-1] \end{bmatrix}^T
\]
of length \( N = 2n \). Let

\[
y_{\text{even}} = \begin{bmatrix} y[0] & y[2] & \ldots & y[2n-2] \end{bmatrix}^T, \quad y_{\text{odd}} = \begin{bmatrix} y[1] & y[3] & \ldots & y[2n-1] \end{bmatrix}^T.
\]

Here we are using the terms \emph{even} and \emph{odd} because we view \( y \) as a function on \( \{0, 1, \ldots, 2n-1\} \). Thus \( y_{\text{even}} \) uses the values of \( y \) at the even integers, while \( y_{\text{odd}} \) uses the values of \( y \) at the odd integers.\(^1\) The splitting of \( y \) into \( y_{\text{even}} \) and \( y_{\text{odd}} \) of half length is called \emph{downsampling}.\(^2\)

\(^1\)But in the MATLAB indexing convention the vector \( y_{\text{even}} \) contains components 1, 3, \ldots, 2n − 1 of the vector \( y \)
Write \( w = e^{2\pi i/N} = e^{\pi i/n} \) and \( z = w^2 = e^{2\pi i/n} \). Then

\[
F_{2n}^H y[j] = \sum_{k=0}^{n-1} w^{-jk} y[k]
\]

(split into even-odd)

\[
= \sum_{k=0}^{n-1} w^{-j(2k)} y[2k] + \sum_{k=0}^{n-1} w^{-j(2k+1)} y[2k+1]
= \sum_{k=0}^{n-1} z^{-jk} y_{\text{even}}[k] + w^{-j} \sum_{k=0}^{n-1} z^{-jk} y_{\text{odd}}[k]
\]

for \( j = 0, 1, 2, \ldots, 2n - 1 \). This shows that

\[
F_{2n}^H y[j] = F_n^H y_{\text{even}}[j] + w^{-j} F_n^H y_{\text{odd}}[j] \quad \text{for } j = 0, 1, \ldots, n - 1 .
\] (26)

Since \( w^n = 1 \) and \( z^n = 1 \), we have \( w^{-(n+j)} = -w^{-j} \) and \( z^{-(n+j)k} = z^{-jk} \). Furthermore, the functions \( F_n^H y_{\text{even}} \) and \( F_n^H y_{\text{odd}} \) are periodic of period \( n \). Thus

\[
F_{2n}^H y[n+j] = F_n^H y_{\text{even}}[j] - w^{-j} F_n^H y_{\text{odd}}[j] \quad \text{for } j = 0, 1, \ldots, n - 1 .
\] (27)

We now express these formulas in concise block-matrix form, just as in the case \( N = 4 \) that we worked out in Example 4.1. Define the \( m \times m \) diagonal matrix

\[
D_m = \begin{bmatrix}
1 & 0 & 0 & \cdots & 0 \\
0 & w^{-1} & 0 & \cdots & 0 \\
0 & 0 & w^{-2} & \cdots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \cdots & w^{-(m-1)}
\end{bmatrix}
\]

(caution: \( w^m = -1 \)). (28)

Note that the diagonal of \( D_m \) only contains half of the \( 2m \)th roots of 1. Let \( P_{2m} \) be the permutation matrix that splits \( y \) into its even and odd components:

\[
P_{2m} = \begin{bmatrix}
\mathbf{e}_1 & \mathbf{e}_3 & \ldots & \mathbf{e}_{2m-1} & \mathbf{e}_2 & \mathbf{e}_4 & \ldots & \mathbf{e}_{2m}
\end{bmatrix}^T, \quad P_{2m} y = \begin{bmatrix}
y_{\text{even}} \\
y_{\text{odd}}
\end{bmatrix}.
\]

Since \( P_{2m} \) only rearranges the components of a vector, calculating \( P_{2m} y \) is free of computational cost. Then, just as in the case \( N = 4 \), the equations (26) and (27) for the components of \( F_{2m} y \) can be written as a single vector equation for \( Y = F_{2m}^H y \):

\[
Y = \begin{bmatrix}
F_m^H y_{\text{even}} + D_m F_m^H y_{\text{odd}} \\
F_m^H y_{\text{even}} - D_m F_m^H y_{\text{odd}}
\end{bmatrix} = G_{2m} \begin{bmatrix}
F_m^H \\
0
\end{bmatrix} P_{2m} y,
\]

where \( G_{2m} = \begin{bmatrix}
I_m & D_m \\
I_m & -D_m
\end{bmatrix} \). The key point in the matrix factorization (29) is that because \( D_m \) is a diagonal matrix, applying the matrix \( G_{2m} \) to a vector only requires \( m \) scalar multiplications (on the last \( m \) components of a vector with \( 2m \) components) followed by \( 2m \) scalar additions, rather than the \((2m)^2\) scalar multiplications and \(2m(2m-1)\) scalar additions that are necessary for a general \( 2m \times 2m \) matrix. The fast Fourier transform (FFT) algorithm calculates \( F_N y \) when \( N \) is a power of 2 by iterating formula (29).

---

2Multiplying a row vector and column vector, each with \( 2m \) components, requires \( 2m \) scalar multiplications followed by \( 2m - 1 \) scalar additions.
Example 4.2. Take $N = 1024 = 2^{10}$. Then by the factorization (29)

$$F_{1024}^H y = G_{1024} \begin{bmatrix} F_{512}^H & 0 \\ 0 & F_{512}^H \end{bmatrix} P_{1024} y.$$ 

The product with $G_{1024}$ needs 512 scalar multiplications. We can use the factorization (29) again, but now with $m = 256$, to express each copy of $F_{512}^H$ in terms of $G_{512}$ and $F_{256}^H$ applied to signals with 256 components. But now we have two copies of $G_{512}$, so we need $2 \cdot 256 = 512$ more scalar multiplications. At the next stage, there are four copies of $G_{256}$, requiring $4 \cdot 128 = 512$ more scalar multiplications. Thus at each stage of the factorization, the number of scalar multiplications remains $512 = 2^9$, and there are 10 stages to get down to $F_1$. So the total scalar multiplication count to calculate $F_{1024}^H y$ by this factorization method is $10 \cdot 2^9$. By contrast, direct evaluation of $F_{1024}^H y$ as a matrix-vector product requires $2^{20}$ scalar multiplications, so the FFT method gives a speedup for multiplications by a factor of $2^{11}/10 = 204.8$.

In general, the factorization argument given in Example 4.2 shows that to calculate $F_N y$ by the FFT matrix factorization method when $N = 2^k$ needs at most $k2^{k-1} = \frac{1}{2}N \log_2 N$ scalar multiplications. The speedup compared to the $N^2$ scalar multiplications needed in a direct matrix-vector product is by a factor of $2N/\log_2 N$. For example, when $N = 2^{20}$ this speedup factor is more than 100,000. The same sort of counting of the number of scalar addition operations needed in the FFT gives an upper bound of $N \log_2 N$. Thus the total number of scalar arithmetic operations in the FFT algorithm is bounded by $(3/2)N \log_2 N$, yielding a comparable speedup factor. Without the FFT algorithm digital signal processing would be impractical.

Exercises

1. Let $S = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$ be the matrix for the shift operator relative to the standard basis for $C^3$. Suppose the matrix $C = \begin{bmatrix} 4 & * & * \\ 7 & * & * \\ 5 & * & * \end{bmatrix}$ satisfies $CS = SC$.

(a) Write $C$ as a polynomial in $S$ and fill in the entries $*$ in $C$.

(b) Let $F$ be the $3 \times 3$ Fourier matrix, and let $w = e^{2\pi i/3}$. Find complex numbers $\lambda_0$, $\lambda_1$, and $\lambda_2$ so that $F^{-1}CF = \begin{bmatrix} \lambda_0 & 0 & 0 \\ 0 & \lambda_1 & 0 \\ 0 & 0 & \lambda_2 \end{bmatrix}$. Express your answer in terms of $w$ and $w^2$ (no complex arithmetic is needed).

2. Let $C$ be an $N \times N$ circulant matrix.

(a) Show that $C^H$ is also a circulant matrix.

(b) If the first column of $C$ is $[c_0 \ c_1 \ c_2 \ \ldots \ c_{N-1}]^T$, show that the first column of $C^H$ is $[\bar{c}_0 \ \bar{c}_{N-1} \ \ldots \ \bar{c}_2 \ \bar{c}_1]^T$.

3. Let $n = 2^k$. Define $\psi(k)$ to be the number of scalar additions needed to calculate $F_n^H c$ by the Fast Fourier Transform (FFT) algorithm (ignore the computational cost of sign changes and sorting a vector
into even and odd components). Note that the product of a row vector and a column vector, each with \( n \) components, needs \( n - 1 \) additions.

(a) Show that \( \psi(1) = 2 \) and that \( \psi(k + 1) = 2\psi(k) + 2(2^k - 1) \).

(b) Use the recursion in (a) to calculate \( \psi(k) \) for \( k = 2, 3, 4 \).

(c) Prove by induction that \( \psi(k) \leq k2^k \) for all positive integers \( k \).

(d) Use the result in Section 4 and (c) to show that the total number of scalar arithmetic operations (multiplications and additions) required for the FFT on vectors of size \( 2^k \) is less than \( (3/2)k2^k \). (Ignore the operations of changing sign and sorting into even and odd components.)