Harmonic Analysis: from Fourier to Haar

María Cristina Pereyra

Lesley A. Ward

Department of Mathematics and Statistics, MSC03 2150, 1 University of New Mexico, Albuquerque, NM 87131-0001, USA
E-mail address: crisp@math.unm.edu

School of Mathematics and Statistics, University of South Australia, Mawson Lakes SA 5095, Australia
E-mail address: Lesley.Ward@unisa.edu.au

2000 *Mathematics Subject Classification.* Primary 42-xx

CHAPTER 6

A whirlwind tour of discrete Fourier and Haar analysis

In this chapter we capitalize on the knowledge acquired while studying Fourier series to develop the simpler Fourier theory in a finite-dimensional setting.

In this context, the *discrete trigonometric basis* for $\mathbb{C}^N$ is a certain set of N orthonormal vectors, which necessarily form an orthonormal basis. The *discrete Fourier transform* is the linear transformation that, for each vector v in $\mathbb{C}^n$, gives us the coefficients of v in the trigonometric basis. Many of the features of the Fourier series theory are still present, but without the nuisances and challenges of being in an infinite-dimensional setting ($L^2(\mathbb{T})$), where one has to deal with integration and infinite sums. The tools required in the finite-dimensional setting are the tools of linear algebra.

For practical purposes, this finite theory is what is needed. Computers deal with finite vectors. The Discrete Fourier Transform (DFT) of a given vector v in $\mathbb{C}^N$ is calculated by applying a certain invertible $N \times N$ matrix to v, and the Discrete Inverse Fourier Transform by applying the inverse matrix to the transformed vector. Surprisingly, one can perform each of these matrix multiplications in order $N \log_2 N$ operations, as opposed to the expected N^2 operations, by using the celebrated *Fast Fourier Transform* (FFT) algorithm; see Section 6.5. This gain in the number of operations is of invaluable importance for numerical applications, especially when dealing with large problems.

In this chapter we introduce another orthonormal basis, the *discrete Haar basis* for $\mathbb{C}^n$, in preparation for later chapters where the Haar bases for $L^2(\mathbb{R})$ and for $L^2([0,1])$ appear. We highlight the similarities of the discrete trigonometric basis and the discrete Haar basis, and their differences. There is also a *Fast Haar Transform* (FHT), of order N operations, described in Section 6.8. The fast Haar transform is a precursor of the *Fast Wavelet Transform* (FWT), studied in more depth in later chapters.

6.1. Fourier series: a summary

We begin by summarizing the ingredients of the theory of Fourier series, as we have developed it in Chapters 1–5.

(1) To each integrable *function* $f : \mathbb{T} \to \mathbb{C}$ we associate a *sequence* $\{a_n\} = \{\widehat{f}(n)\}_{n\in\mathbb{Z}}$ of complex numbers. These numbers are known as the *Fourier coefficients* of f.

(2) The Fourier coefficients are given by the formula

$$a_n = \widehat{f}(n) := \frac{1}{2\pi}\int_{-\pi}^{\pi} f(\theta)e^{-in\theta}\,d\theta, \qquad \text{for } n \in \mathbb{Z}.$$

(3) The *building blocks* of the Fourier series are the *trigonometric functions* $e_n : \mathbb{T} \to \mathbb{C}$ given by

$$e_n(\theta) := e^{in\theta}, \qquad \text{where } n \in \mathbb{Z} \text{ and } \theta \in \mathbb{T}.$$

They are called *trigonometric* rather than *exponential* to emphasize the connection with cosines and sines given by Euler's formula $e^{in\theta} = \cos n\theta + i \sin n\theta$.

Thus the n^{th} Fourier coefficient of f can be written in terms of the usual complex L^2 inner product as

$$\widehat{f}(n) = \frac{1}{2\pi} \int_{-\infty}^{\infty} f(\theta)\overline{e_n(\theta)}\, d\theta =: \langle f, e_n \rangle.$$

(4) The Fourier series associated to f is the doubly infinite series formed by summing the products of the n^{th} Fourier coefficients a_n with the n^{th} trigonometric function $e^{in\theta}$:

$$f(\theta) \sim \sum_{n=-\infty}^{\infty} \widehat{f}(n) e^{in\theta}.$$

We use $\sim$ rather than $=$ in this formula in order to emphasize that for a given point θ, the Fourier series may not sum to the value $f(\theta)$, and indeed may not converge at all.

(5) The set $\{e^{in\theta}\}_{n\in\mathbb{Z}}$ of trigonometric functions forms an *orthonormal basis* for the vector space $L^2(\mathbb{T})$ of square-integrable functions on $\mathbb{T}$. Here *orthonormal* means with respect to the $L^2(\mathbb{T})$ inner product. In particular this means that f equals its Fourier series in $L^2(\mathbb{T})$, or equivalently that the partial Fourier sums $S_N f$ converge in $L^2(\mathbb{T})$ to f:

$$\|f - S_N f\|_{L^2(\mathbb{T})} = \|f - \sum_{|n| \le N} \widehat{f}(n) e^{in\theta}\|_{L^2(\mathbb{T})} \to 0, \quad N \to \infty.$$

Also, the *energy* of an $L^2(\mathbb{T})$ function is preserved by the discrete Fourier transform, in the sense that Parseval's Identity holds:

$$\|f\|^2_{L^2(\mathbb{T})} = \sum_{n=-\infty}^{\infty} |\widehat{f}(n)|^2.$$

In the *Fourier series setting* we have just described, the signal is a *function*, whose domain is the continuously varying quantity $\theta \in \mathbb{T}$. By contrast, in the *discrete setting* of finite Fourier analysis, the signal is a *vector* in $\mathbb{C}^N$, which can be thought of as a function whose domain is the numbers n in the discrete set $\{0, 1, \ldots, N-1\}$.

Physically, the quantities $\theta \in \mathbb{T}$ and $n \in \{0, 1, \ldots, N-1\}$ often represent times, or spatial locations.

In the Fourier series setting, the Fourier transform produces the doubly-infinite sequence of Fourier coefficients. In the discrete setting, the discrete Fourier transform produces another vector in $\mathbb{C}^N$.

Looking ahead, in Chapters 7 and 8 we will develop the theory of the *continuous* or *Fourier transform setting*, where the signal is a function whose domain is the continuously varying quantity $x \in \mathbb{R}$, and the Fourier transform produces another function on $\mathbb{R}$.

6.2. The discrete Fourier basis

We present the analogues in the discrete setting of the five ingredients of Fourier analysis listed above. Fix a positive integer N, and let $\omega := e^{2\pi i/N}$. Recall that $\mathbb{Z}/(N\mathbb{Z}) := \{0, 1, \ldots, N-1\}$. As usual, the superscript t indicates the transpose, so for example $v = [z_0, \ldots, z_{N-1}]^t$ is a column vector.

(1) To each *vector* $v = [z_0, \ldots, z_{N-1}]^t \in \mathbb{C}^N$ we associate a second *vector* $\widehat{v} = [a_0, \ldots, a_{N-1}]^t \in \mathbb{C}^N$. The vector $\widehat{v}$ is known as the *discrete Fourier transform* of v.

(2) The entries of the discrete Fourier transform are given by the formula

$$a_n = \widehat{v}(n) := \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} z_k \omega^{-kn}, \qquad \text{for } n \in \{0, 1, \ldots, N-1\}.$$

(3) The *building blocks* of the discrete Fourier transform are the N discrete trigonometric functions $e_n : \mathbb{Z}/(N\mathbb{Z}) \to \mathbb{Z}/(N\mathbb{Z})$ given by

$$e_n(k) := \frac{1}{\sqrt{N}} \omega^{kn}, \qquad k \in \{0, 1, \ldots, N-1\},$$

for $n \in \{0, 1, \ldots, N-1\}$. Thus the n^{th} Fourier coefficient of v can be written in terms of the $\mathbb{C}^N$ inner product as

$$a_n = \widehat{v}\,(n) = \sum_{k=0}^{N-1} z_k \overline{e_n(k)} =: \langle v, e_n \rangle.$$

(4) The original vector v can be exactly determined from its discrete Fourier transform. The k^{th} entry of v is given by

$$v(k) := z_k = \sum_{n=0}^{N-1} \widehat{v}\,(n) e_n(k).$$

Hence

$$v = \sum_{n=0}^{N-1} \langle v, e_n \rangle e_n.$$

(5) The set $\{e_0, \ldots, e_{N-1}\}$ forms an *orthonormal basis* for the vector space $\mathbb{C}^N$. Here *orthonormal* means with respect to the standard inner product or *dot product* in $\mathbb{C}^N$. Hence the Pythagorean theorem holds:

$$\|v\|^2_{\ell^2(\mathbb{Z}/(N\mathbb{Z}))} = \|v\|^2_{\mathbb{C}^n} = \sum_{n=0}^{N-1} |\widehat{v}\,(n)|^2.$$

In this sense the energy of a vector v in $\mathbb{C}^n$ is the same as the energy of its discrete Fourier transform $\widehat{v}$. Here we have referred to the Hilbert space $\mathbb{C}^N$ with the standard inner product as the space $\ell^2\big(\mathbb{Z}/(N\mathbb{Z})\big)$, to emphasize the similarities with the Hilbert space $L^2(\mathbb{T})$.

Note that as the dimension changes so do the number ω, and the vectors e_n. We should label them with the dimension N, for example $\omega = \omega_N = e^{2\pi i/N}$, $e_n^N(k) = \frac{1}{\sqrt{N}} \omega_N^{nk}$. However we will omit the dependence on N and simply write ω, e_n, and expect the reader to be aware of the underlying dimension.

We give two examples showing the Fourier orthonormal bases obtained for $\mathbb{C}^5$ and $\mathbb{C}^8$ using the scheme described above.

EXAMPLE 6.1. (*Fourier Orthonormal Basis for* $\mathbb{C}^5$) Let $N = 5$, so that

$$\omega = e^{2\pi i/5}.$$

Let

$$f_0 := \begin{bmatrix}1\\1\\1\\1\\1\end{bmatrix}, \quad f_1 := \begin{bmatrix}1\\\omega\\\omega^2\\\omega^3\\\omega^4\end{bmatrix}, \quad f_2 := \begin{bmatrix}1\\\omega^2\\\omega^4\\\omega\\\omega^3\end{bmatrix}, \quad f_3 := \begin{bmatrix}1\\\omega^3\\\omega\\\omega^4\\\omega^2\end{bmatrix}, \quad f_4 := \begin{bmatrix}1\\\omega^4\\\omega^3\\\omega^2\\\omega\end{bmatrix}.$$

Notice the interesting arrangement of the powers of ω in each vector f_l.

The vectors f_0, f_1, f_2, f_3, f_4 are pairwise orthogonal. Dividing by the length $\sqrt{5}$ of each f_l, we obtain an orthonormal basis for $\mathbb{C}^5$, namely $e_o = (1/\sqrt{5})f_0$, $e_1 = (1/\sqrt{5})f_1$, $e_2 = (1/\sqrt{5})f_2$, $e_3 = (1/\sqrt{5})f_3$, $e_4 = (1/\sqrt{5})f_4$. ♢

EXAMPLE 6.2. (*Fourier Orthonormal Basis for* $\mathbb{C}^8$) Let $N = 8$, so that

$$\omega = e^{2\pi i/8} = e^{\pi i/4} = \frac{(1+i)}{\sqrt{2}}.$$

We have pairwise orthogonal vectors f_0, f_1, $\dots$, f_7 given by, respectively,

$$\begin{bmatrix}1\\1\\1\\1\\1\\1\\1\\1\end{bmatrix}, \begin{bmatrix}1\\\omega\\\omega^2\\\omega^3\\\omega^4\\\omega^5\\\omega^6\\\omega^7\end{bmatrix}, \begin{bmatrix}1\\\omega^2\\\omega^4\\\omega^6\\1\\\omega^2\\\omega^4\\\omega^6\end{bmatrix}, \begin{bmatrix}1\\\omega^3\\\omega^6\\\omega\\\omega^4\\\omega^7\\\omega^2\\\omega^5\end{bmatrix}, \begin{bmatrix}1\\\omega^4\\1\\\omega^4\\1\\\omega^4\\1\\\omega^4\end{bmatrix}, \begin{bmatrix}1\\\omega^5\\\omega^2\\\omega^7\\\omega^4\\\omega\\\omega^6\\\omega^3\end{bmatrix}, \begin{bmatrix}1\\\omega^6\\\omega^4\\\omega^2\\1\\\omega^6\\\omega^4\\\omega^2\end{bmatrix}, \begin{bmatrix}1\\\omega^7\\\omega^6\\\omega^5\\\omega^4\\\omega^3\\\omega^2\\\omega\end{bmatrix}.$$

Notice the patterns that arise because N is not prime.

Dividing by the length $\sqrt{8}$ of each f_l, we obtain an orthonormal basis for $\mathbb{C}^8$, namely $e_0 = (1/\sqrt{8})f_0$, $e_1 = (1/\sqrt{8})f_1$, $\dots$, $e_7 = (1/\sqrt{8})f_7$. ♢

EXERCISE 6.3. Verify that the set $\{e_0, e_1, \dots, e_{N-1}\}$ is an orthonormal set in $\mathbb{C}^N$. ♢

We make four remarks. First, the discrete setting has a certain symmetry in that both the original signal and the transformed signal are of the same form: vectors in $\mathbb{C}^N$. In the Fourier series setting, the original signal is a square-integrable function $f : \mathbb{T} \to \mathbb{C}$, while the transformed signal is a doubly infinite sequence $\{a_n\} \in l^2(\mathbb{Z})$. In Chapters 7 and 8 we will develop the Fourier theory in yet another setting, that of the *Fourier transform*. There the original signal and the transformed signal can again be of the same form. For instance, they can both be functions in the Schwartz class $\mathcal{S}(\mathbb{R})$, or they can both be functions in $L^2(\mathbb{R})$. We will also discuss generalizations of the Fourier transform to much larger classes of input signals, such as functions in $L^p(\mathbb{R})$, and, more generally, tempered distributions.

Second, our signal vectors $v = [z_0, \dots, z_{N-1}]^t \in \mathbb{C}^N$ in the discrete setting can be thought of as functions $v : \mathbb{Z}/(N\mathbb{Z}) \to \mathbb{C}$, in the sense that for each number $n \in \mathbb{Z}/(N\mathbb{Z}) := \{0, 1, \dots, N-1\}$, the function v picks out a complex number $v(n) = z_n$. So

$$v = [v(0), v(1), \dots, v(N-1)]^t.$$

This formulation of the input signal is closer to the way we presented the Fourier series setting. With this notation the discrete Fourier transform reads

$$\widehat{v}(m) = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} v(n) e^{-2\pi i n m/N}. \tag{6.1}$$

Third, in the discrete setting there are no difficulties with convergence of series, since all the sums are finite. Thus the subtleties discussed in Chapters 3, 4 and 5 are absent from the setting of the finite Fourier transform.

Fourth, in the discrete setting the identity known as *Parseval's Identity* or *Plancherel's Formula* holds:

$$\sum_{n=0}^{N-1} |v(n)|^2 = \sum_{n=0}^{N-1} |\widehat{v}(n)|^2.$$

This identity is exactly the Pythagorean Theorem for vectors with N entries. Rewritten in a form closer to the way we saw it in the Fourier series setting, Parseval's Identity reads

$$\|v\|_{\ell^2(\mathbb{Z}/N\mathbb{Z})} = \|\widehat{v}\|_{\ell^2(\mathbb{Z}/N\mathbb{Z})}.$$

Here the norm, known as the ℓ^2-norm, is the Euclidean norm induced by the complex inner product. Explicitly, it is given by the square root of the sum of the squares of the absolute values of the entries of the vector v, and thus by the square root of the inner product of v with its complex conjugate.

EXERCISE 6.4. Prove Parseval's Identity in the discrete setting. ◊

As in the Fourier series setting, Parseval's Identity says that the energy of the transformed signal is the same as that of the original signal. In the Fourier series setting, the *energy* means the L^2-norm of a function or the ℓ^2-norm of a doubly infinite series. In the discrete setting, the energy means the ℓ^2-norm of a vector. Parseval's Identity also holds in the Fourier transform setting. For example when the original and transformed signals, f and $\widehat{f}$, are both functions in $L^2(\mathbb{R})$, the energy is represented by the L^2-norm, and

$$\|f\|_{L^2(\mathbb{R})} = \|\widehat{f}\|_{L^2(\mathbb{R})}.$$

ASIDE 6.5. *Some authors work with the orthogonal basis* $\{f_0, f_1, \dots, f_{N-1}\}$ *of* $\mathbb{C}^N$ *defined by*

$$f_l = \sqrt{N}\, e_l,$$

instead of the orthonormal basis $\{e_l\}$. *Then a factor of* $1/N$ *appears in the definition of the coefficients and also in Parseval's Formula (but not in the reconstruction formula). To mirror the Fourier series theory, one could work with the vectors* $\{f_0, f_1, \dots, f_{N-1}\}$ *and consider them to be normalized, by defining the following inner product for vectors* v, $w \in \mathbb{C}^N$*:*

$$\langle v, w\rangle_{nor} := \frac{1}{N} \sum_{n=0}^{N-1} v(n)\overline{w(n)}. \tag{6.2}$$

We made an analogous convention in $L^2(\mathbb{T})$ *when we chose to include the factor* $1/(2\pi)$ *in the definition of the inner product. However, in our formulation of the discrete theory we prefer to use the standard inner product or dot product on* $\mathbb{C}^N$, *not the normalized one. Reader beware: it is wise to check which normalization a given author is using.*

Here are a few exercises dealing with the normalized inner product in $\mathbb{C}^N$.

EXERCISE 6.6. Show that formula (6.2) defines an inner product in $\mathbb{C}^N$. ◊

EXERCISE 6.7. Consider the vectors $\{f_0, f_1, \dots, f_{N-1}\}$. Show that they are orthonormal with respect to the normalized inner product in $\mathbb{C}^N$. ◊

EXERCISE 6.8. Define the normalized discrete Fourier coefficients of $v \in \mathbb{C}^N$ to be

$$\widehat{v}^{\,nor}(n) = \langle v, f_n \rangle_{nor}.$$

Use the theory of orthonormal bases to justify the reconstruction formula

$$v = \sum_{n=0}^{N-1} \widehat{v}^{\,nor}(n) f_n.$$

Verify that Parseval's Identity holds in this setting:

$$\sum_{n=0}^{N-1} |\widehat{v}^{\,nor}(n)|^2 = \|\widehat{v}^{\,nor}\|^2_{\ell^2(\mathbb{Z}/N\mathbb{Z})} = \|v\|^2_{nor} = \frac{1}{N} \sum_{n=0}^{N-1} |v(n)|^2.$$

◊

6.3. A parenthesis on bases and dual bases in $\mathbb{C}^N$

Before discussing the discrete Fourier transform in more detail, let us review some fundamental linear algebra facts about bases.

DEFINITION 6.9. A *basis* for $\mathbb{C}^N$ is a set of N vectors $\{v_1, v_2, \dots, v_N\}$ such that each vector $v \in \mathbb{C}^N$ can be written as a unique linear combination of the elements of the basis: there are complex numbers a_1, a_2, …, a_N such that

$$v = a_1 v_1 + a_2 v_2 + \cdots + a_N v_N. \tag{6.3}$$

The complex numbers a_1, a_2, …, a_N are uniquely determined by the vector v. ◊

Let B denote the $N \times N$ matrix whose j^{th} column is the vector v_j. We write

$$B = \begin{bmatrix} | & & | \\ v_1 & \cdots & v_N \\ | & & | \end{bmatrix},$$

where the vertical line segments are a reminder that the v_j are vectors, not scalars. With this notation, we have the following matrix representation of equation (6.3):

$$v = Ba, \quad \text{where} \quad a = [a_1, a_2, \dots, a_N]^t.$$

As usual, the superscript t indicates the transpose, so here a is a column vector.

THEOREM 6.10. *The following statements are equivalent.*

(1) *A set of N vectors $\{v_1, v_2, \dots, v_N\}$ is a basis for $\mathbb{C}^N$.*
(2) *The vectors $\{v_1, v_2, \dots, v_N\}$ are linearly independent. That is, if $0 = a_1 v_1 + a_2 v_2 + \cdots + a_N v_N$, then $a_1 = \cdots = a_N = 0$.*
(3) *The $N \times N$ matrix B whose columns are the vectors v_j is invertible (also called non-singular). Therefore we can find the coefficients a_1, a_2, …, a_N of a vector v with respect to the vectors $\{v_1, v_2, \dots, v_N\}$ by applying the inverse matrix B^{-1} to v:*

$$a = B^{-1} v.$$

Exercise 6.11. Prove the equivalences in Theorem 6.10. ◊

Given N linearly independent vectors $\{v_1, v_2, \ldots, v_N\}$ in $\mathbb{C}^N$, in other words a basis for $\mathbb{C}^N$, let B be the invertible matrix with columns v_j, and denote by w_j the complex conjugate of the j^{th} row vector of its inverse B^{-1}. Thus

$$B^{-1} = \begin{bmatrix} - & \overline{w_1} & - \\ & \vdots & \\ - & \overline{w_N} & - \end{bmatrix}.$$

The horizontal line segments emphasize that the $\overline{w_j}$ are vectors, not scalars.

With this notation, we see by carrying out the matrix multiplication that

$$B^{-1}v = \begin{bmatrix} - & \overline{w_1} & - \\ & \vdots & \\ - & \overline{w_N} & - \end{bmatrix} \begin{bmatrix} v(1) \\ \vdots \\ v(N) \end{bmatrix} = \begin{bmatrix} \langle v, w_1 \rangle \\ \vdots \\ \langle v, w_N \rangle \end{bmatrix},$$

where $v = [v(1), \ldots, v(N)]^t$.

Exercise 6.12. Verify that the j^{th} entry of the vector $B^{-1}v$ is the (complex!) inner product between the vector v and the vector w_j. ◊

The coefficients of v in the basis $\{v_1, v_2, \ldots, v_N\}$ are exactly the entries of the vector $B^{-1}v$. Therefore, not only are they uniquely determined, but we have an algorithm to find them:

$$a_j = \langle v, w_j \rangle,$$

where the conjugates of the vectors w_j are the rows of B^{-1}. In other words, there is a set of *dual vectors* forming a *dual basis* $\{w_1, w_2, \ldots, w_N\}$ to the basis $\{v_1, v_2, \ldots, v_N\}$, so that *to compute the coefficients of a given vector v in the original basis all we need to do is compute the inner products of v with the dual vectors.*

Exercise 6.13. (*Orthonormality, and Uniqueness of the Dual Basis*) Show that every basis $\{v_1, v_2, \ldots, v_N\}$ and its dual vectors $\{w_1, w_2, \ldots, w_N\}$ satisfy the following *orthonormality* condition:

$$\langle v_k, w_j \rangle = \delta_{k,j}. \tag{6.4}$$

Given a basis $\{v_1, v_2, \ldots, v_N\}$, show that if there are vectors $\{w_1, w_2, \ldots, w_N\}$ with the orthonormality property (6.4), then the matrix whose rows are the conjugates of the vectors $\{w_1, w_2, \ldots, w_N\}$ must be the inverse of the matrix whose columns are the vectors in the basis. That is, the vectors $\{w_1, w_2, \ldots, w_N\}$ are the dual basis for $\{v_1, v_2, \ldots, v_N\}$. ◊

We will encounter dual bases in an infinite-dimensional context when we discuss biorthogonal wavelets (Chapter 10).

In particular, if the basis $\{v_1, v_2, \ldots, v_N\}$ is orthonormal, then the dual vectors coincide with the original basis vectors. In other words, an orthonormal basis is its own dual basis. Equivalently, on the matrix side, if the basis is orthonormal then

$$B^{-1} = \overline{B^t}.$$

Definition 6.14. An $N \times N$ matrix whose columns are orthonormal vectors in $\mathbb{C}^N$ is called a *unitary matrix.* ◊

Unitary matrices are always invertible. The inverse of a unitary matrix U is equal to the conjugate of the transpose of U, that is $U^{-1} = \overline{U^t}$.

6.4. The discrete Fourier transform and its inverse

It is not hard to see that the transformation that maps the vector $v \in \mathbb{C}^N$ to the vector $\widehat{v} \in \mathbb{C}^N$ is linear. We call this transformation the *discrete Fourier transform*, or the *Fourier transform* on $\ell_2\big(\mathbb{Z}/(N\mathbb{Z})\big)$. The latter notation means that we are viewing $\mathbb{C}^N$ as an inner-product vector space whose elements are complex-valued functions defined on the discrete set $\mathbb{Z}/(N\mathbb{Z}) = \{0, 1, \ldots, N-1\}$, with the standard Euclidean inner product.

Definition 6.15. The *Fourier matrix* F_N is the matrix whose columns are formed by the orthogonal Fourier vectors $f_0, \ldots, f_{N-1}$. The entries of the Fourier matrix are

$$F_N(m,n) = e^{-2\pi i m n/N}, \qquad \text{for } m,\, n \in \{0, 1, \ldots, N-1\}.$$

We emphasize that the rows and columns of F_N are labeled from 0 to $N-1$, and not from 1 to N. ♢

The Fourier matrix is *symmetric*: $F_N^t = F_N$.

Example 6.16. (*Fourier Matrix for* $\mathbb{C}^8$) Here is the Fourier matrix F_N in the case when $N = 8$:

$$F_8 = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & \omega & \omega^2 & \omega^3 & \omega^4 & \omega^5 & \omega^6 & \omega^7 \\ 1 & \omega^2 & \omega^4 & \omega^6 & 1 & \omega^2 & \omega^4 & \omega^6 \\ 1 & \omega^3 & \omega^6 & \omega & \omega^4 & \omega^7 & \omega^2 & \omega^5 \\ 1 & \omega^4 & 1 & \omega^4 & 1 & \omega^4 & 1 & \omega^4 \\ 1 & \omega^5 & \omega^2 & \omega^7 & \omega^4 & \omega & \omega^6 & \omega^3 \\ 1 & \omega^6 & \omega^4 & \omega^2 & 1 & \omega^6 & \omega^4 & \omega^2 \\ 1 & \omega^7 & \omega^6 & \omega^5 & \omega^4 & \omega^3 & \omega^2 & \omega^1 \end{bmatrix},$$

where $\omega = \exp(2\pi i/8)$. ♢

The finite Fourier transform is completely determined by the $N \times N$ complex-valued matrix $(1/\sqrt{N})\overline{F_N}$. Specifically, the discrete Fourier transform $\widehat{v}$ of a vector v is given by

$$\widehat{v} = \frac{1}{\sqrt{N}}\,\overline{F_N} v. \tag{6.5}$$

The matrix $(1/\sqrt{N})F_N$ is *unitary*, because its columns are orthonormal vectors. Therefore its inverse is its complex conjugate, so

$$\left(\frac{1}{\sqrt{N}}\,\overline{F_N}\right)^{-1} = \frac{1}{\sqrt{N}}\,F_N.$$

Thus the *discrete inverse Fourier transform* in $\ell_2\big(\mathbb{Z}_N/(N\mathbb{Z})\big)$ can be written as

$$v = \frac{1}{\sqrt{N}}\,F_N \widehat{v}. \tag{6.6}$$

In this finite-dimensional context, we are making an orthogonal change of basis in $\mathbb{C}^N$, from the standard basis to the Fourier basis.

Exercise 6.17. Show that

$$\frac{1}{N} F_N\,\overline{F_N} = \frac{1}{N}\overline{F_N}\,F_N = I_N,$$

where I_N denotes the $N \times N$ identity matrix. ♢

Both F_N and $\overline{F_N}$ are full matrices, in the sense that all their entries are nonzero. So applying one of them to a vector to compute the discrete Fourier transform, or the discrete inverse Fourier transform, requires N^2 multiplications if the vector has no zero entries. Since additions take up an insignificant amount of computer time compared with multiplications, the discrete Fourier transform also has an operation count of the order of N^2 operations.

However, one can dramatically improve on this operation count by exploiting the hidden structure of the Fourier matrix, as we now show.

6.5. The Fast Fourier Transform (FFT)

A key to the practical success of the discrete Fourier transform (6.5) is the existence of a fast algorithm to accomplish it. In the 1960s the American applied mathematician James Cooley (born in 1926) and the American statistician John Wilder Tukey [1915–2000] rediscovered a faster algorithm, which the German mathematician Johann Carl Friedrich Gauss [1777–1855] had found in the early 1800s. This famous *Fast Fourier Transform* (FFT) algorithm reduces the number of multiplications from order N^2 to order $N \log_2 N$. This improvement revolutionized digital signal processing (which was of course completely unforeseen in the time of Gauss). One measure of the influence of the FFT algorithm is that over the period 1945–1988, Cooley and Tukey's paper [**CT**] was the sixth most-cited paper in the Science Citation Index in the areas of mathematics, statistics, and computer science [**Hig**, p.217].

Mathematically, the Fast Fourier Transform is based on a factorization of the Fourier matrix into a product of *sparse* matrices, meaning matrices with many zero entries.

EXERCISE 6.18. The conjugate $\overline{F_4}$ of the 4×4 Fourier matrix F_4 can be written as the product of three sparse matrices, as follows:

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & i & -1 & -i \\ 1 & -1 & 1 & -1 \\ 1 & -i & -1 & i \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & -i \\ 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & i \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{6.7}$$

Use matrix multiplication to check formula (6.7). Can you spot two copies of $\overline{F_2}$ on the right-hand side? ◊

The process of the Fast Fourier Transform is easiest to explain, and to implement, when $N = 2^n$. (If N is not a power of two, one can simply add an appropriate number of zeroes to the end of the signal vector to make N into a power of two, a step known as *zero-padding*.) It begins with a factorization that reduces F_N to two copies of $F_{N/2}$. This reduction is applied to the smaller and smaller matrices $F_{N/2^j}$ for a total of $n = \log_2 N$ steps, until the original matrix is written as a product of $2 \log_2 N$ sparse matrices, half of which can be collapsed into a single permutation matrix.

The total number of multiplications required to apply the Fourier transform to a vector is reduced from order N^2, for the original form of the matrix F_N, to the much smaller order $N \log_2 N$, for the factorized version of F_N.

To understand how this reduction works at the level of individual entries of the transformed vector, we rearrange formula (6.1) for the discrete Fourier transform to express $\widehat{v}(m)$ as a linear combination of two terms A_m and B_m:

$$\begin{aligned}
\widehat{v}(m) &= \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} v(n) e^{\frac{-2\pi i m n}{N}} \\
&= \frac{1}{\sqrt{N}} \left[\sum_{n=0}^{N/2-1} v(2n) e^{\frac{-2\pi i m n}{N/2}} + e^{\frac{-2\pi i m}{N}} \sum_{n=0}^{N/2-1} v(2n+1) e^{\frac{-2\pi i m n}{N/2}} \right] \\
&=: A_m + e^{\frac{-\pi i m}{N/2}} B_m \, .
\end{aligned}$$

The key is to note that $\widehat{v}(m + N/2)$ can be written as another linear combination of the same two terms, namely

$$\begin{aligned}
\widehat{v}\left(m + \frac{N}{2}\right) &= \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} v(n) e^{\frac{-2\pi i (m+N/2) n}{N}} \\
&= \frac{1}{\sqrt{N}} \left[\sum_{n=0}^{\frac{N}{2}-1} v(2n) e^{\frac{-2\pi i m n}{N/2}} + e^{\frac{-2\pi i (m+N/2)}{N}} \sum_{n=0}^{\frac{N}{2}-1} v(2n+1) e^{\frac{-2\pi i m n}{N/2}} \right] \\
&= A_m - e^{\frac{-\pi i m}{N/2}} B_m \, .
\end{aligned}$$

By clever ordering of operations we can reuse A_m and B_m, reducing the total number of operations required. Furthermore, the sums A_m and B_m have the same structure as the sum in $\widehat{v}(m)$, so we can do this process recursively.

In the language of matrices, we have factored the conjugate Fourier matrix $\overline{F_N}$, for N even, into a product of three $N \times N$ matrices:

$$\overline{F_N} = \begin{bmatrix} I_{N/2} & D_{N/2} \\ I_{N/2} & -D_{N/2} \end{bmatrix} \begin{bmatrix} \overline{F_{N/2}} & 0_{N/2} \\ 0_{N/2} & \overline{F_{N/2}} \end{bmatrix} \begin{bmatrix} \text{Even}_{N/2} \\ \text{Odd}_{N/2} \end{bmatrix} . \tag{6.8}$$

Here I_M denotes the $M \times M$ identity matrix, F_M is the $M \times M$ Fourier matrix, 0_M denotes the $M \times M$ zero matrix, Even_M and Odd_M are the $M \times 2M$ matrices that select in order the even and odd entries respectively of the vector $(v(0), v(1), \ldots, v(2M-1))^t$.

EXERCISE 6.19. Compare equation (6.8) with the factorization of $\overline{F_4}$ given in Exercise 6.18. In particular, in equation (6.7) identify the matrices I_2, D_2, $\overline{F_2}$, Even_2, and Odd_2. ♢

We now describe the diagonal $M \times M$ matrix D_M, and the $2M \times 2M$ matrices S_M defined to be

$$S_M = \begin{bmatrix} \text{Even}_M \\ \text{Odd}_M \end{bmatrix},$$

for general M.

The diagonal matrix D_M has diagonal entries $e^{-m\pi i/M}$ for $m = 0, 1, \ldots, M-1$. For example, here are D_2 and D_4:

$$D_2 = \begin{bmatrix} e^0 & 0 \\ 0 & e^{-\pi i/2} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & -i \end{bmatrix},$$

$$D_4 = \begin{bmatrix} e^0 & 0 & 0 & 0 \\ 0 & e^{-\pi i/4} & 0 & 0 \\ 0 & 0 & e^{-2\pi i/4} & 0 \\ 0 & 0 & 0 & e^{-3\pi i/4} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \frac{1-i}{\sqrt{2}} & 0 & 0 \\ 0 & 0 & -i & 0 \\ 0 & 0 & 0 & \frac{-1-i}{\sqrt{2}} \end{bmatrix}.$$

The appearance in D_2 of the fourth root of unity, in D_4 of the eighth root of unity, and so on, illuminates how the Fourier matrix F_N can contain the N^{th} root of unity, while $F_{N/2}$ contains only the $(N/2)^{\text{th}}$ root and the other matrices in the recursion equation contain only real numbers.

The $2M \times 2M$ matrix S_M selects the even-numbered entries (starting with the zeroth entry) of a given $2M$-vector and moves them to the top half of the vector, preserving their order, and it moves the odd-numbered entries of the $2M$-vector to the bottom half of the vector, preserving their order. Here is the 8×8 matrix S_4 acting on a generic vector $v = [v(0), v(1), \ldots, v(7)]$ in $\mathbb{C}^8$:

$$S_4 v = \begin{bmatrix} \text{Even}_4 \\ \text{Odd}_4 \end{bmatrix} v = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} v(0) \\ v(1) \\ v(2) \\ v(3) \\ v(4) \\ v(5) \\ v(6) \\ v(7) \end{bmatrix} = \begin{bmatrix} v(0) \\ v(2) \\ v(4) \\ v(6) \\ v(1) \\ v(3) \\ v(5) \\ v(7) \end{bmatrix}.$$

Another way to think of the effect of the matrix containing $\text{Even}_{N/2}$ and $\text{Odd}_{N/2}$ is that it permutes the columns of the matrix to its left in the recursion equation (6.8), moving the even-numbered columns to the left side of the matrix and the odd-numbered columns to the right side.

Turning to the number of operations required in the Fast Fourier Transform, we see that the first matrix in the decomposition (6.8) requires only N multiplications. The second matrix ostensibly requires $2(N/2)^2 = N^2/2$ multiplications, but it has the conjugate Fourier matrix $\overline{F_{N/2}}$ in the diagonal, and we can use the recursion equation (6.8) again to replace the $(N/2) \times (N/2)$ matrix $\overline{F_{N/2}}$ by

$$\overline{F_{N/2}} = \begin{bmatrix} I_{N/4} & D_{N/4} \\ I_{N/4} & -D_{N/4} \end{bmatrix} \begin{bmatrix} \overline{F_{N/4}} & 0_{N/4} \\ 0_{N/4} & \overline{F_{N/4}} \end{bmatrix} S_{N/4}$$

Next, we can apply the recursion equation to the $(N/4) \times (N/4)$ matrix $\overline{F_{N/4}}$, and keep on reducing in the same way until we cannot subdivide any further, when we reach the 2×2 matrix $\overline{F_2}$.

Exercise 6.20. What three matrices result from applying the recursion equation (6.8) to $\overline{F_2}$? ♢

Exercise 6.21. Verify that for F_N, after two recursive steps the second matrix on the right-hand side of equation (6.8) is decomposed as the product of three

$N \times N$ matrices as follows:

$$\begin{bmatrix} I_{N/4} & D_{N/4} & & \\ I_{N/4} & -D_{N/4} & & \\ & & I_{N/4} & D_{N/4} \\ & & I_{N/4} & -D_{N/4} \end{bmatrix} \begin{bmatrix} F_{N/4} & & & \\ & F_{N/4} & & \\ & & F_{N/4} & \\ & & & F_{N/4} \end{bmatrix} \begin{bmatrix} S_{N/4} & \\ & S_{N/4} \end{bmatrix}.$$

Here, blank spaces represent zero submatrices of appropriate sizes. ♢

EXERCISE 6.22. Write down the decomposition of the conjugate Fourier matrix $\overline{F_8}$ into a product of five matrices, by using equation (6.8) twice. Check your answer by matrix multiplication. How many multiplication operations are required to multiply a vector by the matrix $\overline{F_8}$? How many are required to multiply a vector by the five matrices in your factorization? (Count only multiplications by matrix entries that are neither one nor zero. The operations required to multiply a vector by the Even and Odd matrices are also often ignored in the operation count, since their effect is simply to change the order of the entries of the vector.) ♢

After two iterations, the first two matrices in the decomposition of F_N require only N multiplications each. If $N = 2^j$, then one can do this $j - 1 = \log_2 N$ times, until the block size is 2×2. After $j - 1$ steps, the matrix $\overline{F_N}$ will be decomposed into the product of $\log_2 N - 1$ matrices, each requiring only N multiplications, for a total of $N(\log_2 N - 1)$ multiplications, followed by the product of $j-1$ *scrambling* or permutation matrices, namely the Odd/Even matrices.

Then a miracle occurs. The $j-1$ scrambling matrices, when applied in the right order, become a very simple operation which costs essentially nothing in terms of operation counts. The miracle is observed when describing the numbers 0, 1, …, $N - 1$ in binary notation: "the successive application of the Odd/Even matrices will simply put in place n the m^{th} entry, where m is found by reversing the digits in the binary decomposition of n". In computer science this step is called *bit reversal*.

Let us see the miracle in action when $N = 8 = 2^3$. Denote the successive scrambling matrices by

$$S_4^8 := S_4 = \begin{bmatrix} \text{Even}_4 \\ \text{Odd}_4 \end{bmatrix}, \quad S_2^8 = \begin{bmatrix} S_2 & 0_4 \\ 0_4 & S_2 \end{bmatrix}.$$

We start with the numbers 0, 1, …, 7 in binary notation and apply the matrix S_4^8, followed by the matrix S_2^2:

$$\begin{bmatrix} 000 \\ 001 \\ 010 \\ 011 \\ 100 \\ 101 \\ 110 \\ 111 \end{bmatrix} \overset{S_4^8}{\longrightarrow} \begin{bmatrix} 000 \\ 010 \\ 100 \\ 110 \\ 001 \\ 011 \\ 101 \\ 111 \end{bmatrix} \overset{S_2^8}{\longrightarrow} \begin{bmatrix} 000 \\ 100 \\ 010 \\ 110 \\ 001 \\ 101 \\ 011 \\ 111 \end{bmatrix}.$$

We see that the binary numbers in the right-hand column are obtained from the binary numbers in the left-hand column by reading the binary digits in reverse order.

EXERCISE 6.23. Can you justify the miracle for all $N = 2^j$? Try an argument by induction. Note that in this case the j^{th} miracle matrix M_j is the product of

$j-1$ scrambling $N \times N$ matrices, $M_j = S_2^N S_4^N \dots S_{2^{j-1}}^N$. The scrambling matrix $S_{2^k}^N$ for $k < j$ is a block diagonal matrix with 2^{j-k-1} copies of the $2^{k+1} \times 2^{k+1}$ matrix S_{2^k} on the diagonal. ◊

Even without a miracle, the scrambling $N \times N$ matrices $S_{2^k}^N$ are very sparse. Only N entries of $S_{2^k}^N$ are nonzero, and in fact the nonzero entries are all ones. Applying each of the scrambling matrices requires only N operations, and applying $j-1$ such matrices ($N = 2^j$) requires at most $N(\log_2 N - 1)$ operations. This means that the operation count for the Fourier matrix using this decomposition is no more than $2N(\log_2 N - 1)$ operations. So the algorithm is still of order at most $N \log_2 N$.

EXERCISE 6.24. (*Fast Convolution of Vectors*) Given two vectors v, $w \in \mathbb{C}^N$, define their *discrete circular convolution* by

$$v * w(n) := \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} v(k) w(n-k), \quad n = 0, 1, \dots, N-1.$$

If $n < k$ then $-N \le n - k < 0$. In that case, define $w(n-k) := w(N - (n-k))$, or, equivalently, think of w as being extended periodically over $\mathbb{Z}$ so that $w(n) = w(N+n)$ for all $n \in \mathbb{Z}$. In matrix notation, the linear transformation T_w that maps the vector v into the vector $v * w$ is given by the *circulant matrix*

$$\begin{bmatrix} w(0) & w(N-1) & w(N-2) & \dots & w(1) \\ w(1) & w(0) & w(N-1) & \dots & w(2) \\ w(2) & w(1) & w(0) & \dots & w(3) \\ \vdots & \vdots & \vdots & & \vdots \\ w(N-1) & w(N-2) & w(N-3) & \dots & w(0) \end{bmatrix}.$$

(The term circulant indicates that down each diagonal, the entries are constant.) *A priori*, multiplying a vector by this full matrix requires N^2 operations.

(a) Show that

$$\widehat{v * w}(m) = \widehat{v}(m)\, \widehat{w}(m).$$

(b) Describe a fast algorithm to compute the convolution of two vectors in order $N \log_2 N$ operations. ◊

Convolution with a fixed vector w is a linear transformation given by a circulant matrix, as described in Exercise 6.24. These are also the only *translation-invariant* or *shift-invariant* linear transformations, meaning that if we first shift a vector and then convolve, the output is the same as first convolving and then shifting. More precisely, denote by S_k the linear transformation that shifts a vector $v \in \mathbb{C}^N$ by k units, that is $S_k v(n) := v(n-k)$. Here we are viewing the vector as extended periodically so that $v(n-k)$ is defined for all integers. Then $S_k T_w = T_w S_k$.

EXERCISE 6.25. Show that if $T_w v := v * w$ then $S_k T_w = T_w S_k$. Conversely, show that if a linear transformation T is shift invariant, meaning that $S_k T = T S_k$ for all $k \in \mathbb{Z}$, then there is a vector $w \in \mathbb{C}^N$ such that $Tv = v * w$. ◊

The book by Frazier [**Fra**] takes a linear algebra approach and discusses these ideas and much more in the first three chapters, before moving on to infinite-dimensional space. Chapter 1 is a thorough review of linear algebra with emphasis on changes of basis and translation invariant linear transformations. Chapter 2 discusses discrete Fourier theory. Chapter 3 discusses discrete wavelet theory.

Strang and Nguyen's book [**SN**] also has a linear algebra perspective.

6.6. The discrete Haar basis

In this section we discuss the discrete Haar basis. Unlike the discrete Fourier basis, the Haar basis is localized, a concept that we now describe.

By *localized* we mean, in a not very precise way, that the vector is zero except for a few entries. All entries of the discrete trigonometric vectors are nonzero, and in fact each entry has the same absolute value $1/\sqrt{N}$. Thus the discrete trigonometric vectors are certainly not localized. In particular, since the Fourier coefficients of a vector v are given by the inner product v against a trigonometric vector e_j, all entries of v are involved in the computation.

An example of a basis which is as localized as possible is the *standard basis* in $\mathbb{C}^N$. The standard basis vectors have all but one entry equal to zero; the nonzero entry is 1. Denote by s_j the vector in the standard basis whose j^{th} entry is 1:

$$s_1 = [1, 0, \ldots, 0]^t, s_2 = [0, 1, 0, \ldots, 0]^t, \ldots, s_N = [0, \ldots, 0, 1]^t.$$

The standard basis changes with the dimension, we should tag the vectors $s_j = s_j^N$, but as with the Fourier basis, we will omit the reference to the underlying dimension and hope the reader does not get confused.

Exercise 6.26. Show that the discrete Fourier transform in $\mathbb{C}^N$ of e_j is given by $\widehat{e_j} = s_{j+1}$, for $1 \leq j \leq N$. Start with the case $N = 4$. $\Diamond$

Although the Fourier basis is not localized at all, its Fourier transform is as localized as possible. We say the Fourier basis is *localized in frequency*, but not in space or time.

Exercise 6.27. Verify that the discrete Fourier transform of s_j is the complex conjugate of e_{j-1}, that is $\widehat{s_j} = \overline{e_{j-1}}$. $\Diamond$

Although the standard basis is as localized as possible, its Fourier transform is not localized at all. We say the standard basis is *localized in space or time* but not in frequency.

Exercise 6.28. Let $\{v_1, v_2, \ldots, v_N\}$ be an orthonormal basis for $\mathbb{C}^N$. Show that the discrete Fourier transforms $\{\widehat{v_1}, \widehat{v_2}, \ldots, \widehat{v_N}\}$ form an orthonormal basis for $\mathbb{C}^N$. $\Diamond$

These examples are incarnations of a general principle, the *Uncertainty Principle*, that we discuss in more depth in subsequent chapters. The Uncertainty Principle says that it is impossible for a vector to be simultaneously localized in space and frequency.

The *discrete Haar basis* is intermediate between the discrete Fourier basis and the standard basis, in terms of its space and frequency localization.

We begin with the example of the discrete Haar basis for $\mathbb{C}^8$, where $N = 2^3 = 8$.

EXERCISE 6.29. Define the vectors $\tilde{h}_n$ for $n = 0, 1, \ldots, 7$ by

$$\begin{aligned}
\tilde{h}_0 &:= [1,1,1,1,1,1,1,1]^t,\\
\tilde{h}_1 &:= [-1,-1,-1,-1,1,1,1,1]^t,\\
\tilde{h}_2 &:= [-1,-1,1,1,0,0,0,0]^t,\\
\tilde{h}_3 &:= [0,0,0,0,-1,-1,1,1]^t,\\
\tilde{h}_4 &:= [-1,1,0,0,0,0,0,0]^t,\\
\tilde{h}_5 &:= [0,0,-1,1,0,0,0,0]^t,\\
\tilde{h}_6 &:= [0,0,0,0,-1,1,0,0]^t,\\
\tilde{h}_7 &:= [0,0,0,0,0,0,-1,1]^t.
\end{aligned}$$

Show that the vectors $\{\tilde{h}_0, \tilde{h}_1, \ldots, \tilde{h}_7\}$ are orthogonal vectors in $\mathbb{C}^8$. Show that the (normalized) vectors

$$h_0 = \frac{\tilde{h}_0}{\sqrt{8}}, \quad h_1 = \frac{\tilde{h}_1}{\sqrt{8}}, \quad h_2 = \frac{\tilde{h}_2}{\sqrt{4}}, \quad h_3 = \frac{\tilde{h}_3}{\sqrt{4}},$$
$$h_4 = \frac{\tilde{h}_4}{\sqrt{2}}, \quad h_5 = \frac{\tilde{h}_5}{\sqrt{2}}, \quad h_6 = \frac{\tilde{h}_6}{\sqrt{2}}, \quad h_7 = \frac{\tilde{h}_7}{\sqrt{2}}, \tag{6.9}$$

are orthonormal. They form the discrete Haar basis for $\mathbb{C}^8$. ♢

EXERCISE 6.30. Compute the discrete Fourier transform $\widehat{h}_n$ of each of the vectors in the Haar basis for $\mathbb{C}^8$, displayed in (6.9). Either analytically or using MATLAB, sketch the graphs of each h_n and $\widehat{h}_n$. Notice that the more localized h_n is, the less localized $\widehat{h}_n$ is. ♢

When $N = 2^j$, $j \in \mathbb{N}$, the N Haar vectors are defined according to the analogous pattern. The first non-normalized vector $\tilde{h}_0$ is a string of ones; the second vector $\tilde{h}_1$ has the first half of its entries equal to -1, and the second half equal to 1; the third vector $\tilde{h}_2$ has the first quarter of its entries equal to -1, the second quarter 1, and the rest of the entries zero; and so on. The last non-normalized vector $\tilde{h}_{N-1}$ starts with $N-2$ zeros, and its last two entries are -1 and 1. Again we are abusing notation by not labeling the Haar vectors with the underlying dimension as we should.

EXERCISE 6.31. Write down the discrete Haar basis for $\mathbb{C}^N = \mathbb{C}^{2^j}$ for some $j \geq 4$. ♢

Notice that half of the vectors are very localized, having all entries zero except for two consecutive nonzero entries. These are the 2^{j-1} vectors $\{\tilde{h}_{2^{j-1}}, \tilde{h}_{2^{j-1}+1}, \ldots, \tilde{h}_{2^j-1}\}$. The 2^{j-2} vectors $\{\tilde{h}_{2^{j-2}}, \tilde{h}_{2^{j-2}+1}, \ldots, \tilde{h}_{2^{j-1}-1}\}$ have all entries zero except for four consecutive nonzero entries. The 2^{j-k} vectors $\{\tilde{h}_{2^{j-k}}, \tilde{h}_{2^{j-k}+1}, \ldots, \tilde{h}_{2^{j-k+1}-1}\}$ have all entries zero except for 2^k consecutive nonzero entries. The vectors $\{\tilde{h}_2, \tilde{h}_3\}$ have half of the entries zero. Finally $\tilde{h}_0$ and $\tilde{h}_1$ have all entries nonzero. Furthermore the nonzero entries for $\tilde{h}_1$ consist of a string of consecutive -1s followed by the same number of 1s, and for $\tilde{h}_0$ is just a string of 1s..

We now give a precise definition of the N^{th} Haar basis. We should really tag each of the N vectors with an N to indicate the dimension of $\mathbb{C}$. However, to make the notation less cluttered, we won't bother to do so.

Given $N = 2^j$ and n with $1 \leq n < N$, there is a unique $k \in \mathbb{Z}$ with $0 \leq k < j$ such that

$$2^k \leq n < 2^{k+1}, \quad \text{moreover} \quad n = 2^k + m, \;\; 0 \leq m < 2^k.$$

For $n \geq 1$, define $\tilde{h}_n$, in terms of the unique numbers k and m, $0 \leq k < j$, $0 \leq m < 2^k$, determined by n:

$$\tilde{h}_n(l) = \begin{cases} -1, & \text{if } mN2^{-k} \leq l < (m+\frac{1}{2})N2^{-k}; \\ 1 & \text{if } (m+\frac{1}{2})N2^{-k} \leq l < (m+1)N2^{-k}; \\ 0 & \text{otherwise.} \end{cases}$$

We define $\tilde{h}_0(l) = 1$ for all $0 \leq l \leq N-1$.

Exercise 6.32. Verify that when $N = 8$ the formula above for $\tilde{h_n}(l)$ gives the same vectors as in Example 6.29. ♢

Notice that $\tilde{h}_n(l)$ is nonzero if and only if $m2^{j-k} \leq l < (m+1)2^{j-k}$. In fact those $2^{j-k} = N/2^k$ entries are ± 1. The Haar vectors are defined by

$$h_n = \frac{2^{k/2}}{\sqrt{N}} \tilde{h}_n.$$

Exercise 6.33. Show that the vectors $\{h_n\}_{n=0}^{N-1}$ are orthonormal. ♢

Notice that except for h_0, all the other vectors are rescaled contractions of h_1 and translates of those. See Exercise 6.34.

The parameters k, m, uniquely determined by n so that $n = 2^k + m$, are *scaling* and *translation* parameters respectively. The scaling parameter k, $0 \leq k < j$, tells us that we will decompose the set $\{0, 1, 2, \ldots, 2^j - 1\}$ into 2^k disjoint subsets of $2^{j-k} = N2^{-k}$ consecutive numbers. The translation parameter m, $0 \leq m < 2^k$, indicates which subinterval of length $N2^{-k}$ we are considering, namely the interval $[mN2^{-k}, (m+1)N2^{-k})$.

It would make sense to index the n^{th} Haar vector with the pair (k, m), and we will do so when we study the Haar basis, and some wavelet bases on $L^2(\mathbb{R})$. These are orthonormal bases, $\{\psi_{k,m}\}_{k,m\in\mathbb{Z}}$, indexed by two integer parameters (k, m). The Haar and wavelet bases have the unusual feature that they are found by dilating and translating one function ψ, known as the *wavelet*. More precisely,

$$\psi_{k,m}(x) := 2^{k/2}\psi(2^k x - m).$$

The functions $\psi_{k,m}$ retain the shape of the wavelet ψ. If ψ is concentrated around zero on the interval $[-1, 1]$, then the function $\psi_{k,m}$ is now localized around the point $m2^{-k}$, and concentrated on the interval $[(m-1)2^{-k}, (m+1)2^{-k}]$ of length 2×2^{-k}. As $k \to \infty$ the *resolution* increases (that is, the localization improves), while as $k \to -\infty$ the localization gets worse.

Exercise 6.34. Show that if $N = 2^j$, $n = 2^k + m$, for $0 \leq k < j$, $0 \leq m < 2^k$, and we write $h_{k,m} := h_n$, then

$$h_{k,m}(l) = 2^{k/2} h_{0,0}(2^k l - mN).$$

Here it is understood that $h_{0,0}(z) = h_1(z) = 0$ whenever $z \neq 0, 1, \ldots N-1$. We are padding zeros to the right and left of $h_{0,0}$, not extending it periodically.

♢

6.7. The Discrete Haar Transform

The *discrete Haar transform* is the linear transformation that applied to a vector $v \in \mathbb{C}^N$, $N = 2^j$, gives the coefficients of v in the orthonormal Haar basis. The *inverse discrete Haar transform* is the linear transformation that reconstructs the vector v from its Haar coefficients $\{c_n\}_{n=1}^{N-1}$. Notice that

$$v = c_0 h_0 + c_1 h_1 + \cdots + c_{N-1} h_{N-1}.$$

In matrix notation the inverse discrete Haar transform is given by

$$v = H_N\, c, \quad \text{where } c = [c_0, c_1, \ldots, c_{N-1}]^t,$$

and the Haar matrix H_N is the $N \times N$ matrix whose j^{th} column is the Haar vector h_{j-1}. That is,

$$H_N = \begin{bmatrix} | & & | \\ h_0 & \ldots & h_{N-1} \\ | & & | \end{bmatrix}.$$

The Haar matrix H_N is a unitary matrix (it has orthonormal columns), hence its inverse is its conjugate transpose. The Haar matrix H_N is also a real matrix (the entries are the real numbers $0, \pm 2^{(k-j)/2}$), hence $H_N^{-1} = H_N^t$. We can recover the coefficients $c \in \mathbb{C}^N$ of the vector $v \in \mathbb{C}^N$ by applying the transpose of H_N to the vector v. Therefore the discrete Haar transform is given by

$$c = H_N^t\, v.$$

Notice that either from the above matrix formula or from the fact that the Haar basis is an orthonormal basis, the j^{th} coefficient is calculated by taking the inner product of v with the j^{th} Haar vector:

$$c_j = \langle v, h_j \rangle.$$

EXAMPLE 6.35. Here is the matrix H_8^t:

$$H_8^t = \begin{bmatrix} \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} \\ \frac{-1}{\sqrt{8}} & \frac{-1}{\sqrt{8}} & \frac{-1}{\sqrt{8}} & \frac{-1}{\sqrt{8}} & \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} \\ \frac{-1}{\sqrt{4}} & \frac{-1}{\sqrt{4}} & \frac{1}{\sqrt{4}} & \frac{1}{\sqrt{4}} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{-1}{\sqrt{4}} & \frac{-1}{\sqrt{4}} & \frac{1}{\sqrt{4}} & \frac{1}{\sqrt{4}} \\ \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}.$$

◊

The Haar matrix H_8^t is sparse; it has many zero entries. Counting the nonzero entries, we note that the first and second rows are both full. The third and fourth rows are half-full, so they make one full row together. The fifth, sixth, seventh and eighth rows together make one full row of nonzero entries. Adding, we get $2 + 1 + 1 = 4$ full rows in H_8^t where $n = 8 = 2^3$. In H_{16}^t we get the equivalent of four full rows from the first 8 rows as in H_8^t and the 9th to 16th would make up one more. i.e. $2 + 1 + 1 + 1 = 5$ full rows in H_{16}^t where $N = 16 = 2^4$. The general formula is if $N = 2^j$ we get $j + 1$ full columns, each of length N. So the total number of nonzero entries is $N(j+1)$ where $j = \log_2 N$. Hence, multiplying a vector of length N by H_N takes $N(1+\log_2 N)$ multiplications. This implies that the Discrete Haar Transform can be performed in order $N\log_2 N$ operations. Here is an argument for seeking localized basis vectors: *The brute force operation count is of the same order as the FFT, because of the sparseness of the Haar matrix.* Can this count be improved by using some smart algorithm as we did for the discrete Fourier transform? Yes, the operation count can be brought down to order N operations. We call such an algorithm the *Fast Haar Transform*, and it is the first example of the *Fast Wavelet Transform.*

6.8. The Fast Haar Transform

We illustrate with some examples how one can apply the discrete Haar transform more efficiently. We choose here to argue, as we did for the FFT, in terms of matrix decomposition. We will revisit this algorithm when we discuss the *multiresolution analysis* for wavelets.

First consider $N = 2$. Let us explicitly compute the action of the matrix H_2^t on a vector $v = [v(0), v(1)]^t \in \mathbb{C}^2$.

$$H_2^t v = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}\begin{bmatrix} v(0) \\ v(1) \end{bmatrix} = \begin{bmatrix} \frac{v(1)+v(0)}{\sqrt{2}} \\ \frac{v(1)-v(0)}{\sqrt{2}} \end{bmatrix}.$$

Notice that the output consists of scaled averages and differences of $v(0)$ and $v(1)$. In terms of multiplications, we only need two multiplications, namely $\frac{1}{\sqrt{2}}v(0)$ and $\frac{1}{\sqrt{2}}v(1)$, then we add these two numbers and subtract them. For operation counts, what is costly is the multiplications. *Applying H_2^t requires only 2 multiplications.*

Now consider the case $N = 4$. Let us explicitly compute the action of the matrix H_4^t on a vector $v = [v(0), v(1), v(2), v(3)]^t \in \mathbb{C}^4$:

$$H_4^t v = \begin{bmatrix} \frac{1}{\sqrt{4}} & \frac{1}{\sqrt{4}} & \frac{1}{\sqrt{4}} & \frac{1}{\sqrt{4}} \\ \frac{-1}{\sqrt{4}} & \frac{-1}{\sqrt{4}} & \frac{1}{\sqrt{4}} & \frac{1}{\sqrt{4}} \\ \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 & 0 \\ 0 & 0 & \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}\begin{bmatrix} v(0) \\ v(1) \\ v(2) \\ v(3) \end{bmatrix} = \begin{bmatrix} \frac{v(3)+v(4)+v(1)+v(0)}{\sqrt{4}} \\ \frac{\left(v(3)+v(2)\right)-\left(v(1)+v(0)\right)}{\sqrt{4}} \\ \frac{v(1)-v(0)}{\sqrt{2}} \\ \frac{v(3)-v(2)}{\sqrt{2}} \end{bmatrix}.$$

Denote the scaled averages of pairs of consecutive entries by a_m^1, and their differences by d_m^1. In our case $j = 2$.

$$a_0^1 := \frac{v(1)+v(0)}{\sqrt{2}}, \quad d_0^1 := \frac{v(1)-v(0)}{\sqrt{2}}.$$

$$a_1^1 := \frac{v(3)+v(2)}{\sqrt{2}}, \quad d_1^1 := \frac{v(3)-v(2)}{\sqrt{2}}.$$

Notice that they are the outputs of the 2×2 matrix H_2^t applied to the vectors $[v(0), v(1)]^t, [v(2), v(3)]^t \in \mathbb{C}^2$. With this notation, the output of the matrix $H_4^t v$ is the vector

$$H_4^t v = \begin{bmatrix} \frac{a_1^1+a_0^1}{\sqrt{2}} & \frac{a_1^1-a_0^1}{\sqrt{2}} & d_0^1 & d_1^1 \end{bmatrix}^t.$$

The first two entries are the output of the matrix H_2^t applied to the average vector $a^1 = [a_0^1, a_1^1]$. The last two entries are the output of the 2×2 identity matrix applied to the difference vector $d^1 = [d_0^1, d_1^1]$.

ASIDE 6.36. *The superscript 1 is at this stage unnecessary. However we are computing averages and differences of the averages. We can introduce* $a^0 = a^0_0 := \frac{a^1_1+a^1_0}{\sqrt{2}}$, $d^0 = d^0_0 := \frac{a^1_1-a^1_0}{\sqrt{2}}$. *With this notation we have*

$$H_4^t v = \begin{bmatrix} a^0 & d^0 & d^1 \end{bmatrix}^t.$$

Notice that the vectors d^0 *and* a^0 *have dimension* $1 = 2^0$, *and* d^1 *has dimension* $2 = 2^1$, *so the output has dimension* 4.

In the same way, we can decompose H_4^t as the product of two 4×4 matrices:

$$H_4^t = \begin{bmatrix} \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} & 0 \\ \frac{-1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} H_2^t & 0_2 \\ 0_2 & H_2^t \end{bmatrix}.$$

The second matrix requires 2×2 multiplications, since it involves applying H_2^t twice. The first matrix requires only 2 multiplications, since it involves applying H_2^t once and then the 2×2 identity matrix which involves no multiplications. The total multiplication count for H_4^t is thus $2 \times 2 + 2 = 6 = 2(4 - 1)$.

Notice that the first matrix can be decomposed as the product of a block diagonal matrix (with H_2^t and I_2 (2×2 identity matrix) in the diagonal) with a permutation matrix. In other words,

$$\begin{bmatrix} \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} & 0 \\ \frac{-1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} H_2^t & 0_2 \\ 0_2 & I_2 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

EXERCISE 6.37. Show that H_8^t can be decomposed as the product of three 8×8 matrices as follows:

$$H_8^t = \begin{bmatrix} H_2^t & 0_2 \\ 0_2 & I_6 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} H_4^t & 0_4 \\ 0_4 & H_4^t \end{bmatrix}.$$

Notice that the last matrix on the right hand side requires $2 \times 6 = 12$ multiplications, since it involves applying H_4^t twice. The first matrix on the right hand side requires only 2 multiplications, since it involves applying H_2^t once and then the 6×6 identity matrix I_6 which involves no multiplications. The middle matrix is a permutation matrix, denoted by P_8. It involves no multiplications. The total multiplication count for H_8^t is thus $2 \times 6 + 2 = 14 = 2(8 - 1)$. ♢

We can see a pattern emerging. The matrix H_N^t for $N = 2^j$ can be decomposed as the product of three $N \times N$ matrices. The first and the last are block diagonal matrices while the middle matrix is a permutation matrix:

$$H_N^t = \begin{bmatrix} H_2^t & \\ & I_{N-2} \end{bmatrix} P_N \begin{bmatrix} H_{N/2}^t & 0_{N/2} \\ 0_{N/2} & H_{N/2}^t \end{bmatrix}.$$

Here I_{N-2} is the $(N-2)\times(N-2)$ identity, and P_N is an $N\times N$ permutation matrix given by a permutation of the columns of I_N. Denote by k_N the number of products required to compute multiplication by the matrix H_N^t.

Notice that the last matrix in the decomposition requires $2\times k_{N/2}$ multiplications, since it involves applying $H_{N/2}^t$ twice, and $k_{N/2}$ is the number of multiplications required to compute multiplication by $H_{N/2}^t$. The first matrix requires only 2 multiplications, since it involves applying H_2^t once and then the $(N-2)\times(N-2)$ identity matrix which involves no multiplications. The permutation matrix costs nothing in terms of multiplications. The total multiplication count for H_N^t obeys the following recursive equation

$$k_N = 2\times k_{N/2}+2.$$

Exercise 6.38. Show by induction that if the above decomposition holds for each $N=2^j$, then the number of multiplications required to apply H_N^t is $2(N-1)$. That is, $k_N=2(N-1)$. ♢

We return to the averages and differences, but now in dimension $N=2^j$. Starting with the vector $v\in\mathbb{C}^N$, we create two vectors a^{j-1} and d^{j-1}, each of dimension $N/2=2^{j-1}$, by taking averages and differences of pairs of consecutive entries. We now take the vector $a^{j-1}\in\mathbb{C}^{N/2}$ and create two vectors of dimension $N/4=2^{j-2}$ by the same procedure. We repeat the process until we reach dimension $1=2^0$. The process can be repeated j times, and can be represented by the following tree or *pyramid scheme*:

$$\begin{array}{ccccccccccc} v:=a^j & \to & a^{j-1} & \to & a^{j-2} & \to & \cdots & \to & a^1 & \to & a^0 \\ & \searrow & & \searrow & & \searrow & & \searrow & & \searrow & \\ & & d^{j-1} & & d^{j-2} & & \cdots & & d^1 & & d^0 \end{array}.$$

With this notation it can be seen that

$$H_N^t v = \begin{bmatrix} a^0 & d^0 & d^1 & \dots & d^{j-1}\end{bmatrix}^t.$$

Notice that the vectors a^{j-k} and d^{j-k} have dimension $2^{j-k}=N2^{-k}$, so the vector $\begin{bmatrix} a^0 & d^0 & d^1 & \dots & d^{j-1}\end{bmatrix}$ has the correct dimension

$$1+1+2+2^2+\cdots+2^{j-1}=2^j=N.$$

We will encounter averages and differences again when discussing wavelets in later chapters.

Exercise 6.39. Give a precise description of the permutation matrix P_N in the decomposition of H_N^t. ♢