DISCOVERING TRANSFORMS: A TUTORIAL ON CIRCULANT MATRICES, CIRCULAR CONVOLUTION, AND THE DISCRETE FOURIER TRANSFORM

BASSAM BAMIEH*

Key words. Discrete Fourier Transform, Circulant Matrix, Circular Convolution, Simultaneous Diagonalization of Matrices

AMS subject classifications. 42-01,15-01, 42A85, 15A18, 15A27

Abstract. How could the Fourier and other transforms be naturally discovered if one didn't know how to postulate them? In the case of the Discrete Fourier Transform (DFT), we show how it arises naturally out of analysis of circulant matrices. In particular, the DFT can be derived as the change of basis that simultaneously diagonalizes all circulant matrices. In this way, the DFT arises naturally from a linear algebra question about a set of matrices. Rather than thinking of the DFT as a signal transform, it is more natural to think of it as a single change of basis that renders an entire set of mutually-commuting matrices into simple, diagonal forms. The DFT can then be "discovered" by solving the eigenvalue/eigenvector problem for a special element in that set. A brief outline is given of how this line of thinking can be generalized to families of linear operators, leading to the discovery of the other common Fourier-type transforms.

1. Introduction. The Fourier transform in all its forms is ubiquitous. Its many useful properties are introduced early on in Mathematics, Science and Engineering curricula [1]. Typically, it is introduced as a transformation on functions or signals, and then its many useful properties are easily derived. Those properties are then shown to be remarkably effective in solving certain differential equations, or in analyzing the action of time-invariant linear dynamical systems, amongst many other uses. To the student, the effectiveness of the Fourier transform in solving these problems may seem magical at first, before familiarity eventually suppresses that initial sense of wonder. In this tutorial, I'd like to step back to before one is shown the Fourier transform, and ask the following question: How would one naturally *discover* the Fourier transform rather than have it be *postulated*?

The above question is interesting for several reasons. First, it is more intellectually satisfying to introduce a new mathematical object from familiar and well-known objects rather than having it postulated "out of thin air". In this tutorial we demonstrate how the DFT arises naturally from the problem of *simultaneous diagonalization* of all circulant matrices, which share symmetry properties that enable this diagonalization. It should be noted that simultaneous diagonalization of any class of linear operators or matrices is the ultimate way to understand their actions, by reducing the entire class to the simplest form of linear operations (diagonal matrices) simultaneously. The same procedure can be applied to discover the other close relatives of the DFT, namely the Fourier Transform, the z-Transform and Fourier Series. All can be arrived at by simultaneously diagonalizing a respective class of linear operators that obey their respective symmetry rules.

To make the point above, and to have a concrete discussion, in this tutorial we consider primarily the case of circulant matrices. This case is also particularly useful because it yields the DFT, which is the computational workhorse for all Fourier-type analysis. Given an n-vector $a := (a_0, \dots, a_{n-1})$, define the associated matrix C_a

*Department of Mechanical Engineering, University of California at Santa Barbara, *bamieh@ucsb.edu*. This work is partially supported by NSF Awards CMMI-1763064 and ECCS-1932777.

whose first column is made up of these numbers, and each subsequent column is obtained by a *circular shift* of the previous column

$$C_a := \begin{bmatrix} a_0 & a_{n-1} & a_{n-2} & \cdots & a_1 \\ a_1 & a_0 & a_{n-1} & & a_2 \\ a_2 & a_1 & a_0 & & a_3 \\ \vdots & & \ddots & \ddots & \vdots \\ a_{n-1} & a_{n-2} & a_{n-3} & \cdots & a_0 \end{bmatrix}. \tag{1.1}$$

Note that each row is also obtained from the pervious row by a circular shift. Thus the entire matrix is completely determined by any one of its rows or columns. Such matrices are called *circulant*. They are a subclass of Toeplitz matrices, and as mentioned, have very special properties due to their intimate relation to the Discrete Fourier Transform (DFT) and circular convolution.

Given an n-vector a as above, its DFT $\hat{a}$ is another n-vector defined by

$$\hat{a}_k := \sum_{l=0}^{n-1} a_l \; e^{-i\frac{2\pi}{n}kl}, \qquad k = 0, 1, \ldots, n-1. \tag{1.2}$$

A remarkable fact is that given a circulant matrix C_a, its eigenvalues are easily computed. They are precisely the set of complex numbers $\{\hat{a}_k\}$, i.e. the DFT of the vector a that defines the circulant matrix C_a. There are many ways to derive this conclusion and other properties of the DFT. Most treatments start with the definition (1.2) of the DFT, from which many of its seemingly magical properties are easily derived. To restate the goal of this tutorial, the question we ask here is: what if we didn't know the DFT? How can we arrive at it in a natural manner without needing someone to postulate (1.2) for us?

There is a natural way to think about this problem. Given a class of matrices or operators, one asks if there is a transformation, a change of basis, in which their matrix representations all have the same structure such as diagonal, block diagonal, or other special forms. The simplest such scenario is when a class of matrices can be *simultaneously diagonalized* with the same transformation. Since diagonalizing transformations are made up of eigenvectors of a matrix, then *a set of matrices is simultaneously diagonalizable iff they share a full set of eigenvectors*. An equivalent condition is that they each are diagonalizable, and they all mutually commute. Therefore given a mutually commuting set of matrices, by finding their shared eigenvectors, one finds that special transformation that simultaneously diagonalizes all of them. Thus, finding the "right transform" for a particular class of operators amounts to identifying the correct eigenvalue problem, and then calculating the eigenvectors, which then yield the transform.

An alternative but complementary view of the above procedure involves describing the class of operators using some underlying common symmetry. For example, circulant matrices such as (1.1) have a shift invariance property with respect to circular shifts of vectors. This can also be described as having a shift-invariant action on vectors over $\mathbb{Z}_n$ (the integers modulo n), which is also equivalent to having a shift-invariant action on periodic functions (with period n). In more formal language, circulant matrices represent a class of mutually commuting operators that also commute with the action of the group $\mathbb{Z}_n$. A basic shift operator generates that group, and the eigenvalue problem for that shift operator yields the DFT. This approach has the advantage of being generalizable to more complex symmetries that can be

encoded in the action of other, possibly non-commutative, groups. These techniques are part of the theory of group representations. However, we adopt here the approach described in the previous paragraph, which uses familiar Linear Algebra language and avoids the formalism of group representations. None the less, the two approaches are intimately linked. Perhaps the present approach can be thought of as a "gateway" treatment on a slippery slope to group representations [2, 3] if the reader is so inclined.

This tutorial follows the ideas described earlier. We first (section 2) investigate the simultaneous diagonalization problem for matrices, which is of interest in itself, and show how it can be done constructively. We then (section 3) introduce circulant matrices, explore their underlying geometric and symmetry properties, as well as their simple correspondence with circular convolutions. The general procedure for commuting matrices is then used (section 4) for the particular case of circulant matrices to simultaneously diagonalize them. The traditionally defined DFT emerges naturally out of this procedure, as well as other equivalent transforms. The "big picture" for the DFT is then summarized (section 5). A much larger context is briefly outlined in section 6, where the close relatives of the DFT, namely the Fourier transform, the z-transform and Fourier series are discussed. Those can be arrived at naturally by simultaneously "diagonalizing" families of mutually commuting linear operators. In this case, diagonalization has to be interpreted in a more general sense of conversion to so-called multiplication operators. Finally (subsection 6.2), an example of a non-commutative case is given where not diagonalization, but rather simultaneous *block-diagonalization* is possible. This serves as a motivation for generalizing classical Fourier analysis to so-called non-commutative Fourier analysis which is very much the subject of group representations.

2. Simultaneous Diagonalization of Commuting Matrices. The simplest matrices to study and understand are the diagonal matrices. They are basically uncoupled sets of scalar multiplications, essentially the simplest of all possible linear operations. When a matrix M can be diagonalized with a similarity transformation (i.e. $\Lambda = V^{-1}MV$, where Λ is diagonal), then we have a change of basis in which the linear transformation has that simple diagonal matrix representation, and its properties can be easily understood.

Often one has to work with a *set of transformations* rather than a single one, and usually with sums and products of elements of that set. If we require a different similarity transformation for each member of that set, then sums and products will each require finding their own diagonalizing transformation, which is a lot of work. It is then natural to ask if there exists one basis in which all members of a set of transformations have diagonal forms. This is the simultaneous diagonalization problem. If such a basis exists, then the properties of the entire set, as well as all sums and products (i.e. the algebra generated by that set) can be easily deduced from their diagonal forms.

Definition 2.1. *A set $\mathcal{M}$ of matrices is called* simultaneously diagonalizable *if there exists a single similarity transformation that diagonalizes all matrices in $\mathcal{M}$. In other words, there exists a single non-singular matrix V, such that for each $M \in \mathcal{M}$, the matrix*

$$V^{-1}MV \;=\; \Lambda \qquad \textit{is diagonal.}$$

It is immediate that all sums, products and inverses (when they exist) of elements of $\mathcal{M}$ will then also be diagonalized by this same similarity transformation. Thus a simultaneously diagonalizing transformation, when it exists, would be an invaluable tool in studying such sets of matrices.

When can a given set of matrices be simultaneously diagonalized? The answer is simple to state. First, they each have to be individually diagonalizable as an obvious necessary condition. Then, we will show that *a set of diagonalizable matrices can be simultaneously diagonalized iff they all mutually commute*. We will illustrate the argument in some detail since it gives a procedure for constructing the diagonalizing transformation. In the case of circulant matrices, this construction will yield the DFT. We note that the same construction also yields the z-transform, Fourier transform and Fourier series, but with some slight additional technicalities due to working with operators on infinite-dimensional spaces.

Necessity: We can see that commutativity is a necessary condition because all diagonal matrices mutually commute, and if two matrices are simultaneously diagonalizable, they do commute in the new basis, and therefore they must commute in the original basis. More precisely, let $A = V^{-1}\Lambda_a V$ and $B = V^{-1}\Lambda_b V$ be simultaneously diagonalizable with the transformation V, then

$$\begin{aligned} AB &= \big(V^{-1}\Lambda_a V\big)\big(V^{-1}\Lambda_b V\big) = V^{-1}\Lambda_a\Lambda_b V = V^{-1}\Lambda_b\Lambda_a V \\ &= \big(V^{-1}\Lambda_b V\big)\big(V^{-1}\Lambda_a V\big) = BA \end{aligned} \tag{2.1}$$

What about the converse? If two matrices commute, are they simultaneously diagonalizable? The answer is yes if both matrices are diagonalizable individually (this is of course a necessary condition). The argument is simple if one of the matrices has non-repeated eigenvalues. A little more care needs to be taken in the case of repeated eigenvalues since there are many diagonalizing transformations in that case. We will not need the more general version of this argument in this tutorial.

To begin, let's recap how one constructively diagonalizes a given matrix by finding its eigenvectors. If v_i is a an eigenvector of an $n \times n$ matrix A with corresponding eigenvalue λ_i then we have

$$Av_i \;=\; \lambda_i v_i, \qquad i = 1, \ldots, p, \tag{2.2}$$

where p is the largest number of linearly independent eigenvectors (which can be any number from 1 to n). The relations (2.2) can be compactly rewritten using *partitioned matrix notation* as a single matrix equation

$$\begin{bmatrix} Av_1 & \cdots & Av_p \end{bmatrix} = \begin{bmatrix} \lambda_1 v_1 & \cdots & \lambda_m v_p \end{bmatrix}$$

$$\Updownarrow$$

$$\begin{bmatrix} \quad A \quad \end{bmatrix} \begin{bmatrix} v_1 & \cdots & v_p \end{bmatrix} = \begin{bmatrix} v_1 & \cdots & v_p \end{bmatrix} \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_p \end{bmatrix} \quad \Leftrightarrow \quad AV = V\Lambda, \tag{2.3}$$

where V is a matrix whose columns are the eigenvectors of A, and Λ is the diagonal matrix made up of the corresponding eigenvalues of A.

We say that an $n \times n$ matrix has a *full set of eigenvectors* if it has n linearly independent eigenvectors. In that case, the matrix V in (2.3) is square and nonsingular and $\Lambda \;=\; V^{-1}AV$ is the diagonalizing similarity transformation. Of course not all matrices have a full set of eigenvectors. If the Jordan form of a matrix contains any non-trivial Jordan blocks, then it can't be diagonalized, and has strictly less than n linearly independent eigenvectors. We can therefore state that *a matrix is diagonalizable iff it has a full set of eigenvectors*, i.e. diagonalization is equivalent (in a constructive sense) to finding n linearly independent eigenvectors.

The case of simple (non-repeated) eigenvalues. Now consider the problem of simultaneous diagonalization. It is clear from the above discussion that two matrices can be simultaneously diagonalized iff they *share a full set of eigenvectors*. Consider the converse of the argument (2.1), and assume that A has (simple) non-repeated eigenvalues. This means that

$$Av_i \;=\; \lambda_i v_i, \qquad i = 1, \ldots, n, \quad \text{and } \lambda_i \neq \lambda_j \text{ if } i \neq j.$$

Consider any matrix B that commutes with A. Let B act on each of the eigenvectors by Bv_i and observe that

$$A\,\big(Bv_i\big) \;=\; B\;Av_i \;=\; B\;\lambda_i v_i \;=\; \lambda_i\;\big(Bv_i\big). \tag{2.4}$$

Thus Bv_i is an eigenvector of A with eigenvalue λ_i. Since those eigenvalues are distinct, and the corresponding eigenspace is one dimensional, Bv_i must be a scalar multiple of v_i

$$Bv_i \;=\; \gamma_i v_i.$$

Thus v_i is an eigenvector of B, but possibly with an eigenvalue γ_i different from λ_i. In other words, the eigenvectors of B are exactly the unique (up to scalar multiples) eigenvectors of A. We summarize this next.

LEMMA 2.2. *If a matrix A has simple eigenvalues, then A and B are simultaneously diagonalizable iff they commute. In that case, the diagonalizing basis is made up of the eigenvectors of A.*

This statement gives a constructive procedure for simultaneously diagonalizing a set $\mathcal{M}$ of mutually commuting matrices. If we can find one matrix $A \in \mathcal{M}$ with simple eigenvalues, then find its eigenvectors, those will yield the simultaneously diagonalizing transformation for the entire set. This is the procedure used for circulant matrices in section 4, where the "shift operator" S or its adjoint S^* play the role of the matrix with simple eigenvalues. The diagonalizing transformation for S^* yields the standard DFT. We will see that we can also produce other, equivalent versions of the DFT if we use eigenvectors of S instead, or eigenvectors of S^p with (p, n) co-prime.

3. Structural Properties of Circulant Matrices. The structure of circulant matrices is most clearly expressed using modular arithmetic. In some sense, modular arithmetic "encods" the symmetry properties of circulant matrices. We begin with a geometric view of modular arithmetic by relating it to rotations of roots of unity. We then show the "rotation invariance" of the action of circulant matrices, and finally connect that with circular convolution.

3.1. Modular Arithmetic, $\mathbb{Z}_n$, and Circular Shifts. To understand the symmetry properties of circulant matrices, it is useful to first study and establish some simple properties of the set $\mathbb{Z}_n := \{0, 1, \cdots, n-1\}$ of integers *modulo* n. The arithmetic in $\mathbb{Z}_n$ is *modular arithmetic*, that is, we say k *equals* l *modulo* n if $k - l$ is an integer multiple of n. The following notation can be used to describe this formally

$$\begin{array}{rcl} k & = & l \pmod n \\ \text{or } k & \equiv_n & l \end{array} \qquad \Longleftrightarrow \qquad \exists m \in \mathbb{Z}, \text{ s.t.} \quad k - l \;=\; m\,n$$

Thus for example $n \equiv_n 0$, and $n + 1 \equiv_n 1$ and so on. There are two equivalent ways to define (and think) about $\mathbb{Z}_n$, one mathematically formal and the other graphical. The first is to consider the set of all integers $\mathbb{Z}$ and regard any two integers k and

l such that $k - l$ is a multiple of n as equivalent, or more precisely as members of the same *equivalence class*. The infinite set of integers $\mathbb{Z}$ becomes a finite set of equivalence classes with this equivalence relation. This is illustrated in Figure 3.1a

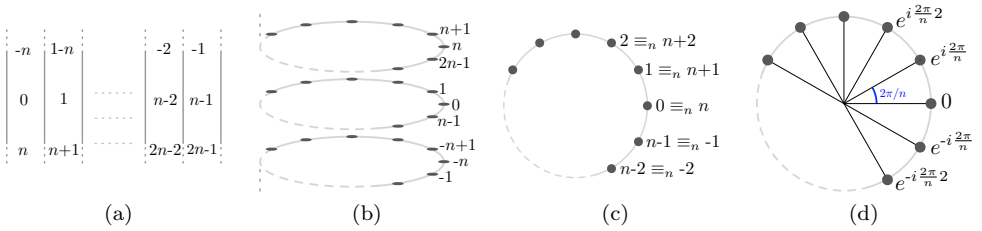


Fig. 3.1: (a) Definition of $\mathbb{Z}_n$ as the decomposition of the integers $\mathbb{Z}$ into *equivalence classes* each indicated as a "vertical bin". Two integers in $\mathbb{Z}$ belong to the same equivalence class (and represent the same element of $\mathbb{Z}_n$) if they differ by an integer multiple of n. (b) Another depiction of the decomposition where two integers that are vertically aligned in this figure belong to the same equivalence class. The arithmetic in $\mathbb{Z}_n$ is just angle addition in this diagram. For example, $(-1) + (n + 1) \equiv_n 0 \equiv_n n$. (c) Using the set $\{0, 1, \cdots, n - 1\}$ as $\mathbb{Z}_n$. A few equivalent members are shown, and the arithmetic of Z_n is just angle addition here. This can be thought of as the "top view" of (b). (d) The nth roots of unity $\rho_m := \exp\left(i\frac{2\pi}{n}m\right)$ lying on the unit circle in the complex plane. Identifying ρ_m with $m \in \mathbb{Z}_n$ shows that complex multiplication on $\{\rho_m\}$ (which corresponds to angle addition) is equivalent to modular addition in $\mathbb{Z}_n$.

where elements of $\mathbb{Z}_n$ are arranged in "vertical bins" which are the equivalence classes. Each equivalence class can be identified with any of its members. One choice is to identify the first one with the element 0, the second one with 1, and so on up to the n'th class identified with the integer $n - 1$. Figure 3.1c also shows how elements of $\mathbb{Z}_n$ can be arranged on a discrete circle so that the arithmetic in $\mathbb{Z}_n$ is identified with angle addition. One more useful isomorphism is between $\mathbb{Z}_n$ and the nth roots of unity $\rho_m := e^{i\frac{2\pi}{n}m}$, $m = 0, \ldots, n - 1$. The complex numbers $\{\rho_m\}$ lie on the unit circle each at a corresponding angle of $\frac{2\pi}{n}m$ counter-clockwise from the real axis (Figure 3.1d). Complex multiplication on $\{\rho_m\}$ corresponds to addition of their corresponding angles, and the mapping $\rho_m \to m$ is an isomorphism from complex multiplication on $\{\rho_m\}$ to modular arithmetic in $\mathbb{Z}_n$.

Using modular arithmetic, we can write down the definition of a circulant matrix (1.1) by specifying the kl'th entry[1] of the matrix C_a as

$$(C_a)_{kl} \;:=\; a_{k-l}, \quad k, l \in \mathbb{Z}_n, \tag{3.1}$$

where we use (mod n) arithmetic for computing $k - l$. It is clear that with this definition, the first column of C_a is just the sequence $a_0, a_1, \cdots, a_{n-1}$. The second column is given by the sequence $\{a_{k-1}\}$ and is thus $a_{-1}, a_0, \cdots, a_{n-2}$, which is exactly the sequence $a_{n-1}, a_0, \cdots, a_{n-2}$, i.e. a circular shift of the first column. Similarly each subsequent column is a circular shift of the column preceding it.

Finally, it is useful to visualize an n-vector $x := (x_0, \ldots, x_{n-1})$ as a set of numbers arranged at equidistant points along a circle, or equivalently as a *function* on the discrete circle. This is illustrated in Figure 3.2. Note the difference between this

[1] Here, and in this entire tutorial, matrix rows and columns are indexed from 0 to $n - 1$ rather than the more traditional 1 through n indexing. This alternative indexing significantly simplifies notation, and corresponds more directly to modular arithmetic.

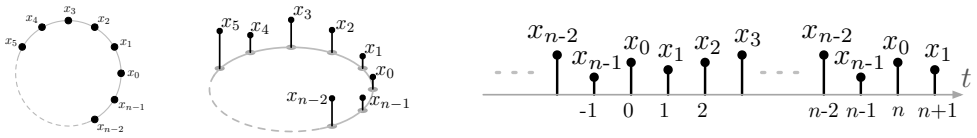


Fig. 3.2: A vector $x := (x_0, \ldots, x_{n-1})$ visualized as (a) a set of numbers arranged counter-clockwise on a discrete circle, or equivalently (b) as a function $x : \mathbb{Z}_n \longrightarrow \mathbb{C}$ on the discrete circle $\mathbb{Z}_n$. (c) An n-periodic function on the integers $\mathbb{Z}$ can equivalently be viewed as a function on $\mathbb{Z}_n$ as in (b).

figure and Figure 3.1, which depicts the elements of $\mathbb{Z}_n$ and modular arithmetic. Figure 3.2 instead depicts *vectors* as a set of numbers arranged in a discrete circle, or as *functions on* $\mathbb{Z}_n$. A function on $\mathbb{Z}_n$ can also be thought of as a *periodic* function (with period n) on the set of integers $\mathbb{Z}$ (Figure 3.2.c). In this case, periodicity of the function is expressed by the condition

$$x_{t+n} = x_t \quad \text{for all } t \in \mathbb{Z}. \tag{3.2}$$

It is however more natural to view periodic functions on $\mathbb{Z}$ as just functions on $\mathbb{Z}_n$. In this case, periodicity of the function is simply "encoded" in the modular arithmetic of $\mathbb{Z}_n$, and condition (3.2) does not need to be explicitly stated.

3.2. Symmetry Properties of Circulant Matrices. Amongst all circulant matrices, there is a special one. Let S and its adjoint S^* be the *circular shift operators* defined by the following action on vectors

$$\begin{aligned} S\ (x_0,\ \cdots,\ x_{n\text{-}2},\ x_{n\text{-}1}) &= \ (x_{n\text{-}1},\ x_0,\ \cdots,\ x_{n\text{-}2}) \\ S^*\ (x_0,\ x_1,\ \cdots,\ x_{n\text{-}1}) &= \ (x_1,\ \cdots,\ x_{n\text{-}1},\ x_0)\,. \end{aligned}$$

S is therefore called the *circular right-shift operator* while S^* is the *circular left-shift operator*. It is clear that S^* is the inverse of S, and it is easy to show that it is the adjoint of S. The latter fact also becomes clear upon examining the matrix representations of S and S^*

$$Sx = \begin{bmatrix} 0 & & & 1 \\ 1 & & & \\ & \ddots & \ddots & \\ & & 1 & 0 \end{bmatrix} \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_{n\text{-}1} \end{bmatrix} = \begin{bmatrix} x_{n\text{-}1} \\ x_0 \\ \vdots \\ x_{n\text{-}2} \end{bmatrix}, \quad S^*x = \begin{bmatrix} 0 & 1 & & \\ & \ddots & \ddots & \\ & & & 1 \\ 1 & & & 0 \end{bmatrix} \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_{n\text{-}1} \end{bmatrix} = \begin{bmatrix} x_1 \\ \vdots \\ x_{n\text{-}1} \\ x_0 \end{bmatrix},$$

which shows that S^* is indeed the transpose (and therefore the adjoint) of S. Note that both matrix representations are circulant matrices since $S = C_{(0,1,0,\ldots,0)}$ and $S^* = C_{(0,\ldots,0,1)}$ in the notation of (1.1). The actions of S and S^* expressed in terms of vector indices are

$$(Sx)_k \ := \ x_{k-1}, \qquad (S^*x)_k \ := \ x_{k+1}, \qquad\qquad k \in \mathbb{Z}_n, \tag{3.3}$$

where modular arithmetic is used for computing vector indices. For example $(Sx)_0 = x_{0-1} \equiv_n x_{n\text{-}1}$.

An important property of S is that it commutes with any circulant matrix. One way to see this is to observe the for any matrix M, left (right) multiplication by S

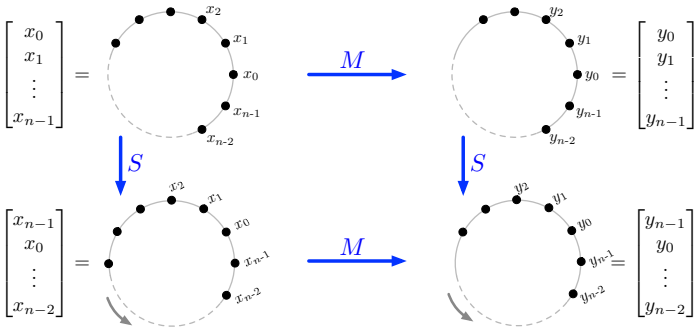


Fig. 3.3: Illustration of the *circular shift-invariance* property of the matrix-vector product $y = Mx$ in a commutative diagram. Sx is the circular shift of a vector, depicted also as a counter-clockwise rotation of the vector components arranged on the discrete circle. A matrix M has the shift invariance property if $SM = MS$. In this diagram, this means that the action of M on the rotated vector Sx is equal to acting on x with M (to produce $y = Mx$) first, and then rotating the resulting vector to yield Sy. A matrix M has this shift-invariance property iff it is circulant.

amounts to row (column) circular permutation. A brief look at the circulant structure in (1.1) shows that a row circular permutation gives the same matrix as a column circular permutation. Therefore, for any circulant matrix C_a, we have $SC_a = C_aS$. A more detailed argument is as follows.

To see this, note that the matrix representation of S implies its ij'th entry is given by $(S)_{ij} = \delta_{i-j-1}$. Now let C_a be any circulant matrix, and observe that

$$(SC_a)_{ij} = \sum_l S_{il}\,(C_a)_{lj} = \sum_l \delta_{i-l-1}\ a_{l-j} = \sum_l \delta_{(i-1)-l}\ a_{l-j} = a_{i-1-j},$$
$$(C_aS)_{ij} = \sum_l (C_a)_{il}\,S_{lj} = \sum_l a_{i-l}\ \delta_{l-j-1} = \sum_l a_{i-l}\ \delta_{l-(j+1)} = a_{i-j-1},$$

where (3.1) is used for the entries of C_a. Thus S commutes with any circulant matrix. The converse is also true (see Exercise Appendix A.1), and we state these conclusions in the next lemma.

LEMMA 3.1. *A matrix M is circulant iff it commutes with the circular shift operator S, i.e. $SM = MS$.*

Note a simple corollary that a matrix is circulant iff it commutes with S^* since

$$SM = MS \quad \Longleftrightarrow \quad S^*\,SM\,S^* = S^*\,MS\,S^* \quad \Longleftrightarrow \quad MS^* = S^*M,$$

which could be an alternative statement of the Lemma. The fact that a circulant matrix commutes with S could have been used as a definition of a circulant matrix, with the structure in (1.1) derived as a consequence. Commutation with S also expresses a *shift invariance* property. If we think of an n-vector x as a function on $\mathbb{Z}_n$ (Figure 3.2.b), then $SMx = MSx$ means that the action of M on x is shift invariant. Geometrically, Sx is a counter-clockwise rotation of the function x in Figure 3.2.b. $S\,(Mx) = M\,(Sx)$ means that rotating the result of the action of M on x is the same as rotating x first and then acting with M. This property is illustrated graphically in Figure 3.3.

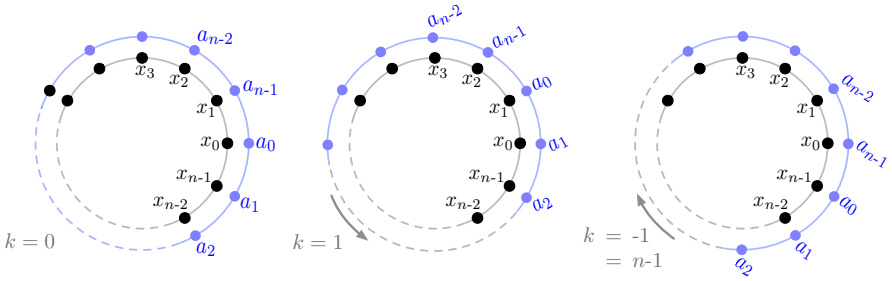


Fig. 3.4: Graphical illustration of circular convolution $y_k = \sum_{l=0}^{n-1} a_{k-l} x_l$ for $k = 0, 1, -1$ respectively. The a-vector is arranged in reverse orientation, and then each y_k is calculated from the dot product of x and the rotated, reverse-oriented a-vector rotated by k steps counter clockwise.

3.3. Circular Convolution. We will start with examining the matrix-vector product when the matrix is circulant. By analyzing this product, we will obtain the circular convolution of two vectors. Let C_a by some circulant matrix, and examine the action of such a matrix on any vector $x = (x_0,\ x_1,\ \cdots,\ x_{n-1})$. The matrix-vector multiplication $y = C_a x$ in detail reads

$$y \;=\; \begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_{n-1} \end{bmatrix} \;=\; \begin{bmatrix} a_0 & a_{n-1} & \cdots & a_1 \\ a_1 & a_0 & & a_2 \\ \vdots & & \ddots & \vdots \\ a_{n-1} & a_{n-2} & \cdots & a_0 \end{bmatrix} \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_{n-1} \end{bmatrix} = C_a\ x. \tag{3.4}$$

Using $(C_a)_{kl} = a_{k-l}$, this matrix-vector multiplication can be rewritten as

$$y_k \;=\; \sum_{l=0}^{n-1} (C_a)_{kl}\, x_l \;=\; \sum_{l=0}^{n-1} a_{k-l}\ x_l. \tag{3.5}$$

This can be viewed as an operation on the two vectors a and x to yield the vector y, and allows us to reinterpret the matrix-vector product of a circulant matrix as follows.

Definition 3.2. *Given two n-vectors a and x, their* circular convolution $y = a \star x$ *is another n-vector defined by*

$$y \;=\; a \star x \qquad \Leftrightarrow \qquad y_k \;=\; \sum_{l=0}^{n-1} a_{k-l}\ x_l, \tag{3.6}$$

where the indices in the sum are evaluated modulo n.

Comparing (3.5) with (3.6), we see that multiplying a vector by a circulant matrix is equivalent to convolving the vector with the vector defining the circulant matrix

$$y \;=\; C_a\ x \;=\; a \star x. \tag{3.7}$$

The sum in (3.6) defining circular convolution has a nice circular visualization due to modular arithmetic on $\mathbb{Z}_n$. This is illustrated in Figure 3.4. The elements of x are arranged in a discrete circle counter-clockwise, while the elements of a are arranged in a circle clockwise (the reverse orientation is because elements of x are

indexed like x_l while those of a are indexed like $a_{\cdot -l}$ in the definition (3.6)). For each k, the array a is rotated counter-clockwise by k steps (Figure 3.4 shows cases for three different values of k). The number y_k in (3.6) is then obtained by multiplying the x and rotated a arrays element-wise, and then summing. This generates the n numbers $y_0, \ldots, y_{n-1}$.

From the definition, it is easy to show (see Exercise Appendix A.3) that circular convolution is associative and commutative.

- *Associativity:* for any three n-vectors a, b and c we have

$$a \star (b \star c) \;=\; (a \star b) \star c$$

- *Commutativity:* for any two n-vectors a and b

$$a \star b \;=\; b \star a$$

The above two facts have several interesting implications. First, since convolution is commutative, the matrix-vector product (3.4) can be written in two equivalent ways

$$C_a\; x \;=\; a \star x \;=\; x \star a \;=\; C_x\; a.$$

Applying this fact in succession to two circulant matrices

$$C_b C_a\; x \;=\; C_b\,(a \star x) \;=\; b \star (a \star x) \;=\; (b \star a) \star x \;=\; C_{b \star a}\; x.$$

This means that the product of any two circulant matrices C_b and C_a is another circulant matrix $C_{b \star a}$ whose defining vector is $b \star a$, the circular convolution of the defining vectors of C_b and C_a respectively. We summarize this conclusion and an important corollary of it next.

THEOREM 3.3. 1. *Circular convolution of any two vectors can be written as a matrix-vector product with a circulant matrix*

$$a \star x \;=\; C_a\; x \;=\; C_x\; a.$$

2. *The product of any two circulant matrices is another circulant matrix*

$$C_a C_b \;=\; C_{a \star b}.$$

3. *All circulant matrices mutually commute since for any two C_a and C_b*

$$C_a C_b \;=\; C_{a \star b} \;=\; C_{b \star a} \;=\; C_b C_a.$$

The set of all n-vectors forms a commutative algebra under the operation of circular convolution. The above shows that the set of $n \times n$ circulant matrices under standard matrix multiplication is also a commutative algebra isomorphic to n-vectors with circular convolution.

4. Simultaneous Diagonalization of all Circulant Matrices Yields the DFT . In this section, we will *derive* the DFT as a byproduct of diagonalizing circulant matrices. Since all circulant matrices mutually commute, we recall Lemma 2.2 and look for a circulant matrix that has simple eigenvalues. The eigenvectors of that matrix will then give the simultaneously diagonalizing transformation.

The shift operator is in some sense the most fundamental circulant matrix, and is therefore a good candidate for an eigenvector/eigenvalue decomposition. The eigenvalue problem for S will turn out to be the simplest one. Note that we have two options. To find eigenvectors of S or alternatively of S^*. We begin with S^* since this will end up yielding the classically defined DFT.

4.1. Construction of Eigenvectors/Eigenvalues of S^*. Let w be an eigenvector (with eigenvalue λ) of the shift operator S^*. Note that it is also an eigenvector (with eigenvalue λ^l) of any power $(S^*)^l$ of S^*. Applying the definition (3.3) to the relation $S^*w = \lambda w$ will reveal that an eigenvector w has a very special structure

$$\begin{array}{rcl} S^*w &=& \lambda w \\ (S^*)^l w &=& \lambda^l w \end{array} \quad \Longleftrightarrow \quad \begin{array}{rcll} w_{k+1} &=& \lambda\, w_k, & k \in \mathbb{Z}_n, \\ w_{k+l} &=& \lambda^l\, w_k, & k \in \mathbb{Z}_n,\ l \in \mathbb{Z}, \end{array} \tag{4.1}$$

i.e. each entry w_{k+1} of w is equal to the previous entry w_k multiplied by the eigenvalue λ. These relations can be used to compute all eigenvectors/eigenvalues of S^*. First, observe that although (4.1) is valid for all $l \in \mathbb{Z}$, this relation "repeats" for $l \geq n$. In particular, for $l = n$ we have for each index k

$$w_{k+n} \;=\; \lambda^n w_k \quad \Longleftrightarrow \quad w_k \;=\; \lambda^n w_k \tag{4.2}$$

since $k + n \equiv_n k$. Now since the vector $w \neq 0$, then for at least one index k, $w_k \neq 0$, and the last equality implies that $\lambda^n = 1$, i.e. any eigenvalue of S must be an *nth root of unity*

$$\boxed{\lambda^n = 1 \quad \Longleftrightarrow \quad \lambda \;=\; \rho_m \;:=\; e^{i\frac{2\pi}{n}m}, \quad m \in \mathbb{Z}_n.}$$

Thus we have discovered that the n eigenvalues of S^* are precisely the n distinct nth roots of unity $\{\rho_m,\ m = 0, \ldots, n-1\}$. Note that any of the nth roots of unity can be expressed as a power of the first nth root: $\rho_m = \rho_1^m$ (recall Figure 3.1d).

Now fix $m \in \mathbb{Z}_n$ and compute $w^{(m)}$, the eigenvector corresponding to the eigenvalue ρ_m. Apply the last relation in (4.1) $w_{k+l} = \lambda^l w_k$, and use it to express the entries of the eigenvector $w^{(m)}$ in terms of the first entry $(k = 0)$

$$w^{(m)}_{l+0} = \lambda^l w_0 \quad \Leftrightarrow \quad w^{(m)}_l = \rho_m^l w_0 \quad \Leftrightarrow \quad w^{(m)} = w_0 \left(1,\ \rho_m,\ \rho_m^2,\ \ldots,\ \rho_m^{n\text{-}1}\right). \tag{4.3}$$

Note that w_0 is a scalar, and since eigenvectors are only unique up to multiplication by a scalar, we can set $w_0 = 1$ for a more compact expression for the eigenvector. In addition, ρ_m in (4.3) could be any of the nth roots of unity, and thus that expression applies to all of them, yielding the n eigenvectors. We summarize the previous derivations in the following statement.

LEMMA 4.1. *The circular left-shift operator S^* on $\mathbb{R}^n$ has n distinct eigenvalues. They are the nth roots of unity $\rho_m := e^{i\frac{2\pi}{n}m} = \rho_1^m =: \rho^m$, $m \in \mathbb{Z}_n$. The corresponding eigenvectors are*

$$w^{(m)} \;=\; \left(1,\ \rho^m,\ \rho^{2m},\ \ldots,\ \rho^{m(n\text{-}1)}\right), \qquad m = 0, \ldots, n-1, \tag{4.4}$$

Note that the eigenvectors $\left\{w^{(m)}\right\}$ are indexed with the same index as the eigenvalues $\{\lambda_m = \rho_m\}$. It is useful and instructive to visualize the eigenvalues and their corresponding eigenvectors as specially ordered sets of the roots of unity. Which roots of unity enter into any particular eigenvector, as well as their ordering, is determined by the algebra of rotations of roots of unity. This is illustrated in detail in Figure 4.1

4.2. Eigenvalues Calculation of a Circulant Matrix Yields the DFT. Now that we have calculated all the eigenvectors of the shift operator in Lemma 4.1, we can use them to find the eigenvalues of any circulant matrix C_a. Recall that since any circulant matrix commutes with S^*, and S^* has distinct eigenvalues, then C_a has

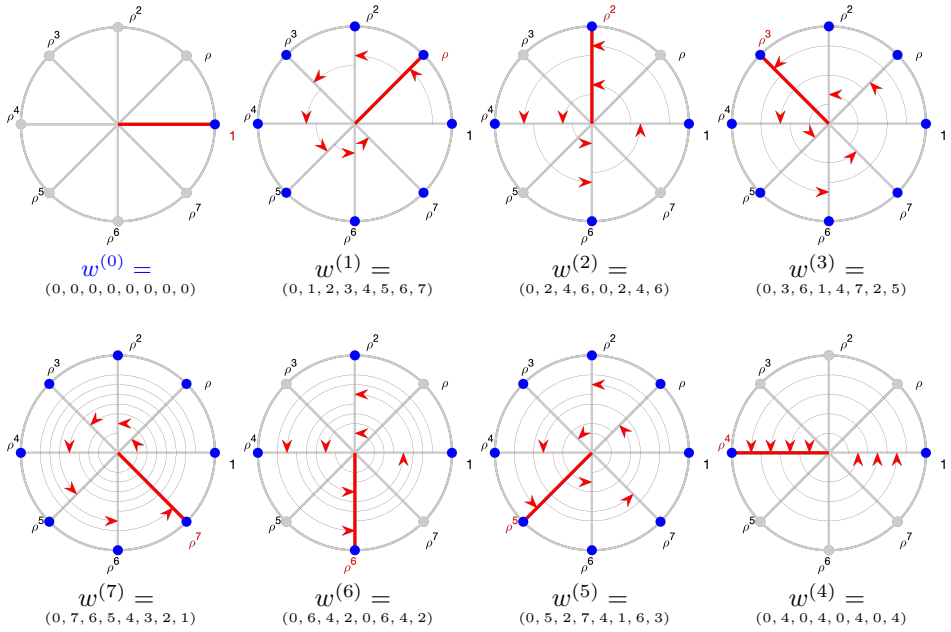


Fig. 4.1: Visualization of the eigenvalues and eigenvectors of the left shift operator S^* for the case $n = 8$. The eigenvalues (red straight lines and red labels), and elements of the corresponding eigenvectors (blue dots) are all points on the unit circle of the complex plane. ρ is the n'th root of unity (here $\rho = e^{i\pi/4}$). For each $m \in \mathbb{Z}_n$, ρ^m is an eigenvalue with eigenvector $w^{(m)} = \left(1, \rho^m, \rho^{2m}, \ldots, \rho^{m(n-1)}\right)$, where each element is a rotation of the previous element by ρ^m (curvy red arrows). For compactness of notation, *vectors are denoted above by powers of ρ, e.g.* $w^{(4)} = \left(1, \rho^2, \rho^4, \rho^6, 1, \rho^2, \rho^4, \rho^6\right) = (0, 2, 4, 6, 0, 2, 4, 6)$. Notice the pattern that which powers of ρ appear in $w^{(m)}$ depends on the least common factor (lcf) of m and n, e.g. in $w^{(4)}$ that number is $8/\text{lcf}(4, 8) = 2$. For (m, n) co-prime, all powers of ρ appear in $w^{(m)}$, though with permuted ordering (see $w^{(1)}, w^{(3)}, w^{(5)}, w^{(7)}$).

the same eigenvectors as those (4.4) previously found for S^* (by Lemma 2.2). Thus we have the relation

$$(4.5)\quad C_a\ w^{(m)} = \lambda_m\ w^{(m)} \quad \Leftrightarrow \quad \begin{bmatrix} a_0 & a_{n-1} & \cdots & a_1 \\ a_1 & a_0 & & a_2 \\ \vdots & & \ddots & \vdots \\ a_{n-1} & a_{n-2} & \cdots & a_0 \end{bmatrix} \begin{bmatrix} 1 \\ \rho_m \\ \vdots \\ \rho_m^{n-1} \end{bmatrix} = \lambda_m \begin{bmatrix} 1 \\ \rho_m \\ \vdots \\ \rho_m^{n-1} \end{bmatrix},$$

where $\{\lambda_m\}$ are the eigenvalues of C_a (not the eigenvalues of S^* found in the previous section). Each row of the above equation represent essentially the same equation (but multiplied by a power of ρ_m). The first row is the easiest equation to work with

$$\begin{aligned} \lambda_m &= a_0 + a_{n-1}\ \rho_m + \cdots + a_1\ \rho_m^{n-1} \\ &= a_0 + a_1\ \rho_m^{-1} + \cdots + a_{n-1}\ \rho_m^{-(n-1)} \\ &= \sum_{l=0}^{n-1} a_l\ \rho_m^{-l} \ = \ \sum_{l=0}^{n-1} a_l\ \rho^{-ml} \ = \ \boxed{\sum_{l=0}^{n-1} a_l\ e^{-i\frac{2\pi}{n}ml} \ =: \ \hat{a}_m,} \end{aligned} \tag{4.6}$$

which is precisely the classically-defined DFT (1.2) of the vector a.

We therefore conclude that any circulant matrix C_a is diagonalizable by the basis (4.4). Its n eigenvalues are given by $(\hat{a}_0, \hat{a}_1, \ldots, \hat{a}_{n-1})$ from (4.6), which is the DFT of the vector $(a_0, a_1, \ldots, a_{n-1})$. In this way, the DFT arises from a formula for computing the eigenvalues of any circulant matrix.

One might ask what the conclusion would have been if the eigenvectors of S have been used instead of those of S^*. A repetition of the previous steps but now for the case of S would yield that the eigenvalues of a circulant matrix C_a are given by

$$\mu_k \;=\; \sum_{l=0}^{n-1} a_l \; e^{i\frac{2\pi}{n}kl}, \qquad\qquad k = 0, 1, \ldots, n-1. \tag{4.7}$$

While the expressions (4.6) and (4.7) may at first appear different, the sets of numbers $\{\lambda_m\}$ and $\{\mu_k\}$ are actually equal. So in fact, the expression (4.7) gives the same set of eigenvalues as (4.6) but arranged in a different order since $\mu_k = \lambda_{-k}$.

$$\lambda_{-k} \;=\; \sum_{l=0}^{n-1} a_l \; e^{-i\frac{2\pi}{n}(-k)l} \;=\; \sum_{l=0}^{n-1} a_l \; e^{i\frac{2\pi}{n}kl} \;=\; \mu_k.$$

Along with the two choices of S and S^*, there are also other possibilities. Let p be any number that is coprime with n. It is easy to show (Exercise Appendix A.2) that a $n \times n$ matrix is circulant iff it commutes with S^p. In addition, the eigenvalues of S^p are distinct (see Figure 4.1). Therefore the eigenvectors of S^p (rather than those of S) can be used to simultaneously diagonalize all circulant matrices. This would yield yet another transform distinct from the two transforms (4.6) or (4.7). However, the set of numbers produced from that transform will still be the same as those computed from the previous two transforms, but arranged in a different ordering.

5. The Big Picture. Let C_a be a circulant matrix made from a vector a as in (1.1). If we use the eigenvectors (4.4) of S^* as columns of a matrix W, the n eigenvalue/eigenvector relationships (4.5) $C_a w^{(m)} = \lambda_m w^{(m)}$ can be written as a single matrix equation as follows

$$\begin{aligned} C_a \begin{bmatrix} w^{(0)} & \vdots & \cdots & \vdots & w^{(n\text{-}1)} \end{bmatrix} &= \begin{bmatrix} w^{(0)} & \vdots & \cdots & \vdots & w^{(n\text{-}1)} \end{bmatrix} \begin{bmatrix} \hat{a}_0 & & & \\ & \ddots & & \\ & & & \hat{a}_{n-1} \end{bmatrix}, \\ \iff \qquad C_a\, W &= W\, \operatorname{diag}(\hat{a}), \end{aligned} \tag{5.1}$$

where we have used the fact (4.6) that the eigenvalues of C_a are precisely $\{\hat{a}_m\}$, the elements of the DFT of the vector a.

It is easy to verify that the columns of W are mutually orthogonal[2], and thus W is a unitary matrix (up to a rescaling) $W^*W = WW^* = nI$, or equivalently $W^{-1} = \frac{1}{n}W^*$. Since the matrix W is made up of the eigenvectors of S^*, which in turn are made up of various powers of the roots of unity (4.4), it has some special structure which is worth examining

$$W := \begin{bmatrix} w^{(0)} & \vdots & \cdots & \vdots & w^{(n\text{-}1)} \end{bmatrix} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & \rho & \cdots & \rho^{n\text{-}1} \\ \vdots & \vdots & & \vdots \\ 1 & \rho^{n\text{-}1} & \cdots & \rho^{(n\text{-}1)(n\text{-}1)} \end{bmatrix}.$$

[2]This also follows from the fact that the columns of W are the eigenvectors of S^*, and since S^* is a normal matrix, it has mutually orthogonal eigenvectors.

The matrix W is symmetric, W^* is thus the matrix W with each entry replaced by its complex conjugate. Furthermore, since for each root of unity $\left(\rho^k\right)^* = \rho^{-k}$, we can therefore write

$$W^* = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & \rho^{-1} & \cdots & \rho^{-(n-1)} \\ \vdots & \vdots & & \vdots \\ 1 & \rho^{-(n-1)} & \cdots & \rho^{-(n-1)(n-1)} \end{bmatrix}.$$

Also observe that multiplying a vector by W^* is exactly taking its DFT. Indeed the m'th row of W^*x is

$$\hat{x}_m \;=\; \begin{bmatrix} 1 & \rho^{-m} & \cdots & \rho^{-m(n-1)} \end{bmatrix} \begin{bmatrix} x_0 \\ \vdots \\ x_{n-1} \end{bmatrix},$$

which is exactly the definition (1.2) of the DFT. Similarly, multiplication by $\frac{1}{n}W$ is taking the inverse DFT

$$x_l \;=\; \frac{1}{n}\sum_{k=0}^{n-1} \hat{x}_k\, \rho^{kl} \;=\; \frac{1}{n}\sum_{k=0}^{n-1} \hat{x}_k\, e^{i\frac{2\pi}{n}kl}.$$

Multiplying both sides of (5.1) from the right by W^{-1} gives the diagonalization of C_a which can be written in several equivalent forms

$$\begin{aligned} C_a &= W\, \mathrm{diag}\,(\hat{a})\, W^{-1} \;=\; W\, \mathrm{diag}\,(\hat{a})\, \left(\tfrac{1}{n}W^*\right) \\ &= \left(\tfrac{1}{n}W\right)\, \mathrm{diag}\,(\hat{a})\, W^* \;=\; \left(\tfrac{1}{\sqrt{n}}W\right)\, \mathrm{diag}\,(\hat{a})\, \left(\tfrac{1}{\sqrt{n}}W^*\right). \end{aligned} \tag{5.2}$$

The diagonalization (5.2) can be interpreted as follows in terms of the action of a circulant matrix C_a on any vector x

$$C_a\, x \;=\; \underbrace{\left(\tfrac{1}{n}W\right)\, \underbrace{\mathrm{diag}\,(\hat{a})\, \underbrace{W^*\, x}_{\text{DFT of } x}}_{\text{multiply by } \hat{a} \text{ entrywise}}}_{\text{inverse DFT}}$$

Thus the action of C_a on x, or equivalently the circular convolution of a with x, can be performed by first taking the DFT of x, then multiplying the resulting vector component-wise by $\hat{a}$ (the DFT of the vector a defining the matrix C_a), and then taking an inverse DFT. In other words, the diagonalization of a circulant matrix is equivalent to converting circular convolution to component-wise vector multiplication through the DFT. This is illustrated in Figure 5.1.

Note that in the literature there is an alternative form for the DFT and its inverse

$$\hat{x}_k \;=\; \frac{1}{\sqrt{n}}\sum_{l=0}^{n-1} x_l\, e^{-i\frac{2\pi}{n}kl}, \qquad\qquad x_l \;=\; \frac{1}{\sqrt{n}}\sum_{k=0}^{n-1} \hat{x}_k\, e^{i\frac{2\pi}{n}kl},$$

which is sometimes preferred due to its symmetry (and is also truly unitary since with this definition $\|x\|_2 = \|\hat{x}\|_2$). This "unitary" DFT corresponds to the last diagonalization given in (5.2). We do not adopt this unitary DFT definition here since it complicates[3] the statement that the eigenvalues of C_a are precisely the entries of $\hat{a}$.

[3] If the unitary DFT is adopted, the equivalent statement would be that the eigenvalues of C_a are the elements of the entries of $\sqrt{n}\,\hat{a}$.

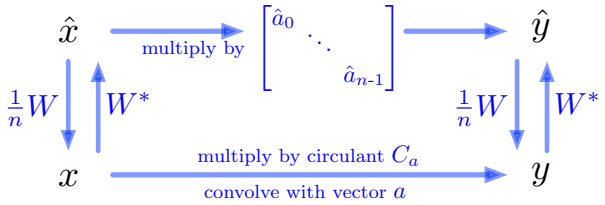


Fig. 5.1: Illustration of the relationships between circulant matrices, circular convolution and the DFT. The matrix-vector multiplication $y = C_a x$ with the circulant matrix C_a is equivalent to the circular convolution $y = a \star x$. The DFT is a linear transformation W^* on vectors with inverse $\frac{1}{n}W$. It converts multiplication by the circulant matrix C_a into multiplication by the diagonal matrix $\mathrm{diag}\,(\hat{a}_0, \ldots, \hat{a}_{n-1})$ whose entries are the DFT of the vector a defining the matrix C_a.

We summarize the algebraic aspects of the big picture in the following theorem.

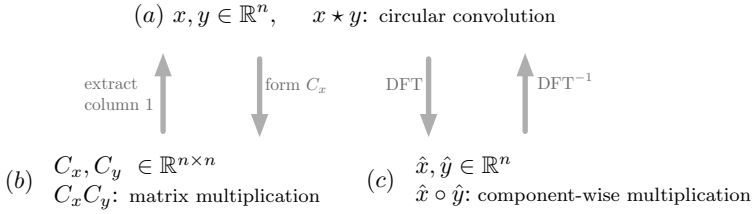
THEOREM 5.1. *The following sets are isomorphic commutative algebras*

(a) The set of n-vectors is closed under circular convolutions and is thus an algebra with the operations of addition and convolution.

(b) The set of $n \times n$ circulant matrices is an algebra under the operations of addition and matrix multiplication.

(c) The set of n-vectors is an algebra under the operations of addition and component-wise multiplication.

The above isomorphisms are depicted by the following diagram



6. Further Comments and Generalizations. We end by briefly sketching two different ways in which the procedures described in this tutorial can be generalized. The first is generalizations to families of mutually commuting infinite matrices and linear operators. These families are characterized by commuting with shifts of functions defined on "time-axes" which can be identified with groups or semi-groups. This yields the familiar Fourier transform, Fourier series, and the z-transform. A second line of generalization is to families of matrices that do not commute. In this case we can no longer demand simultaneous diagonalization, but rather simultaneous *block diagonalization* whenever possible. This is the subject of group representations, but we will only touch on the simplest of examples by way of illustration. The discussions in this section are meant to be brief sketches to motivate the interested reader into further exploration of the literature.

6.1. Fourier Transform, Fourier Series, and the z-Transform. First we recap what these classical transforms are. They are summarized in Table 6.1. In a Signals and Systems course [1], these concepts are usually introduced as transforms on temporal signals, so we will use that language to refer to the independent variable as time, although it can have any other interpretation. As is the theme of this tutorial, the starting point should not be the signal transform, but rather the systems,

or operators, that act on them and their respective invariance properties. We now formalize these properties.

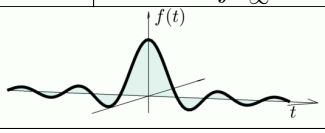
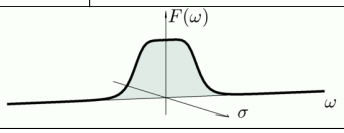
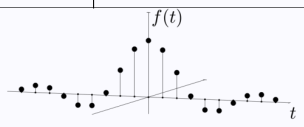
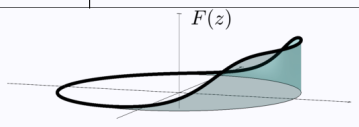
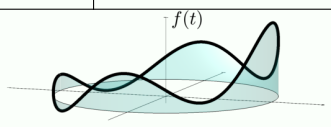
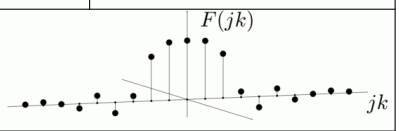
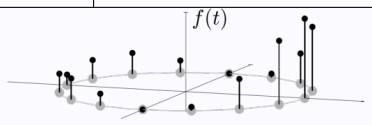
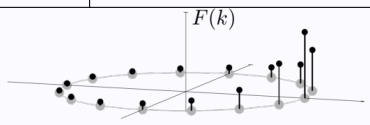
Time Axis	**Transform**	**Frequency Axis (Frequency "Set")**
$t \in \mathbb{R}$	Fourier Transform $F(\omega) := \int_{-\infty}^{\infty} f(t)\ e^{-j\omega t} dt$	$j\omega \in j\mathbb{R}$ *imaginary axis of* $\mathbb{C}$
		
$t \in \mathbb{Z}$	z-Transform (bilateral) $F(z) := \sum_{t\in\mathbb{Z}} f(t)\ z^{-t}$	$z = e^{j\theta} \in \mathbb{T}$ *unit circle of* $\mathbb{C}$
		
$t \in \mathbb{T}$	Fourier Series $F(k) := \int_{0}^{2\pi} f(t)\ e^{-jkt} dt$	$jk \in j\mathbb{Z}$, *integers of imaginary axis of* $\mathbb{C}$
		
$t \in \mathbb{Z}_n$	Discrete Fourier Transform (DFT) $F(k) := \sum_{t=0}^{n-1} f(t)\ e^{-j\frac{2\pi}{n}kt}$	$k \in \mathbb{Z}_n$ *n-roots of unity*
		

Table 6.1: A list of Fourier-type transforms. The left column lists the time axis over which a signal is defined, the middle column lists the common name and expression for the transform, and the right column lists the frequency axis, or more precisely the "set of frequencies" associated with that transform. The set of frequencies is typically identified with a subset of the complex plane $\mathbb{C}$. Note how all the transforms have the common form of integrating or summing the given signal against a function of the form e^{-st}, where the set of values ("frequencies") of $s \in \mathbb{C}$ is different for different transforms.

The time axes are the integers $\mathbb{Z}$ for discrete time and the reals $\mathbb{R}$ for continuous time. Moreover, the discrete circle $\mathbb{Z}_n$ and the continuous circle $\mathbb{T}$ are the time axes for discrete and continuous-time periodic signals respectively. A common feature of the time axes $\mathbb{Z}$, $\mathbb{R}$, $\mathbb{T}$ and $\mathbb{Z}_n$ is that they all are *commutative groups*. In fact, they are the basic commutative groups. All other (so-called locally compact) commutative groups are made up of group products of those basic four [4].

Let's denote by t, τ, T elements of those groups $\mathbb{G} = \mathbb{R},\ \mathbb{Z},\ \mathbb{T},$ or $\mathbb{Z}_n$. For any function $f : \mathbb{G} \to \mathbb{R}$ (or $\mathbb{C}$) defined on such a group, there is a natural "time shift" operation

$$(S_T f)(t) \ := \ f(t-T), \tag{6.1}$$

which is the right shift (delay) of f by T time units. All that is needed to make sense of this operation is that for $t, T \in \mathbb{G}$, we have $t - T \in \mathbb{G}$, and that is guaranteed by the group structure for any of those four time sets. Now consider a linear operator $A : \mathbb{C}^{\mathbb{G}} \longrightarrow \mathbb{C}^{\mathbb{G}}$ acting on the vector space $\mathbb{C}^{\mathbb{G}}$ of all scalar-valued functions on $\mathbb{G}$, which can be any of the four time sets. We call such an operator *time invariant* (or *shift invariant*) if it commutes with all possible time-shift operations (6.1), i.e.

$$\forall T \in \mathbb{G}, \quad S_T A \ = \ A S_T. \tag{6.2}$$

To conform with traditional terminology, we refer to such shift-invariant linear operators as *Linear Time-Invariant (LTI) systems*. They are normally described as differential or difference equations with a forcing term, or as convolutions of signals amongst other representations. However, only the shift-invariance property (6.2), and not the details of those representations, is what's important in discovering the appropriate transform that simultaneously diagonalizes such operators.

There are additional technicalities in generalizing the previous techniques to sets of *linear operators* on infinite dimensional spaces rather than matrices. The procedure however is very similar. We identify the class of operators to be analyzed. This involves a shift (time) invariance property, which then implies that they all mutually commute. The "eigenvectors" of the shift operators give the simultaneously diagonalizing transform. The complication here is that eigenvectors may not exists in the classical sense (they do in the case of Fourier series, but not in the other cases). In addition, diagonalization will not necessarily correspond to finding a new basis of the vector space. In both the Fourier and z-transforms, the number of "linearly independent eigenfunctions" is not even countable, so they can't be thought of as forming a basis. Fortunately, it is easy to circumvent these difficulties by generalizing the concept of diagonal matrices to *multiplication operators*. For linear operators on infinite-dimensional spaces, these play the same role as the diagonal matrices do on finite-dimensional spaces.

DEFINITION 6.1. *Let Ω be some set, and consider the vector space $\mathbb{C}^{\Omega}$ of all scalar-valued functions on Ω. Given a particular scalar-valued function $a : \Omega \longrightarrow \mathbb{C}$, we define the associated* multiplication operator $M_a : \mathbb{C}^{\Omega} \longrightarrow \mathbb{C}^{\Omega}$ *by*

$$\forall x \in \Omega, \qquad (M_a f)(x) \ := \ a(x)\ f(x),$$

i.e. the point-wise multiplication of f by a.

If $\Omega = \{1, \ldots, n\}$, then $\mathbb{C}^{\Omega} = \mathbb{C}^n$, the space of all complex n-vectors, and M_a is simply represented by the diagonal matrix whose entries are made up of the entries of the vector a. The concept introduced above is however much more general. Note that it is immediate from the definition that all multiplication operators mutually commute, just like all diagonal matrices mutually commute.

Now we generalize the concept of diagonalizing matrices. Diagonalizing an operator, when possible, is done by converting it to a multiplication operator.

DEFINITION 6.2. *A linear operator $A : \mathcal{H} \longrightarrow \mathcal{H}$ on a vector space $\mathcal{H}$ is said to be* diagonalizable *if there exists a function space $\mathbb{C}^\Omega$, and an invertible transformation $V : \mathbb{C}^\Omega \longrightarrow \mathcal{H}$ that converts A into a multiplication operator M_a*

$$VAV^{-1} \;=\; M_a,$$

for some function $a : \Omega \longrightarrow \mathbb{C}$. The function a is referred to as the symbol *of the operator A, and V is the* diagonalizing *transformation.*

Thus in contrast to the case of matrices, we may have to move to a different vector space to diagonalize a general linear operator.

In some cases, we can still give a diagonalization an interpretation as a basis expansion in the same vector space. It provides a helpful contrast to consider such an example. Let an operator $A : \mathcal{H} \to \mathcal{H}$ have a countable set of eigenfunctions $\{v_m\}$ that span a Hilbert space $\mathcal{H}$. Assume in addition that A is normal, and thus the eigenfunctions are mutually orthonormal. An example of this situation the case of shift invariant operators on $\mathbb{T}$, which corresponds to the 3rd entry in Table 6.1 (i.e. Fourier series). In that case we can take $\Omega = \mathbb{Z}$, and thus $\mathbb{C}^\mathbb{Z}$ is the space of all complex-valued bilateral sequences. In addition we can add a Hilbert space structure by using ℓ^2 norms and consider $\ell^2(\mathbb{Z})$ as the space of sequences. The diagonalizing transformation[4] is $V : \ell^2(\mathbb{Z}) \longrightarrow L^2(\mathbb{T})$ described as follows. Let $v_k(t) := e^{jkt}/\sqrt{2\pi}$ be the Fourier series elements. They are an orthonormal basis of $L^2(\mathbb{T})$. Consider any square-integrable function $f \in L^2(\mathbb{T})$ with Fourier series

$$F_k := \langle v_k, f\rangle = \frac{1}{\sqrt{2\pi}} \int_0^{2\pi} e^{-jkt}\; f(t)\; dt, \qquad f(t) = \sum_{k\in\mathbb{Z}} F_k\; v_k(t) = \frac{1}{\sqrt{2\pi}} \sum_{k\in\mathbb{Z}} F_k e^{jkt}.$$

The mapping $V : \ell^2(\mathbb{Z}) \longrightarrow L^2(\mathbb{T})$ then simply maps each function f to its bilateral Fourier series coefficients

$$\{\ldots, F_{\text{-}1}, F_0, F_1, \ldots\} \quad \leftrightarrow \quad f = \sum_{k\in\mathbb{Z}} F_k v_k.$$

Plancherel's theorem guarantees that this mapping is a bijective isometry. Any shift-invariant operator A on $L^2(\mathbb{T})$ (e.g. those that can be written as circular convolutions) then has a diagonalization as a doubly infinite matrix

$$VAV^{-1} \;=\; \begin{bmatrix} \ddots & & & & \\ & \hat{a}_{-1} & & & \\ & & \hat{a}_0 & & \\ & & & \hat{a}_{\text{-}1} & \\ & & & & \ddots \end{bmatrix},$$

where the sequence $\{\hat{a}_k\}$ is made up of the eigenvalues of A.

The example (Fourier series) just discussed is a very special case. In general, we have to consider diagonalizing using a multiplication operator on an uncountable domain. Examples of these are the Fourier and z-transforms, where the diagonalizations are multiplication operators on $\mathbb{C}^\mathbb{R}$ and $\mathbb{C}^\mathbb{T}$ respectively. Note that both sets $\mathbb{R}$ and $\mathbb{T}$ are uncountable, thus an interpretation of the transform as a basis expansion is not possible. None the less, multiplication operators provide the necessary generalization of diagonal matrices.

[4] In this case the function space is $\mathbb{C}^\mathbb{N}$, the space of semi-infinite sequences. We need to add a Hilbert space structure in order to make sense of converges of infinite sums, and the choice $\ell^2(\mathbb{N})$ provides additional nice properties such as Parseval's identity. We do not discuss these here.

6.2. Simultaneous Block-Diagonalization and Group Representations. The simplest example of a non-commutative group of transformations is the so-called *symmetric group* $\mathbb{S}_3$ of all permutations of ordered 3-tuples. Consider the ordered 3-tuple $(0,1,2)$ and the following "circular shift" and "swap" operations on it

$$\begin{array}{rclcrcl} I\,(0,1,2) & := & (0,1,2)\,, & & s_{01}\,(0,1,2) & := & (1,0,2)\,, \\ c\,(0,1,2) & := & (2,0,1)\,, & & s_{12}\,(0,1,2) & := & (0,2,1)\,, \\ c^2\,(0,1,2) & := & (1,2,0)\,, & & s_{20}\,(0,1,2) & := & (2,1,0)\,, \end{array} \tag{6.3}$$

where I is the identity (no permutation) operation, and s_{ij} is the operation of swapping the i and j elements. The group operation is the composition of permutations. Note that the first three permutations $\left\{I, c, c^2\right\}$ are isomorphic to $\mathbb{Z}_2$ as a group and thus mutually commute. The swap and shift operations in general do not mutually commute. A little investigation shows that the six elements

$$\left\{I,\ c,\ c^2,\ s_{01},\ s_{12},\ s_{20}\right\} \ = \ \mathbb{S}_3$$

do indeed form the group of all permutations of a 3-tuple.

A *representation* of a group is an isomorphism between the group and a set of matrices (or linear operators) with the composition operation between them being standard matrix multiplication. With a slight abuse of notation (where we use the same symbol for the group element and its representer) we have the following representation of $\mathbb{S}_3$ as linear operators on $\mathbb{R}^3$ (i.e. as matrices on 3-vectors)

$$I = \left[\begin{smallmatrix} 1&0&0\\0&1&0\\0&0&1 \end{smallmatrix}\right], c = \left[\begin{smallmatrix} 0&1&0\\0&0&1\\1&0&0 \end{smallmatrix}\right], c^2 = \left[\begin{smallmatrix} 0&0&1\\1&0&0\\0&1&0 \end{smallmatrix}\right], s_{01} = \left[\begin{smallmatrix} 0&1&0\\1&0&0\\0&0&1 \end{smallmatrix}\right], s_{12} = \left[\begin{smallmatrix} 1&0&0\\0&0&1\\0&1&0 \end{smallmatrix}\right], s_{20} = \left[\begin{smallmatrix} 0&0&1\\0&1&0\\1&0&0 \end{smallmatrix}\right]. \tag{6.4}$$

Those matrices acting on a vector (x_0, x_1, x_2) will permute the elements of that vector according to (6.3).

Now we ask the question: *can the set of matrices in* $\mathbb{S}_3$ *(identified as* (6.4)*) be simultaneously diagonalized?* Recall that commutativity is a necessary condition for simultaneous diagonalizability, and since this set is not commutative, the answer is no. The failure of commutativity can be seen from the following easily established relation between shifts and swaps

$$s_{kl} c \ = \ c\ s_{k+1,l+1}, \qquad k,l \in \mathbb{Z}_2$$

(i.e. the arithmetic for $k+1$ and $l+1$ should be done in $\mathbb{Z}_2$).

The lack of commutativity precludes simultaneous diagonalizability. However, it is possible to simultaneously block-diagonalize all elements of $\mathbb{S}_3$ so they all have the following block-diagonal form

$$\left[\begin{array}{c:cc} * & 0 & 0 \\ \hdashline 0 & * & * \\ 0 & * & * \end{array}\right]. \tag{6.5}$$

In some sense, this the simplest form one can hope for when analyzing all members of $\mathbb{S}_3$ (and the algebra generated by it). This block diagonalization does indeed reduce the complexity of analyzing a set of 3×3 matrices to analyzing sets of at most 2×2 matrices. While this might not seem significant at first, it can be immensely useful in certain cases. Imagine for example doing symbolic calculations with 3×3 matrices. This typically yields unwieldy formulas. A reduction to symbolic calculations for 2×2

matrices can give significant simplifications. Another case is when infinite-dimensional operators can be block diagonalized with finite-dimensional blocks. This is the case when one uses Spherical Harmonics to represent rotationally invariant differential operators. In that case the representation has finite-dimensional blocks, though with increasing size.

Block Diagonalization and Invariant Subspaces. Let's first examine how block diagonalization can be interpreted geometrically. Given an operator $A : \mathcal{V} \longrightarrow \mathcal{V}$ on a vector space, we say that a subspace $\mathcal{V}_o \subseteq \mathcal{V}$ is *A-invariant* if $A\mathcal{V}_o \subseteq \mathcal{V}_o$ (i.e. for any $v \in \mathcal{V}_o$, $Av \in \mathcal{V}_o$). Note that the span of any eigenvector of A (the so-called eigenspace) is an invariant subspace of dimension 1. Finding invariant subspaces is equivalent to *block triangularization*. Let $\mathcal{V}_1$ be some complement of $\mathcal{V}_o$ (i.e. $\mathcal{V} = \mathcal{V}_o \oplus \mathcal{V}_1$), then with respect to that decomposition, A has the form

$$A = \left[\begin{array}{c|c} A_{11} & A_{12} \\ \hline 0 & A_{22} \end{array}\right]. \tag{6.6}$$

Note that in general, the complement subspace will not be A-invariant. If it were, then $A_{12} = 0$ above, and that form of A would be block diagonal. Thus *block diagonalization amounts to finding an A-invariant subspace $\mathcal{V}_o$, as well as a complement $\mathcal{V}_1$ of it such that $\mathcal{V}_1$ is also A-invariant.*

Now observe the following facts which are immediately obvious (at least for matrices) from the the form (6.6). If A is invertible, then $\mathcal{V}_o$ is also A^{-1}-invariant since the inverse of an upper-block-triangular matrix is also upper-block-triangular. If we choose $\mathcal{V}_1 = \mathcal{V}_o^\perp$, the orthogonal complement of $\mathcal{V}_o$, then $\mathcal{V}_o$ is A-invariant iff $\mathcal{V}_o^\perp$ is A^*-invariant (this can be seen from (6.6) by observing that A^* is block-lower-triangular). Finally, in the special case that A is unitary (i.e. $AA^* = A^*A = I$, and therefore $A^{-1} = A^*$), it follows from the previous two observations that for a unitary A, any A-invariant subspace $\mathcal{V}_o$ is such that its orthogonal complement $\mathcal{V}_o^\perp$ is automatically A-invariant. Therefore, for unitary matrices, block triangularization is equivalent to block diagonalization, which can be done by finding invariant subspaces and their orthogonal complements.

Block Diagonalization of $\mathbb{S}_3$. Now we return to the matrices of $\mathbb{S}_3$ (6.5) and show how they can be simultaneously block diagonalized. Note that all the matrices are unitary, and therefore once all the common invariant subspaces are found, they are guaranteed to be mutually orthogonal. The easiest one to find is the vector $(1, 1, 1)$. Note that it is an eigenvector of all members of $\mathbb{S}_3$ with eigenvalue 1 (since obviously any permutation of the elements of this vector produce the same vector again). This is an eigenspace of dimension 1. There is not another shared eigenspace of dimension 1 since then we would have simultaneous diagonalizability, and we know that is precluded by the lack of commutativity. We thus have to simply find the 2-dimensional orthogonal complement of the span of $(1, 1, 1)$. There are several choices for its basis. One of them is as follows

$$v_1 := \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \quad v_2 := \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}, \quad v_3 := \begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix}.$$

Notice that the vectors $\{v_i\}$ are mutually orthogonal, which simplifies calculations that finally give the elements of $\mathbb{S}_3$ in this new basis as

$$c = \left[\begin{array}{c|cc} 1 & 0 & 0 \\ \hline 0 & -1 & -1 \\ 0 & 1 & 0 \end{array}\right], c^2 = \left[\begin{array}{c|cc} 1 & 0 & 0 \\ \hline 0 & 0 & -1 \\ 0 & -1 & -1 \end{array}\right], s_{01} = \left[\begin{array}{c|cc} 1 & 0 & 0 \\ \hline 0 & 0 & 1 \\ 0 & 1 & 0 \end{array}\right], s_{12} = \left[\begin{array}{c|cc} 1 & 0 & 0 \\ \hline 0 & 1 & -1 \\ 0 & 1 & -1 \end{array}\right], s_{20} = \left[\begin{array}{c|cc} 1 & 0 & 0 \\ \hline 0 & -1 & -1 \\ 0 & 0 & 1 \end{array}\right], \tag{6.7}$$

which are all indeed of the form (6.5).

It is more common in the literature to perform the above analysis in the language of group representations, specifically as decomposing a given representation into its component *irreducible representations*. Block diagonalization is then an observation about the matrix form that the representation takes after that decomposition. For the student proficient in linear algebra, but perhaps not as familiar with group theory, a more natural motivation is to start as done above from the block-diagonalization problem as the goal, and then use group representations as a tool to arrive at that goal.

What has been done above can be restated using group representations as follows. A representation of a group $\mathbb{G}$ is a group homomorphism $\rho : \mathbb{G} \longrightarrow \mathrm{GL}(V)$ into the group $\mathrm{GL}(V)$ of invertible linear transformations of a vector space V. Assume for simplicity that $\mathbb{G}$ is finite, ρ is injective, V is finite dimensional, and that all transformations $\rho(\mathbb{G})$ are unitary. The matrices (6.4) of $\mathbb{S}_3$ are in fact the images of an injective, unitary homomorphism $\rho : \mathbb{S}_3 \longrightarrow \mathrm{GL}(3)$ into the group of all non-singular transformations of $\mathbb{R}^3$.

A representation is said to be *irreducible* if there are no non-trivial invariant subspaces common to all transformations $\rho(\mathbb{G})$. In other words, all elements of $\rho(\mathbb{G})$ cannot be simultaneously block diagonalized. As we demonstrated, (6.4) is indeed reducible. More formally, let $\rho_i : \mathbb{G} \longrightarrow \mathrm{GL}(V_i)$, $i = 1, 2$ be two given representations. Their *direct sum* $\rho_1 \oplus \rho_2 : \mathbb{G} \longrightarrow \mathrm{GL}(V_1 \oplus V_2)$ is the representation formed by the "block diagonal" operator

$$\begin{bmatrix} \rho_1(g) & 0 \\ 0 & \rho_2(g) \end{bmatrix}, \tag{6.8}$$

with the obvious generalization to more than two representations. If a representation is reducible, then the existence of a common invariant subspace means that it can be written as the direct sum of so-called "subrepresentations" as in (6.8). Thus *simultaneous block-diagonalization into the smallest dimension blocks* is equivalent to the *decomposition of a given representation into the direct sum of irreducible representations*. This is what we have done for the representation (6.4) of $\mathbb{S}_3$ by finding the two common invariant subspaces (which contain no proper further subspaces that are invariant) and thus brining all of them into the block diagonal form (6.7). In general, it is a fact that any representation of a finite group (more generally, of a compact group) can be decomposed as the direct sum of irreducible representations [5, 6].

REFERENCES

[1] A. V. Oppenheim, A. S. Willsky, and S. H. Nawab, "Signals and systems, vol. 2," *Prentice-Hall Englewood Cliffs, NJ*, vol. 6, no. 7, p. 10, 1983.

[2] R. Plymen, "Noncommutative fourier analysis," 2010.

[3] M. E. Taylor and J. Carmona, *Noncommutative harmonic analysis*. American Mathematical Soc., 1986, no. 22.

[4] W. Rudin, *Fourier analysis on groups*. Courier Dover Publications, 2017.

[5] P. Diaconis, "Group representations in probability and statistics," *Lecture Notes-Monograph Series*, vol. 11, pp. i–192, 1988.

[6] J.-P. Serre, *Linear representations of finite groups*. Springer Science & Business Media, 2012, vol. 42.

Appendix A. Exercises.

A.1. Circulant Structure. Show that any matrix M that commutes with the shift operator S must be a circulant matrix, i.e. must have the structure shown

in (1.1), or equivalently (3.1).

Answer: Starting from the relation $SM = MS$, and using the definition $(S)_{ij} = \delta_{i-j-1}$ compute

$$(SM)_{ij} = \sum_l S_{il}\,(M)_{lj} = \sum_l \delta_{i-l-1}\;\;(M)_{lj} = (M)_{i-1,j}\,,$$
$$(MS)_{ij} = \sum_l (M)_{il}\,S_{lj} = \sum_l (M)_{il}\;\;\delta_{l-j-1} = (M)_{i,j+1}\,.$$

Note that since the indices $i-j-1$ of the Kroenecker delta are to be interpreted using modular arithmetic, then the indices $i-1$ and $j+1$ of M above should also be interpreted with modular arithmetic. The statements

$$(M)_{i-1,j} = (M)_{i,j+1} \quad\Leftrightarrow\quad (M)_{i-1,j-1} = (M)_{i,j} \quad\Leftrightarrow\quad (M)_{i,j} = (M)_{i+1,j+1}$$

then mean that the i'th column is obtained from the previous $i-1$ column by circular right shift of it.

Alternatively, the last statement above implies that for any k, $(M)_{i,j} = (M)_{i+k,j+k}$, i.e. that entries of M are constant along "diagonals". Now take the first column of M as $m_i := (M)_{i,0}$, then

$$(M)_{ij} \;=\; (M)_{i-j,j-j} \;=\; (M)_{i-j,0} \;=\; m_{i-j}.$$

Thus all entries of M are obtained from the first column by circular shifts as in (3.1).

A.2. co-Prime Powers of the Shift. Show that an $n\times n$ matrix M is circulant iff it commutes with S^p where (p,n) are coprime.

Answer: If M is circulant, then it commutes with S and also commutes with any of its powers S^p. The other direction is more interesting.

The basic underlying fact for this conclusion has to do with modular arithmetic in Z_n. If (p,n) are coprime, then there are integers a, b that satisfy the Bezout identity

$$ap + bn \;=\; 1,$$

which also implies that ap is equivalent to 1 mod n since $ap = 1 - bn$, i.e. it is equal to a multiple of n plus 1. Therefore, there exists a power of S^p, namely S^{ap} such that

$$S^{ap} \;=\; S. \tag{A.1}$$

Thus if M commutes with S^p, then it commutes with all of its powers, and namely with $S^{ap} = S$, i.e. it commutes with S, which is the condition for M being circulant.

Equation (A.1) has a nice geometric interpretation. S^p is a rotation of the circle in Figure 3.3 by p steps. If p and n were not coprime, then regardless of how many times the rotation S^p is repeated, there will be some elements of the discrete circle that are not reachable from the 0 element by these rotations (examine also Figure 4.1 for an illustration of this). The condition p and n coprime insures that there is some repetition of the rotation S^p, namely $(S^p)^a$ which gives the basic rotation S. Repetitions of S then of course generate all possible rotations on the discrete circle. In other words, p and n coprime insures that by repeating the rotation S^p, all elements of the discrete circle are eventually reachable from 0.

A.3. Commutativity and Associativity. Show that circular convolution (3.6) is commutative and associative.

Answer: *Commutativity:* Follows from

$$(a \star b)_k \;=\; \sum_l a_l b_{k-l} \;=\; \sum_l a_{k-j} b_j \;=\; (b \star a)_k$$

where we used the substitution $j = k - l$ (and consequently $l = k - j$) .
Associativity: First note that $(b \star c)_i = \sum_j b_j c_{i-j}$, and compare

$$(a \star (b \star c))_k = \sum_l a_l \, (b \star c)_{k-l} \;=\; \sum_l a_l \left(\sum_j b_j c_{k-l-j} \right) \;=\; \sum_{l,j} a_l b_j c_{k-l-j}$$

$$((a \star b) \star c)_k = \sum_j (a \star b)_j \, c_{k-j} \;=\; \sum_j \left(\sum_l a_l b_{j-l} \right) c_{k-j} \;=\; \sum_{l,j} a_l b_{j-l} c_{k-j}.$$

Relabeling $j - l =: i$ (and therefore $j = l + i$) in the second sum makes it

$$\sum_{l,i} a_j b_i c_{k-(l+i)} \;=\; \sum_{l,i} a_j b_i c_{k-l-i},$$

Which is exactly the first sum, but with a different labeling of the indices.