PATTERNED MATRICES

BY

PERSI DIACONIS

TECHNICAL REPORT NO. 320
APRIL 1989

PREPARED UNDER THE AUSPICES
OF
NATIONAL SCIENCE FOUNDATION GRANT DMS96-00235

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
PATTERNED MATRICES

Persi Diaconis
Dept. of Mathematics
Harvard University

Abstract

If the pattern in a matrix can be captured as invariance under a group, the spectral properties of the matrix can be written down using Fourier analysis. Such "generalized circulants" are closed under sum, product, transpose, conjugation. They all commute and are simultaneously diagonalizable. Applications to Markov Chains, Radon transforms, and coding theory are presented. The ideas extend to give natural classes of matrices whose powers are circulants.
1. Introduction.

This is a study of patterned matrices with spectral properties deducible from their group of symmetries. The easiest examples are circulants. These are matrices of form

\[
\begin{array}{cccc}
  a & b & c & d \\
  d & a & b & c \\
  c & d & a & b \\
  b & c & d & a \\
\end{array}
\]

where each row is derived from the row above by shifting right cyclically.

Section 2 lists the well known properties of circulants. In brief, the class of circulants is closed under sum, product, inverse and transpose. All circulants commute. They are simultaneously diagonalized by the Fourier matrix and thus have explicitly given eigenvalues and eigenvectors. If you run into a circulant in the course of a problem you are happy to make its acquaintance.

Circulants are invariant under the cyclic group in the sense that

\[
M_{ij} = M_{\pi(i)\pi(j)}
\]

for \(\pi(i) = i + 1 \pmod{n}\). It is natural to define the symmetry group of a matrix as the set of all permutations \(\pi\) leaving it invariant. One can also begin with a group \(G\) of permutations and consider all invariant matrices as \(G\)-circulants. The well known properties of circulants all have analogs which now involve the representation theory of the group. These topics are developed in section 3.

Section 3 then specializes to matrices of form

\[
M_{st} = f(ts^{-1})
\]

for \(s, t \in G\). These convolution matrices are invariant if and only if the function \(f\) is constant on conjugacy classes: \(f(sts^{-1}) = f(t)\). An example from \(S_3\) is

\[
\begin{array}{cccccc}
  a & b & b & c & c & b \\
  b & a & c & b & b & c \\
  b & c & a & b & b & c \\
  c & b & b & a & c & b \\
  c & b & b & c & a & b \\
  b & c & c & b & b & a \\
\end{array}
\]  

(1-1)
There is a strong pattern in this matrix. The second row can be derived from the first by switching \((1, 2)\) \((3, 4)\) \((5, 6)\). The same operations give row 4 from 3 and 6 from 5. All such matrices commute. Their eigenstructure can be explicitly given using the character theory of the symmetric group. Similarly, every finite group gives rise to a simple collection of matrices by \(M_{st} = f(ts^{-1})\) where \(f\) is constant on conjugacy classes.

I encountered these generalized circulants in joint work with Mehrdad Shahshahani analyzing repeated shuffling of a deck of cards. They proved useful in the analysis of Radon transform in joint work with Ron Graham. They had earlier been applied in work of Rothaus and Thompson (1966) to show the non-existence of perfect codes. More recently Chilling (1987b) and others have applied the spectral decomposition to finite group theory. Andersson and Pearlman (1988) use these ideas to get generalizations of Hadamard’s inequality. These applications are discussed in sections 3C and 4E.

Section 4 outlines the most broadly useful generalization: If the matrices are indexed by a finite set \(X\) on which \(G\) acts transitively then matrices of form \(M_{xy} = f(x^{-1}y)\) can be defined. For these to be well defined, \(f\) must satisfy \(f(k_1xk_2) = f(x)\), for all \(k\) in \(K \subset G\) the isotropy subgroup. All such patterned matrices commute if and only if \((G, X)\) form what is called a Gelfand pair. The eigenvalues are given as explicit spherical transforms of the matrix entries. Everything can be computed. This includes the constant on conjugacy set up but there are hundreds of other examples that occur naturally, including matrices indexed by the vertices of distance transitive graphs. An example, analyzed further in section 4 is

\[
\begin{array}{cccccc}
  a & b & b & b & b & c \\
  b & a & b & c & b & b \\
  b & b & a & b & c & b \\
  b & c & b & a & b & b \\
  b & b & c & b & a & b \\
  c & b & b & b & b & a \\
\end{array}
\]

Section 5 presents examples such as the following

\[
\begin{array}{cccccc}
  a & b & c & d & e \\
  c & d & e & a & b \\
  e & a & b & c & d. \\
  b & c & d & e & a \\
  d & e & a & b & c \\
\end{array}
\]
Here each row is shifted by 2. These are matrices whose powers are circulants. They arise naturally in applications and also have a nice theory.

The final section lists some open problems and directions for further research.
2. Circulants.

**A. Basic Properties.** An $n \times n$ matrix with complex entries of form

\[
\begin{array}{cccc}
  a_0 & a_1 & \cdots & a_{n-1} \\
  a_{n-1} & a_0 & \cdots & a_{n-2} \\
  a_{n-2} & a_{n-1} & \cdots & a_{n-3} \\
  \vdots & \vdots & \ddots & \vdots \\
  a_1 & a_2 & \cdots & a_0 \\
\end{array}
\]

is called a circulant. These matrices have a long history carefully surveyed by Muir (1890-1923). They are still a topic of research interest as the reader can discover by browsing through recent years of the journal *Linear Algebra and its Applications*.

A lovely introduction to circulants is given by Davis (1979). This book discusses numerous applications. It seems like a worthwhile project to see in what way the properties described by Davis extend to the generalizations suggested here.

One way to organize and understand properties of circulants recognizes them as the matrix form of convolution: Treat the index set $0, 1, 2, \cdots n - 1$ as the integers mod $n$. Then a circulant has form

\[
M_{ij} = f(j - i) \quad \text{with} \quad f(k) = a_k, \quad \text{subscripts mod } n.
\]

Multiplying $M$ on the left by the row vector $(g(0), g(1) \cdots g(n - 1))$ yields

\[
(2-1) \quad \sum_{i=0}^{n-1} g(i)f(j - i) \quad \text{as } j^{th} \text{ entry.}
\]

It is natural to interpret (2-1) in the language of probability. Suppose $g(0), g(1) \cdots g(n - 1)$ are the respective chances of a random walk taking initial step $0, 1, \cdots, n - 1$, and $f(0), \cdots, f(n - 1)$ the chance of the second step in a walk. Then, the chance that after two steps the walk winds up at $j$ is given by (2-1) – the convolution of $g$ and $f$. This will be denoted $f * g(j)$ in what follows.

The Fourier transform of $f$ at frequency $j$ is

\[
(2-2) \quad \hat{f}(j) = \sum_{k=0}^{n-1} e^{2\pi ijk/n} f(k).
\]
As usual, \( \widehat{f} \ast g(j) = \widehat{f}(j) \cdot \widehat{g}(j) \) and \( f \) can be recovered from \( \widehat{f} \) by

\[
(2-3) \quad f(k) = \frac{1}{n} \sum_{j=0}^{n-1} e^{-2\pi ijk/n} \widehat{f}(j).
\]

Fix a function \( f \) and define a circulant by \( M_{ij} = f(j-i) \). Take the function \( g_k(j) = e^{2\pi ijk/n} \). Then

\[
(g_kM)_j = \sum_i g_k(i)f(j-i) = \sum_i g_k(j-i)f(i) = g_k(j)\sum_i g_k(-i)f(i).
\]

Thus the complex exponentials \( g_k(\cdot) \) are left eigenvectors of \( M \) with eigenvalue \( \widehat{f}(-k) \). The usual summing of the complex exponentials to 0 implies that the Fourier matrix

\[
(2-4) \quad F_{jk} = e^{2\pi ijk/n} \sqrt{n}
\]

is unitary and the spectral decomposition becomes the following: \( M \) is a circulant with first row \( f(0), f(1) \cdots f(n-1) \) if and only if

\[
(2-5) \quad M = F^* \text{Diag}(\widehat{f}(0), \widehat{f}(-1) \cdots \widehat{f}(-(n-1)))F.
\]

Here \( \widehat{f} \) is defined at (2-2) and \( F \) is the Fourier matrix defined at (2-4).

All properties of circulants follow from this spectral representation:

(2-6) The class of circulants is closed under sum, product, transpose, conjugation and inverse. Davis (1979) discusses generalized inverses of circulants.

(2-7) All circulant commute. They are simultaneously diagonalized by the Fourier matrix \( F \) of (2-4).

B. Some Applications. Circulants appear in many mathematical problems. Davis (1979) gives several examples. The present section lists some applied problems where circulants make a natural appearance.

Example 2 – 1. Random Walk. Consider a particle constrained to hop about on \( n \) points arranged in a circle. At each time the particle hops left or right with probability \( \frac{1}{2} \).
This is a cyclic version of the classical drunkard's walk. Index the points as 0, 1, 2, \cdots, n-1. The chance of moving from i to j is thus

\[ M_{ij} = \begin{cases} 
\frac{1}{2} & \text{if } |i - j| = 1 \\
0 & \text{otherwise.}
\end{cases} \]

The matrix \( M \) is a circulant with first row \((0, \frac{1}{2}, 0, \cdots, 0, \frac{1}{2})\). Using the explicit diagonalization one can answer a question such as "What happens after many steps?" Answer: the walk becomes uniformly distributed. More specifically, the chance that a walk starting at i winds up at j after \( k \) steps is \( \frac{1}{n} \) plus an error term that becomes exponentially small provided \( k \gg n^2 \). This assumes \( n \) is odd (otherwise the walk is at an even position after an even number of steps).

One may ask other questions such as "How long on average before the walk returns to where it started?" Answer: about \( n^2 \). "How long before the walk hits every point?" Answer: about \( n^2 \). These and other classical problems about random walk are discussed in Feller (1968). Diaconis (1988) gives details using the tools developed here.

If \( X_k \) is the position of the walk at time \( k \), the walk can be described as

\[ X_k = X_{k-1} + \epsilon_k \text{ (mod } n), \]

where \( \epsilon_k \) are random variables taking value \( \pm 1 \) with probability \( \frac{1}{2} \). Using formula (2-8) the disturbance terms can be generalized to have an arbitrary distribution \( P\{\epsilon_k = j\} = f(j) \), say. The associated Markov chain is still a circulant with first row \((f(0), \cdots, f(n-1))\).

**Example 2.2 Smoothing Data.** Consider \( x_0, x_1, x_2, \cdots, x_{n-1} \) a sequence of observations collected in time. A plot of \( i \) versus \( x_i \) may be smoothed replacing \( x_i \) by e.g., \( (x_{i-1} + x_i + x_{i+1})/3 \) or \( (x_{i-1} + 2x_i + x_{i+1})/4 \). Such linear smoothers can be represented as linear maps on sequences of length \( n \). The matrices of linear smoothers are circulants. In this context, the eigenvalues of the associated circulants are important in studies of stability and invertibility – a "smooth" is a 1-1 encoding of data if and only if no eigenvalue vanishes.

**Example 2.3 Covariance Matrices.** If \( X_0, X_1, \cdots, X_{n-1} \) is a series of random variables, the covariance between \( X_i \) and \( X_j \) is defined by

\[ R_{ij} = E(X_i - \mu_i)(X_j - \mu_j) \text{ with } \mu_i = E(X_i). \]
The collection of covariances are often collected in a covariance motion $\Sigma$. This has $(i,j)$ entry $R_{ij}$. In applied work $\Sigma$ is to be estimated from data. This is a hard job and a variety of simplifying assumptions are commonly employed. Stationarity implies that $R_{ij}$ is a function of $i - j$ only. If this is interpreted cyclically, $\Sigma$ is a circulant. The cyclic assumption is sometimes valid from physical considerations – for example, $X_i$ could be readings of seismic recorders placed in a circle around Mt. St. Helens.

Andersson (1986) is a classical reference to circularly symmetric covariance matrices. Perlman (1987) gives a modern discussion of this and its generalizations.

**Example 2.4  Cyclic Codes.** A binary code is a collection of binary vectors of length $n$. It is called *cyclic* if it is closed under cyclic rearrangements. For example, $\{000, 001, 010, 100\}$ is a cyclic code. A code is *quasi-cyclic* if there is some integer $s$ such that every cyclic shift of a code word by $s$ is again a code word. Such codes have an elegant theory. They include as important the important examples of the $BCH$ codes.

The assignment of a given input message to a code word is often a linear operation. The matrix of this linear map is called the generator matrix of the code. By convention the generator matrix is of the form $[I|A]$, where $I$ is an identity matrix and $A$ is a matrix. There is a rich class of naturally occurring codes whose generator matrix has $A$ a circulant. Chapter (16, §§6,7) of Macwilliams and Sloane (1977) discusses these results. In particular, they show that such "double circulant" codes are automatically quasi-cyclic with $s = 2$.

**Further examples.** Ablow and Brenner (1963) show how circulants arise naturally in a class of mechanical problems involving a system of springs. I believe there are dozens of other examples and would be pleased to hear from others on this matter.
3. Circulants indexed by groups.

One natural generalization of circulants replaces the integers mod $n$ by a finite group $G$. Let $f$ be a real or complex function defined on $G$. Define a $|G| \times |G|$ matrix $M$ via

$$M_{st} = f(ts^{-1}) \quad \text{for } s, t \in G.$$  

An elegant development results from taking $f$ to be constant on conjugacy classes:

$$(3-2) \quad f(sts^{-1}) = f(t) \quad \text{for all } s, t \in G.$$  

Such functions are called class functions. The associated matrices will be called class circulants.

**Remarks.** 1) For the integers mod $n$ (or any Abelian group) $sts^{-1} = t$ for every $s, t \in G$. Thus all functions are class functions.

2) For non-Abelian groups, sizable numbers of elements can be lumped together in a conjugacy class. For example, $S_3$ – the 6-element symmetric group – has 3 conjugacy classes $\{id\}, \{(1,2),(1,3),(2,3)\}, \{(1,2,3),(1,3,2)\}$. Here, permutations are being written in cycle notation, so $(1,2,3)$ takes 1 to 2 and 2 to 3 and 3 to 1. In $S_3$ a class function only takes on 3 values. Call them $a, b, c$. If the group is written in the order above, the $6 \times 6$ matrix is shown in (1-1) above.

The main result of this section shows that the class circulants are closed under sum, product, transpose, conjugation and inverse. All class circulants commute and are explicitly simultaneously diagonalizable. These results are proved in section B below which gives a crash course in elementary group representation theory. Section C gives examples and applications. Section D relates the definition above to work of Chillag (1987a,b, 1988) who gives applications to group theory.

**B. A crash course in group representations.**

Let $G$ be a finite group. A representation assigns matrices to group elements in such a way that the product of group elements is assigned to the product of their matrices. Formally, a representation is a homomorphism $\rho : G \to GL(V)$. Thus $\rho(st) = \rho(s)\rho(t)$. Here $V$ is a vector space over the real or complex numbers of dimension $d_\rho < \infty$. 

9
For \( f : G \to \mathbb{C} \) a function define the Fourier transform at \( \rho \) by

\[
\hat{f}(\rho) = \sum_{s \in G} f(s)\rho(s).
\]

If \( g \) is another function define convolution by

\[
f * g(s) = \sum_{t \in G} f(st^{-1})g(t).
\]

As usual, Fourier transform turns convolution into multiplication

\[
\hat{f} \ast \hat{g}(\rho) = \hat{f}(\rho)\hat{g}(\rho).
\]

This is easily checked by multiplying the matrices on the right hand side.

A representation is irreducible if 0 and \( V \) are the only invariant subspaces of \( V \). Thus if \( W \) is a subspace of \( V \) such that \( \rho(s)W \subset W \) for each \( s \in G \) then \( W = 0 \) or \( V \).

Example. Take \( G = S_n \) the symmetric group. Every mathematician is familiar with 3 representations:

(i) The trivial representation: \( V \) is 1-dimensional, \( \rho(s)V = V \).

(ii) The alternating representation: \( V \) is 1-dimensional, \( \rho(s)V = \text{sgn}(s)V \).

(iii) The \( n \)-dimensional representation: \( V \) is \( n \)-dimensional, generated by the standard basis vectors \( e_1, e_2, \ldots, e_n \). If \( s \) is a permutation, define \( \rho(s)e_i = e_{s(i)} \) and extend by linearity. The matrix of the linear map \( \rho(\cdot) \) is the usual permutation matrix in this basis.

The \( n \)-dimensional representation is not irreducible: If \( W \) is all multiples of \( e_1 + \cdots + e_n \), then \( W \) is invariant under permuting coordinates. Similarly \( W^\perp = \{x \in \mathbb{C}^n : x_1 + \cdots + x_n = 0\} \) is invariant and can be shown to be irreducible. There are other irreducible representations of \( S_n \).

The Fourier inversion theorem gives a recipe for reconstructing \( f \) from \( \hat{f}(\rho) \) at all irreducible representations of \( G \):

\[
f(s) = \frac{1}{|G|} \sum_{\rho} d_\rho \text{tr}(\rho(s^{-1})\hat{f}(\rho)).
\]
Example. Take \( G = \mathbb{Z}(n) \) the integers mod \((n)\). There are \( n \) irreducible representations; all are 1-dimensional. A 1-dimensional linear map is given by multiplying by a complex number. Here the \( j^{th} \) representation

\[
\rho_j(k) = e^{2\pi i j k / n}.
\]

The Fourier transform at the \( j^{th} \) representation becomes the usual discrete Fourier transform:

\[
\hat{f}(\rho_j) = \sum_{k=0}^{n-1} e^{2\pi i j k / n} f(k).
\]

The Fourier inversion theorem becomes

\[
f(k) = \frac{1}{n} \sum_{j=0}^{n-1} e^{-2\pi i j k / n} \hat{f}(\rho_j).
\]

The kind of elementary representation theory needed to follow this paper is developed in the first 30 pages of Serre (1977). Other fine references are Lederman (1988), Naimark-Stern (1982), or the encyclopedic Curtis-Reiner (1962).

To describe the diagonalization of convolution matrices enumerate the group as \( G = \{s_1, s_2, \cdots, s_N\} \). Let \( F_{ij} = f(s_j s_i^{-1}) \) with \( f \) an arbitrary function. Let \( \rho_1, \rho_2, \cdots, \rho_k \) be the irreducible representations. These may be taken as unitary, so \( \rho(s) \) is a unitary matrix for every \( s \in G \).

The diagonalization is built up in stages. For each representation \( \rho_j \), define a \( d_j^2 \times d_j^2 \) block diagonal matrix:

\[
M_j = \begin{pmatrix}
\hat{f}(\rho_j) & 0 \\
0 & \ddots & \hat{f}(\rho_j)
\end{pmatrix}.
\]

Let

\[
(3-3) \quad M = \begin{pmatrix}
M_1 & 0 \\
\ddots & \ddots \\
0 & M_k
\end{pmatrix}.
\]

This \( M \) is \( N \times N \) because \( d_1^2 + d_2^2 + \cdots + d_k^2 = N \) (Serre (1978, page 18).
Let
\[
\psi_j(s) = \sqrt{\frac{d_j}{N}} (\rho_j(s)_{11}, \ldots, \rho_j(s)_{d_j,d_j})
\]
\[
\varphi(s) = (\psi_1(s), \psi_2(s), \ldots, \psi_k(s))^t.
\]
This \( \varphi \) is an \( N \times 1 \) column. Finally let \( \Phi \) be the \( N \times N \) matrix with columns \( \varphi(s_i) \)

\[
(3-4) \quad \Phi = (\varphi(s_1), \varphi(s_2), \ldots, \varphi(s_N)).
\]

**Theorem 1.** Let \( f \) be a function from a finite group \( G \) into \( \mathbb{C} \). Let \( F_{st} = f(ts^{-1}) \) be an \( N \times N \) matrix with \( N = |G| \). Then

\[
(3-5) \quad F = \Phi^*M\Phi
\]

with \( \Phi^*\Phi = I \), a unitary matrix defined in (3-4) above and \( M \) as in (3-3).

**Proof:** The decomposition follows from the Fourier inversion theorem:

\[
F_{ab} = \frac{1}{N} \sum_{j=1}^{K} d_j \text{tr}[\hat{f}(\rho_j)\rho_j(s_b)\rho_j(s_a^{-1})]
\]

\[
= \frac{1}{N} \sum_{j=1}^{K} d_j \text{tr}[\rho_j(s_a^{-1})\hat{f}(\rho_j)\rho_j(s_b)]
\]

\[
= \sum_{j=1}^{K} \psi_j(s_a)^* M_j \psi_j(s_b).
\]

The last equality follows by expanding the trace. The property \( \Phi^*\Phi = I \) follows from the orthogonality of the matrix entries of the irreducible representations. This is one of the standard corollaries of Schur's lemma (see, e.g., Serre (1978, pg. 14)).

**Remark.** The decomposition of \( M \) is not a diagonalization, indeed the matrices \( \hat{f}(\rho_j) \) that make up \( M_j \) can be arbitrary: just choose them and define \( f \) through the Fourier inversion theorem. In the following corollary, class functions are shown to have diagonal \( \hat{f}(\rho_j) \) and in section 4 Gelfand pairs are shown to have diagonal transforms.

**Corollary 1.** Let \( F_{st} = f(ts^{-1}) \) with \( f \) a class function (i.e., \( f(sts^{-1}) = f(t) \)). Then \( M \) can be unitarily diagonalized as

\[
M = \Phi^*D\Phi
\]

12
with $\Phi$ unitary as defined in (3-4) above, and $D$ a diagonal matrix made of blocks $\Delta_1, \Delta_2, \cdots, \Delta_k$ with

$$\Delta_i = c_i I_{d_i^2 \times d_i^2}$$

and

$$c_i = \frac{1}{d_i} \sum_{t \in G} f(t)\overline{\chi_i(t)}; \quad \chi_i(t) = Tr(\rho_i(t)) = -the\ character\ of\ \rho\ at\ t.$$ 

**PROOF:** Let $\rho$ be a unitary representation and observe that $\hat{f}(\rho)$ commutes with $\rho(s)$:

$$\rho(s)\hat{f}(\rho)\rho(s^{-1}) = \sum_{t \in G} f(t)\rho(sts^{-1}) = \hat{f}(\rho).$$

Schur’s lemma now implies that $\hat{f}(\rho) = cI$. Calculate the constant by taking traces:

$$d\rho c = \sum f(s)\chi_\rho(s).$$

Thus in Theorem 1, $M_i = c_i I_{d_i^2 \times d_i^2}$. ■

**Remarks:** The explicit diagonalization was used in Rothaus and Thompson (1966) for the symmetric group. Diaconis and Shahshahani (1981) use it in general. It was written out in this explicit form in Mathews (1985). It has frequently been rediscovered, see, e.g., Zieschang (1988). I presume it is folklore going back to the beginnings of the subject.

**Corollary 2.** Let $G$ be a finite group with $|G| = N$. Let $f$ be a class function with associated matrix $F_f = f(ts^{-1})$. The class of such matrices is closed under sum, product, transpose, conjugation, and inverse. All such $F_f$ commute and are simultaneously diagonalizable by the matrix $\Phi$ of (3-4).

**C. Applications.**

The explicit decomposition has been applied to problems of random walk, perfect codes, and inversion of Radon transforms. These applications are detailed in Chapter 3 of Diaconis (1988) and will only be treated briefly here.

Consider first the problem of random transpositions studied by Diaconis and Shahshahani (1981). Begin with $n$ cards face down on a table. Card 1 is at the left, card 2 next, and card $n$ at the right of the row. Cards are repeatedly mixed by random transpositions. Thus the left hand chooses a card at random, the right hand chooses a card at random (so
left = right with probability $\frac{1}{n}$) and then the two cards are transposed. This corresponds to the probability

$$f(\pi) = \begin{cases} \frac{1}{n} & \text{if } \pi = id \\ \frac{2}{n^2} & \text{if } \pi \text{ is a transposition} \\ 0 & \text{otherwise.} \end{cases}$$

The identity and the transpositions form conjugacy classes and so $f(\pi)$ is a class function. If this problem is analyzed as a Markov chain, the transition matrix is $M_f$ and the explicit diagonalization allows one to show $\frac{1}{2} n \log n + cn$ transpositions are necessary and suffice to mix up $n$ cards.

Rothaus and Thompson (1966) used exactly the same ideas to study the existence of perfect codes in the symmetric group. Here the group is treated as a metric space using the minimum number of transpositions required to bring $\pi$ to $\sigma$ as a metric. One seeks a code – that is, a set of permutations – such that balls of a fixed radius about each code-word are disjoint and cover $S_n$. Such a code is called perfect. Taking the unit ball gives as indicator function

$$f(\pi) = \begin{cases} 1 & \text{if } \pi = id \text{ or a transportation} \\ 0 & \text{otherwise.} \end{cases}$$

If $g$ is the indicator function of the code, then perfectness translates into $f \ast g(\pi) \equiv 1$. Fourier transforming this relation, gives strong restrictions on $\hat{g}(\rho)$. Rothaus and Thompson used these results to show that perfect codes on $S_n$ do not exist for any $n$ such that a prime $p$ divides $1 + \binom{n}{2}$ with $p > \sqrt{n} + 2$. Laura Chihara (1987) has used these ideas to show that perfect codes do not exist for other classical finite groups.

A third application of the explicit diagonalization appears in work of Diaconis and Graham (1985) on Radon transforms. Very roughly, the Radon transform replaces a function $f$ with an average and then one asks about uniqueness and inversion. If $f$ is defined on a group and

$$Rf(s) = \sum_{t \in B} f(st)$$

where $B$ is a Ball in a metric on $G$, then the Radon transform can be represented as a convolution matrix. If the metric is bi-invariant, $d(a,b) = d(sa, sb) = d(as, as)$, the matrix will be a class circulant. The transform is invertible if and only if the numbers $c_i$ of Corollary 1 are non-zero.
This brief sketch omits many rich mathematical details. Morrison (1986) and Fill (1988) are nice examples of how far the simple ideas outlined above can be pushed.

**D. Chillag's contribution.** In a series of papers, David Chillag (1987a,b, 1988) has developed a parallel theory. With present notation, it is easy to relate his results to the development of section B. Let $A$ be the space of class functions on a finite group $G$. This is a linear space over the complex numbers. It can be considered as a commutative algebra in two ways. The classical way to form an algebra is by convolution since the convolution of two class functions is a class function.

A second natural way to form a commutative algebra uses the pointwise product of two functions. This leads to Chillag's generalized circulants. To explain, take the characters of the irreducible representations as a basis for the algebra $A$ under pointwise product. These are orthonormal under the usual inner product

$$
(f|g) = \frac{1}{|G|} \sum_{s \in G} f(s)g^*(s).
$$

Fix a class function $f$ and treat it as a linear map on $A$ by

$$
L_f(g)(x) = f(x)g(x).
$$

In the basis given by the characters $L_f$ has matrix with $i,j$ entry

$$
M_{ij} = \langle f \chi_i | \chi_j \rangle.
$$

Chillag calls these matrices generalized circulants. They are simultaneously diagonalizable as

$$
M = X^{-1}DX.
$$

With $X$ the character matrix having $i,j$ entry $\chi_i(c_j)$ the value of the $i^{th}$ character at the $j^{th}$ conjugacy class. The matrix $D$ is diagonal with $(i,i)$ element $f(c_i)$. Note that the matrix $X$ is not unitary in non-Abelian examples.

From this representation one sees that all the usual properties of circulants hold. One way to understand these circulants is to look back at Theorem 1. The matrix $M_k$ has a
lot of redundancy, with repeated blocks down the diagonal. Especially for class functions, one can reduce this redundancy by looking only at the class functions.

One can also consider $f$ acting on $A$ by convolution. This gives a linear map which also leads to simultaneously diagonalizable matrices. This is essentially what was done in section B. It is important to note that the two algebra structures on $A$ are "isomorphic" (both are commutative semi-simple algebras of the same dimension). The Fourier transform takes one algebra to the other. Giving the isomorphism explicitly is a classical unsolved problem of identifying characters and conjugacy classes in a natural way. See Mackey (1975).

Having two ways of writing the same thing can be enormously useful. Chillag applies his ideas to group theory in Chillag (1987b). The results are refreshing and thought-provoking.

A. Gelfand pairs. The examples above show that a natural generalization of circulants can be developed using group theory. There are other possible generalizations. In this section a natural general theory is developed.

Let $X$ be a finite set. Let $M_{xy}$ be a matrix with complex entries indexed by $X$. Let $G$ be a group of permutations of $X$ preserving $M$. Thus $s \in G$ implies $M_{sxsy} = M_{xy}$ for all $x, y$. For example, the set of all permutations preserving $M$ always forms a group.

To start a theory, suppose $G$ acts transitively on $X$ so that for each $x, y$ in $X$ there is $s \in G$ such that $sx = y$. Fix $x_0 \in X$. Let $K = \{s : sx_0 = x_0\}$. Identify $X$ with cosets of $K$: choose $id, x_1, x_2, \ldots$ in $G$ such that $G = K \cup x_1K \cup x_2K \cup \cdots$ as a disjoint union. Identify $x_0$ with the identity. There is no essential dependence on the choice of $x_0$. Any $s \in G$ can be uniquely written $s = x_i k$ for some $x_i$.

Define a function $f$ on $G$ by

$$f(xk) = f(x) = M_{x, x_0}.$$ 

By definition $f$ is right invariant under $K$; $f(sk) = f(s)$. The definition also yields that $f$ is $K$ left invariant; $f(ks) = f(s)$.

Going backward; if $f$ is given as a $K$-bi-invariant function on $G$. Let $L(X)$ be the vector space of complex valued functions on $X$. Define a linear mapping $T_f : L(X) \to L(X)$ by

$$T_f g(x) = \sum_{y \in X} g(y)f(y^{-1}x).$$

Observe that $T$ is well defined. If $L(X)$ is given its usual basis of delta functions $\delta_z(x) = 1$ if $x = z$, zero otherwise, the matrix of $T_f$ for this basis is $M_{xy} = f(y^{-1}x)$. If $g$ above is $K$ left invariant so is $T_f g$.

This gives a 1-1 correspondence between matrices on $X = G/K$-invariant under $G$ and $K$-bi-invariant functions on $G$. In this correspondence, matrix product corresponds to convolution. Let the algebra of $G$-invariant matrices be denoted $H(G, K)$. These objects are often called Hecke algebras. Curtis and Reiner (1962) (1986 II.D) contains a clear
elementary treatment. Tackas (1986) contains an extensive bibliography. Hecke algebras are an important part of modern group theory and are still in a very active state of development.

**Example.** (Class circulants). Let \( X \) itself be a group, whose elements are denoted \( x, y, \ldots \). Let \( G \) be the product group \( X \times X \). This acts on \( X \) by \((s,t)x = sx t^{-1}\). Taking \( x_0 = id \). The isotropy subgroup \( K \) is seen to be the diagonal subgroup \( \{(s,s)\} \). A matrix \( M_{xy} \) is \( G \) invariant if and only if \( M_{xy} = f(y^{-1}x) \) with \( f \) constant on conjugacy classes. ■

The group \( G \) operates on \( X \) and so on \( L(X) \). Thus \( L(X) \) is a representation of \( G \). By Mashkes' Theorem (Serre 1977, p. 7) \( L(X) \) decomposes into a direct sum of invariant irreducible subspaces.

\[
(4-1) \quad L(X) = V_0 \oplus V_1 \oplus \cdots \oplus V_j.
\]

Define the pair \((G,K)\) to be a *Gelfand pair* if no \( V_i \) is isomorphic to a \( V_j \) for \( i \neq j \). We say the direct sum decomposition is multiplicity free. Gelfand pairs have an extensive modern development. Recent surveys are given by Conway and Sloane (1987) who give applications to packing and covering. Letac (1981) and Diaconis (1988) are surveys aimed at probabilists and statisticians.

The connection with circulants comes in the following theorem proved, for example, in Diaconis (1988, Chapter 3F).

**Theorem 2.** Let \( G \) be a finite group acting transitively on a finite set \( X \) with isotropy subgroup \( K \). The following are equivalent

a) The algebra of \( G \) invariant matrices is commutative.

b) \( G/K \) is a Gelfand pair.

c) There exists a unitary matrix \( U \) such that for every invariant \( M \)

\[
M = U^* D U
\]

with \( D \) a block diagonal matrix having \( j \) blocks (\( j \) appears in \( 4-1 \) above) the \( i^{th} \) block is itself diagonal with only the \((1,1)\) entry nonzero. This entry
is \( \sum_{x} s_{i}(x) M_{x,x_{0}} \), where the spherical function \( s_{i} \) is the unique \( K \)-bi-invariant function in \( V_{i} \) having \( s_{i}(x_{0}) = 1 \).

**Remarks.**

1. There are literally hundreds of classes of Gelfand pairs known. Diaconis (1988; 3G) gives an annotated bibliography. The class functions form a commutative algebra and so a Gelfand pair. Distance transitive graphs generate an interesting class of examples where the eigenvalues and eigenvectors can be explicitly given. See Biggs (1974) and Bannai-Ito (1984,1986). Stanton (1984) gives an extremely clear discussion of Gelfand pairs arising from Coxeter groups where the spherical functions are classically studied orthogonal polynomials.

2. Invariant matrices can be block diagonalized without the Gelfand pair condition. Start with a matrix \( M_{xy} \). Let \( G \) be the largest group such that \( M_{zz} = M_{xy} \). Take \( K \) as the isotropy subgroup. Suppose

\[
L(X) = V_{0} \oplus V_{1} \cdots \oplus V_{j} = W_{0} \oplus W_{1} \oplus \cdots \oplus W_{t},
\]

where \( W_{i} \) is a direct sum of isomorphic irreducible representations. Suppose that \( W_{i} \) lumps together \( \ell_{i} \) of the \( V \)'s. Then, there is a unitary \( U \) such that every invariant \( M \) can be written

\[
M = U^{*} D U
\]

with \( D \) block diagonal, the \( i^{th} \) block of form \( \begin{pmatrix} A & 0 \\ 0 & 0 \end{pmatrix} \) with \( A \ell_{i} \times \ell_{i} \) arbitrary. There are \( \ell + 1 \) blocks. The \( i^{th} \) block can be given reasonably explicitly in terms of Fourier analysis. See, for example, Dieudonné (1978). An example is presented in the last section of this paper. The result follows from Theorem 1 in section 3 and Frobenius reciprocity.

3. It is an interesting algorithmic question to calculate the symmetry group of a given matrix. If the group is to be transitive, all rows (and columns) must be permutations of the first row. Fix a row (say, the first). The set of permutations arising from the rows is one set of candidates (if all entries in the first row are distinct, these must all work).

**B. An Example.** Fix integers \( k, n \), \( 0 < k \leq n/2 \). Let the set \( X \) consist of the \( k \)-element subsets of \( \{1,2,\cdots,n\} \). Thus \( |X| = \binom{n}{k} \). The symmetric group \( S_{n} \) acts transitively on \( X \). The subgroup fixing \( \{1,2,\cdots,k\} \) is \( S_{k} \times S_{n-k} \). Thus \( X \) can be identified with \( S_{n}/S_{k} \times S_{n-k} \).

\[19\]
When $n = 4, k = 2$, $X$ can be identified with the vertices of an octahedron.

Random walk on this graph gives rise to a matrix of form

$$
\begin{pmatrix}
    a & b & b & b & b & c \\
    b & a & b & c & b & b \\
    b & b & a & b & c & b \\
    b & c & b & a & b & b \\
    b & b & c & b & a & b \\
    c & b & b & b & b & a
\end{pmatrix}.
$$

This is invariant under $S_4$ and can be explicitly diagonalized using Theorem 2.

For general $n, k$, this example arose in joint work with Shahshahani in analyzing a model of diffusion proposed by Bernoulli and Laplace. Consider two urns. The left contains $k$ red balls, the right contains $n - k$ black balls. At each time, a ball is chosen at random from each urn and the two balls are switched. It is intuitively clear that after many switches the urns are all mixed up, about half red and half black. Using the machinery above, Shahshahani and I showed that order $n \log n$ switches are required. The paper gives an explicit diagonalization. Here the spherical functions are "dual Hahn polynomials".

**C. Greenhalgebras.** One way to interpolate between conjugacy invariance and bi-invariance under a subgroup consider a finite group $G$ and subgroups $H \supset K$ with $K$ normal in $H$. Let $H(G; H, K)$ be the set of complex functions on $G$ conjugacy invariant under $H$ (so $f(s) = f(hs h^{-1})$) and bi-invariant under $K$ (so $f(s) = f(k_1sk_2)$). These form an algebra under convolution.

Taking $H = G, K = id$ gives conjugacy invariance. Taking $K = H$ gives bi-invariance. Shahshahani and I discovered that this algebra is commutative for the example $S_n \supset S_k \times S_{n-k} \supset S_k$. See also Travis (1974), Andy Greenhalgh (1989) and Hirschman (1972) give necessary and sufficient conditions on $G, H, K$ to insure commutativity in general: Any
irreducible representation of $H$ that restricts to an isotypic direct sum of $1$-dimensional representations of $K$ must induce up to a multiplicity free representation of $G$.

Charles Dunkl (1972) has pointed out that this can be recast in the language of Gelfand pairs: form the product group $G \times K$ with subgroup $H \times K$. The algebra of bi-invariant functions in this product set up is isomorphic to $H(G; H, K)$.

Greenhalgh carried out the details of the following random walk example: consider $n$ balls labelled 1, 2, \cdots $n$. Place $k$ of them in a rack in order 1, 2, \cdots $k$. Place the remaining $n - k$ in a bag. Each time pick a ball from the bag and one from the rack and switch. All of the spherical functions underlying the diagonalization of this Markov chain are computable from character theory. Greenhalgh gives a sharp form to the following rough summary – $n \log n + cn$ steps are necessary and suffice to mix things up.

**D. Some matrix theory.** Working in a different setting, David Chillag (1988) has pointed out that many of the computations can be usefully regarded as change of basis formulas. Let $A$ be a commutative algebra of dimension $k$ over $\mathbb{C}$ (assumed semisimple) with $\Sigma = \{\sigma_1, \cdots, \sigma_k\}$ and $\Delta = \{\delta_1, \cdots, \delta_k\}$ a pair of bases of $A$. For $a \in A$ define a $k \times k$ matrix $M(a, \Sigma, \Delta)$ by $a\sigma_i = \Sigma m_{ij}(a; \Sigma, \Delta)\delta_j$.

A standard result (e.g., Suprunenko and Tyshkevich (1968)) is that $A$ has a basis $\mathcal{E} = \{\epsilon_1, \cdots, \epsilon_k\}$ such that $a\epsilon_i = a(i)\epsilon_i$ for all $a$ (so $\mathcal{E}$ is a basis of eigenvectors of all $a$ simultaneously), $\epsilon_i\epsilon_j = 0$ for $i \neq j$, and $\epsilon_i^2 = \epsilon_i$. It is convenient to take one of the bases above as this $\mathcal{E}$. If a $k \times k$ matrix $X$ is defined by $\sigma_i = \sum_{j=1}^k x_{ij} \epsilon_j$, Chillag shows that $a \mapsto M(a, \Sigma, \Sigma)$ is an isomorphism of algebras. Hence, for any basis $\Sigma$, the $M(a, \Sigma, \Sigma)$ may be regarded as "circulants"– they are closed under sum, product, they all commute and are simultaneously diagonalized by the matrix $X$:

$$M(a, \Sigma, \Sigma) = X \text{ Diag}(a(1), a(2), \ldots, a(k))X^{-1}.$$  

In the setting of Gelfand pairs, $A$ may be taken as the algebra of bi-invariant functions, $\mathcal{E}$ the spherical functions, and $\Sigma$ the indicator functions of double cosets. This gives the usual Fourier inversion theorem.

The diagonalizations that have been natural in my applications have involved matrices indexed by groups or homogeneous spaces. This brief section shows that (as Letac (1981)
has noted) one can also work with matrices of size the dimension of the algebra. This throws away some redundancy, but often also some ease of interpretation. Chillag (1988) gives many variants of these change-of-basis formulas which can be carried over to the present setting.

E. A final example. One way to produce matrices invariant under a group is to start with a fixed matrix and symmetrize it. An elegant collection of applications of this idea appears in Andersson and Perlman (1988). They work with Hermitian matrices $M$. If $G$ is a finite group acting as permutations, define permutation matrices $\pi(s)$ for $s \in G$ and

$$\tilde{M} = \frac{1}{|G|} \sum_{s \in G} \pi(s)M\pi^*(s).$$

Thus $\pi(s)\tilde{M}\pi(s)^* = \tilde{M}$ so $\tilde{M}$ can be explicitly diagonalized (if $G$ acts transitively).

If $\lambda(M)$ is the vector of eigenvalues of $M$, it is a standard inequality that $\lambda(M)$ majorizes $\lambda(\tilde{M})$. Since the eigenvalues of $\tilde{M}$ can be calculated in closed form (at least in principle) this gives a collection of inequalities for the original eigenvalues.

Andersson and Perlman adapted this idea from an idea of Marshall and Olkin. These last authors worked with $\pi(s)$ ranging over the group of all $n!2^n$ sign change matrices. Then $\tilde{M} = Diag(M_{11}, M_{22}, \cdots M_{nn})$. Thus any symmetric convex function of the eigenvalues of $M$ is smaller than the same function applied to the diagonal entries. In particular, for positive definite symmetric matrices one gets Hadamard's inequality

$$\det(M) \leq \Pi M_{jj}.$$ 

In this last application the sign change group acts in a more general way than permutations. Most of the theory developed here goes over to such generalized symmetry. It seems like a worthwhile project to work this out carefully.
5. Matrices whose Powers are Circulants.

A. Garden variety circulants indexed by \( \mathbb{Z}(n) \) have each row shifted right by 1. Consider a matrix which has rows shifted by 2:

\[
\begin{array}{ccccc}
  a & b & c & d & e \\
  d & e & a & b & c \\
  b & e & d & e & a \\
  e & a & b & c & d \\
  c & d & e & a & b \\
\end{array}
\]

These matrices can be explicitly Jordanized. They have the property that an appropriate power is a circulant. For \( \mathbb{Z}(P) \), these results were derived in Friedman (1958), (1961) and Ablow-Brenner (1963). They are nicely exposted by Davis (1979, Chapter 6). As explained below the Jordanization also follows from standard representation theory.

These results can be generalized to the kinds of generalized circulants introduced in previous sections. The results are easiest to exposition for circulants derived from convolutions in section 3.

Let \( G \) be a finite group. Let \( A : G \rightarrow G \) be a homomorphism. Thus \( A(st) = A(s)A(t) \). Let \( f : G \rightarrow \mathbb{C} \) be any function.

Define a \( |G| \times |G| \) matrix \( M \) indexed by \( G \) as

\[
(5-1) \quad M_{st} = f(t \ A(s^{-1})).
\]

Example. If \( G = \mathbb{Z}(n) \), and \( A(j) = aj \). Then the matrix \( M \) has each row shifted right by \( a \). More general groups and homomorphisms have rows more general permutation of the first row.

Lemma. Let \( M \) be defined at (5-1). If \( A^k = Id \) then the \( k^{th} \) power of \( M \) is a convolution matrix which can be unitarily block diagonalized as in theorem 1 of section 3.

Proof: A motivated proof can be given using the language of Markov chains. Define a chain taking values on \( G \) by \( X_0 = id, X_n = A(X_{n-1}) \ast t_n \) with \( t_i \) independent and identically distributed random variables \( P\{t_i = s\} = f(s) \). This assumes \( f \) is a probability
but the identity proved will hold for all probabilities and so for all \( f \). Express the \( X_n \) in terms of \( t_i \):

\[
X_0 = id, \quad X_1 = t_1, \quad X_2 = A(t_1) \cdot t_2, \quad X_3 = A^2(t_1) \cdot A(t_2) \cdot t_3, \ldots
\]

\[
(5-2) \quad X_n = A^n(t_1) \cdot A^{n-1}(t_2) \cdots t_n.
\]

Now \( A^k = id \), so

\[
X_{kj} = \underbrace{A^{k-1}(t_1) \cdot A^{k-2}(t_2) \cdots t_k}_{\delta_1} \underbrace{A^{k-1}(t_{k+1}) \cdots t_{2k}}_{\delta_2} \underbrace{A^{k-1}(t_{(j-1)k+1}) \cdots t_{jk}}_{\delta_j}.
\]

This represents \( X_{kj} \) as the \( j \)-fold convolution of a basic step \( \delta \), where \( \delta_i \) are independent and identically distributed since \( t_i \) are.

**Remarks.**
1. If \( f \) is a class function, so is the image law of \( f \) under \( A^j \). It follows that the law of \( \delta_i \) above is a class function and so has a diagonal Fourier transform. Thus for class functions the matrix (5-1) has its \( k^{th} \) power unitarily diagonalizable. The eigenvalues of the \( k^{th} \) power are explicitly available using character theory. The eigenvalues of the original matrix are thus \( k^{th} \) roots of these values. For applications this may be all the information needed. A bit more generally, (5-2) expresses the law of \( X_n \) as a convolution of class functions for any \( n \).

2. As a special case, if \( G \) is Abelian, all functions are class functions so there is no restriction. In this case it is easy to give the explicit Jordanization of the original matrix. Briefly, the matrix is the matrix of a linear transformation of convolution by \( f \) on the vector space of all functions on the group with basis of delta functions at the group elements. Make the unitary change of basis to a basis of the characters. The homomorphism \( A \) acts on these, splitting them into orbits. Here \( A\chi(s) = \chi(As) \). These orbits span subspaces invariant under \( M \). On such a subspace \( Mx = \hat{f}(x) \cdot Ax \), so the matrix of \( M \) has \( \hat{f}(x) \) below the diagonal in the upper right hand corner as \( x \) ranges through the characters in the orbit. Such matrices are of the form \( \pi D \) with \( \pi \) a basic circulant and \( D \) diagonal. Davis (1979) gives their explicit Jordanization.
3. The shifted circulants of this section can be put into a more group theoretic framework. The powers of $A$ generate a cyclic group $c(k)$ which acts on $G$. Form the semi-direct product of $c(k)$ and $G$. This acts on $L(G)$ to give a representation. The discussion above amounts to an explicit decomposition of this representation along the lines of the Mackey-Wigner method of “little subgroups”. See Serre (1977, Chapter 8).

4. I have not thoroughly explored conditions on $f$ which give a nice form for $M$ in the Gelfand pair setting. One condition that works: take $A$ an automorphism that preserves the isotropy subgroup $K$. A natural probabilistic generalization involves random walk on homogeneous space $X = G/K$. Each time, convolve with a bi-invariant perturbation and then twist with a fixed $s \in G - K$.

B. An example. On the integers mod $p$, consider the family of processes

$$X_n = a_n X_{n-1} + b_n \pmod{p}$$

where $(a_n, b_n)$ are independent. Such processes arise in analyzing random number generators with varying multipliers. For example, the first congruential generator employed by Lehmer was $X_n = a X_{n-1} \pmod{p}$. This worked at a constant rate but was called upon at varying times (depending on the time required to complete other loops in the program). This results in an effectively random $a$.

A second familiar specialization has $a_n \equiv 1$. This is then a random walk (mod $p$). Suppose $p$ is odd and $b_n = \pm 1$ with probability $\frac{1}{2}$. Then standard techniques show that the walk tends to a uniform distribution. More quantitatively:

$$\sum |P\{X_n - j\} - \frac{1}{p}| \leq \alpha e^{-\beta n/p^2} \text{ for universal } \alpha, \beta.$$ 

This estimate shows $n$ has to be larger than $p^2$ for the walk to effectively forget where it started. This is proved in detail in Chung-Diaconis-Graham (1987) or in Diaconis (1988).

Varying slightly, consider $X_n = 2X_{n-1} + b_n \pmod{p}$, with $b_n = 0, \pm 1$ with probability $\frac{1}{3}$. Chung-Diaconis-Graham (1987) show this deterministic doubling dramatically speeds up the walk: order $\log p \log \log p$ steps are necessary and suffice to have the walk become random.
The matrix of this last Markov chain has the form described in this section. If $2^k = 1 \pmod{p}$ its $k^{th}$ power is a circulant. This forms the basis for the calculations in Chung-Diaconis-Graham (1987).

The argument of lemma 1 above allows random values of $a_n$ to be dealt with: In a basis given by the characters, the matrix of the Markov chain is a band-diagonal matrix. For example, $X_n = a^n X_{n-1} + b_n$ with $t_n = \pm 1$ with probability $\frac{1}{2}$ and $b_n$ arbitrary leads to Jacobi matrices. The extensive theory of orthogonal polynomials can now be called upon to give estimates for eigenvalues. This theory is at present in active development by myself and my students.
REFERENCES


