The Asymptotic Toeplitz-Circulant Eigenvalue Problem

Raymond H. Chan          Gilbert Strang
Department of Mathematics  Department of Mathematics
University of Hong Kong    M.I.T.

ABSTRACT

We study the solution of symmetric positive definite Toeplitz systems \( Ax = b \) by the preconditioned conjugate gradient method. The preconditioner is a circulant matrix \( C \) which copies the middle diagonals of \( A \), and each iteration uses the Fast Fourier Transform. Convergence is governed by the eigenvalues of \( C^{-1}A \) -- a Toeplitz-circulant eigenvalue problem--and it is fast if those eigenvalues are clustered. We find the limiting behavior of the eigenvalues as the dimension increases, and we prove that they cluster around \( \lambda = 1 \). For a wide class of problems the error after \( q \) conjugate gradient steps decreases like \( r^{q^2} \).

ACKNOWLEDGEMENT

We are grateful for the support of National Science Foundation grants DCR84-05506 and DCR86-02563 (RHC) and DMS84-03222 (GS) and of Army Research Office grants DAAG29-83-K0025 and DAAL03-86-K0171 (GS).
Introduction

This paper discusses a class of linear systems $Ax = b$. The matrix $A$ has the **Toeplitz Property**: down each diagonal its entries are constant. The $i,j$ entry is $a_{i-j}$, and we assume symmetry and positive definiteness. Such systems are fundamental in signal processing and time series, where the convolution form reflects invariance in time or in space (stationarity or homogeneity). Toeplitz matrices also arise directly from constant-coefficient partial differential equations, and from integral equations with a convolution kernel, when those equations are made discrete.

With periodicity, these problems can be solved quickly by Fourier transform. The convolution becomes a multiplication and deconvolution is straightforward. In the nonperiodic case, which is analogous to a problem on a finite interval (or on a bounded region in the multidimensional case), this direct solution is lost. The inverse of a Toeplitz matrix is not Toeplitz, because of the presence of a boundary and the absence of periodicity. Nevertheless the matrix $A$ is determined by only $n$ coefficients $a_0, \ldots, a_{n-1}$, rather than $n^2$. Algorithms that exploit the Toeplitz property are much faster than the $n^3/6$ operations of symmetric elimination, and **direct methods** based on the Levinson recursion formula are in constant use. Superfast direct methods, which replace Levinson's $O(n^2)$ by $O(n \log^2 n)$, are becoming competitive in practical calculations. The second author has proposed an **iterative method** [1] which it is hoped
will be fast and flexible.

Our goal in this paper is to analyze the convergence of the iterations. They use a preconditioner: the Toeplitz matrix is replaced by a circulant matrix. It retains the Toeplitz property and adds periodicity. Each diagonal in the lower triangular part wraps around into a diagonal in the upper triangular part, and the entries satisfy \( c_{ij} = c_{i-j} = c_{i-j+n} \). The distinction between Toeplitz and circulant matrices is seen (in the symmetric case) in

\[
A = \begin{bmatrix}
a_0 & a_1 & \cdots & a_{n-2} & a_{n-1} \\
a_1 & a_0 & a_1 & \cdots & a_{n-2} \\
a_2 & a_1 & a_0 & \cdots & \cdots \\
\vdots & \vdots & \vdots & \ddots & a_1 \\
a_{n-1} & a_2 & a_1 & a_0 & \cdots
\end{bmatrix}
\quad \text{and} \quad
C = \begin{bmatrix}
c_0 & c_1 & \cdots & c_2 & c_1 \\
c_1 & c_0 & c_1 & \cdots & c_2 \\
c_2 & c_1 & c_0 & \cdots & \cdots \\
\vdots & \vdots & \vdots & \ddots & c_1 \\
c_1 & c_2 & c_1 & c_0
\end{bmatrix}
\]

The diagonals containing \( c_1 \) reappear in the corners, where the matrix \( A \) has a new (and probably smaller) entry \( a_{n-1} \). To go from \( A \) to \( C \) will require changing about \( n^2/4 \) entries, and the key question in analyzing convergence will be the eigenvalues of \( C^{-1}A \).

Multiplication by a circulant \( C \) is identical to discrete convolution. The linear system \( Cz = b \) is the convolution equation \( c \ast z = b \), where \( c \) is the first column of \( C \). After a discrete Fourier transform it becomes \( \hat{c} \hat{z} = \hat{b} \). Therefore \( \hat{z} \) is given by a component-by-component division, and \( z \) is recovered from the inverse Fourier transform.
The components of \( \hat{c} \) are proportional to the eigenvalues of 
\( C \), and this convolution rule is the diagonalization of the 
circulant matrix \([2]\): \( z = F^{-1}A^{-1}Fb \). This is one of the rare 
instances in which a linear system is solved by diagonalizing 
the coefficient matrix! Normally elimination is much faster, 
but the Fourier matrix \( F \) (whose entries are the complex roots 
of unity \( F_{jk} = w^{jk} = \exp 2\pi ijk/n \), and whose columns are the 
eigenvectors of every circulant matrix) is very special.

The speed of the iterative method depends on the fact that 
multiplication by \( F \) and \( F^{-1} \)--the discrete Fourier transform 
multiplications and its inverse--can be done so quickly. Those 
are computed by the Fast Fourier Transform, which dominates each step of the 
iteration. It requires only \( n \log n \) multiplications, and the 
calculations can be done in parallel. It applies directly to 
\( C \) and our goal is to apply it also to \( Ax = b \)--reaching the 
required tolerance in a number of steps which in the best case 
is independent of \( n \).

We mention that multiplication by a Toeplitz matrix \( A \) (but 
not inversion) is also quick by the FFT. The matrix is extended 
to a circulant \( A^* \) of order \( 2n \), the vector \( d \) is completed to 
\( d^* \) by \( n \) zeros, and \( Ad \) appears in the first \( n \) components 
of \( A^*d^* \)--which is another discrete convolution. The goal is 
to replace \( A \) by \( C \) in any linear system to be solved, and to 
use \( A \) itself only in matrix multiplications.

This is exactly what is achieved by the ordinary iterative 
method \( Cx_{n+1} = (C-A)x_n + b \), and also by the preconditioned
conjugate gradient method. There is a Toeplitz multiplication on the right side and a circulant inversion on the left. We will see that the ordinary iterations can diverge; they depend on the extreme eigenvalues of $C^{-1}A$, which are not in close control. However the conjugate gradient method can be very effective. Its convergence rate also depends on the eigenvalues $\lambda_1$ of $C^{-1}A$, but not exclusively on $\lambda_1$ and $\lambda_n$. Conjugate gradient convergence is fast when the eigenvalues are clustered, and that is the property established in this paper. Thus we want to show that the circulant matrix satisfies, for large $n$, the two essential requirements for a good preconditioner:

1. $Cz = d$ can be solved quickly and stably (Theorem 1 will estimate $\|C^{-1}\|$).

2. $C$ is close to $A$ (the eigenvalues of $C^{-1}A$ are clustered near 1).

For completeness we list the steps of the preconditioned conjugate gradient method, which gives the exact solution at step $n$—but it is treated as an iterative method and stops earlier. Each iteration contains the periodic linear system with coefficient matrix $C$, the multiplication by $A$, and the two inner products which appropriately orthogonalize the directions $d_j$. Starting from $x_0 = 0$ and $r_0 = b$,

Solve $Cz_{j-1} = r_{j-1}$

$z_j = z^T_{j-1}r_{j-1}/z^T_{j-2}r_{j-2}$ (except $z_1 = 0$)

d_j = z_{j-1} + z_jd_{j-1}$ (except $d_1 = z_0$)

$\alpha_j = z^T_{j-1}r_{j-1}/d^T_jA d_j$

$x_j = x_{j-1} + \alpha_jd_j$

$r_j = r_{j-1} - \alpha_jAd_j$
The Eigenvalues of $C^{-1}A$

We want to choose $C$ close to $A$. The simplest construction is to copy the central diagonals of $A$ and bring them around to complete the circulant. Starting from the first column $a_0,\ldots,a_{n-1}$ of $A$, with $n = 2m$, the first column of $C$ is $a_0,\ldots,a_m,\ldots,a_1$. If $A$ decays quickly away from the main diagonal, then $C$ starts to do the same but increases again as we approach the corner. By substituting the vector $(1,0,\ldots,0,-1)$ into the Rayleigh quotient $x^TAx/x^TCx$, we see that the largest eigenvalue of $C^{-1}A$ is at least

$$\frac{a_0-a_{n-1}}{a_0-a_1} \leq \lambda_{\text{max}}(C^{-1}A). \quad (1)$$

This can easily exceed 2, in which case the ordinary iteration $C_{n+1} = (C-A)x_n + b$ will fail. The iterating matrix $I - C^{-1}A$ has $1-\lambda_{\text{max}}$ outside the unit circle. However the conjugate gradient method can compensate for any single outlying eigenvalue in a single iteration. The question is whether many other eigenvalues are far from unity, when corners of order $m = n/2$ are different in $C$ and $A$.

Our first results were experimental [1]. With diagonal entries $a_k = 1/(1+k)$ the eigenvalues for $n = 12$ are $.707, .957, \ldots, 1.047, 1.880$. The largest and smallest make ordinary iteration too slow, but the other eigenvalues are clustered around 1. As the order $n$ was increased, they all approached limiting values—although it was not always clear (say for $\lambda_4$) whether the limit was 1. In these experiments, and in others with different diagonals $a_k$, the $x_j$ converged quickly to $x = A^{-1}b$.

The next results were theoretical [3]. The matrices with geometrically decreasing diagonals $a_k = t^k$ had exhibited a remarkable pattern in the computations, and the eigenvalues (and eigenvectors) of $C^{-1}A$ could be verified analytically. The extremes are $1/(1+t)$ and $1/(1-t)$, and $\lambda = 1$ is a double eigenvalue. What was striking was that there were only two other eigenvalues of $C^{-1}A$. For a matrix of order
1024, each of them is repeated 510 times! Those eigenvalues are exponentially close to 1: \( \lambda = 1/(1+t^{n/2}) \) and \( \lambda = 1/(1-t^{n/2}) \). Convergence of conjugate gradients is extremely fast, and our goal is to see when this exponential clustering can be predicted.

These matrices \( A \) were studied by Kac-Murdoch-Szego [4], who observed that they have a special property: \( A^{-1} \) is tridiagonal. The generating function of \( A \) is

\[
\begin{align*}
 f &= \sum_{-\infty}^{\infty} a_k e^{ik\theta} = \sum_{-\infty}^{\infty} t^{|k|} e^{ik\theta} = \frac{1-t^2}{(1-te^{i\theta})(1-te^{-i\theta})}. 
\end{align*}
\]

It is real and positive, so that \( A \) is symmetric positive definite (for \( |t| < 1 \)). It is the reciprocal of a three-term polynomial, which underlies the fact that \( A^{-1} \) is banded and that only two limiting eigenvalues are different from 1.

The ideal approach is to learn about the spectrum of \( C^{-1}A \) from this function \( f(\theta) \). We recognize that in practice the matrices are finite and the very distant diagonals \( a_k \) will not be used. But the asymptotic properties appear to be decisive, and all the information about \( C \) and \( A \) is in \( f \). The goal is to turn a problem in operator theory into a problem in function theory.

**Uniform Invertibility of \( C \)**

Suppose that the Toeplitz matrices \( A_n \), of order \( n \), are finite sections of a fixed singly infinite positive definite matrix \( A_\infty \). The \( i,j \) entries of \( A_n \) and \( A_\infty \) are \( a_{i-j} \), and the associated function
\[ f(\theta) = \sum_{-\infty}^{\infty} a_k e^{ik\theta} \]

is real and positive. We will assume that the sequence \( a_k \) is in \( \ell^1 \), so that \( f \) belongs to the Wiener class: \( \sum |a_k| < \infty \). Then the function \( 1/f \) associated with \( A_{-1}^{-1} \) belongs to the same class, and a more precise analysis becomes possible.

The first step is to consider \( C \). The construction which copies the middle diagonals of \( A \) does not guarantee invertibility of \( C \).

Example 1: \[
\begin{bmatrix}
2 & -1 & 0 & 0 \\
-1 & 2 & -1 & 0 \\
0 & -1 & 2 & -1 \\
0 & 0 & -1 & 2 \\
\end{bmatrix}
\]

For this "second-difference matrix" the construction produces a singular \( C \). That occurs whenever \( A \) comes from discretizing an operator with no zero-order term. Dirichlet boundary conditions leave \( A \) invertible while periodic boundary conditions make \( C \) singular. It is a case when sines should replace exponentials.

The matrix \( S \) whose entries are \( \sin \frac{jk\pi}{n+1} \) has the eigenvectors of this \( A \) as its columns. The Fast Sine Transform carries out multiplication by \( S \) and \( S^{-1} \) in \( n \log n \) steps. Therefore the new preconditioner can be \( SDS^{-1} \), where the diagonal matrix \( D \) has entries \( d_j = \sum a_k e^{ik\theta}, \theta = \pi/(n+1) \).

In this example \( A_{-1} \) is only semi-definite, and \( f(\theta) = 2 - 2 \cos \theta \) is only nonnegative. Iteration is not needed because the matrix is banded, but for a full matrix with \( \sum a_k = 0 \)
the idea may be useful—conjugate gradients preconditioned by a sine transform. In this paper we stay with the positive definite case $f > 0$.

$$\begin{bmatrix}
.7 & 1/2 & 1/4 & 1/8 \\
1/2 & .7 & 1/4 & 1/8 \\
1/4 & 1/2 & .7 & 1/4 \\
1/8 & 1/4 & 1/2 & .7 \\
\end{bmatrix} \quad \text{and} \quad \begin{bmatrix}
.7 & 1/2 & 1/4 & 1/8 \\
1/2 & .7 & 1/4 & 1/8 \\
1/4 & 1/2 & .7 & 1/4 \\
1/8 & 1/4 & 1/2 & .7 \\
\end{bmatrix}$$

Example 2: $A = \begin{bmatrix}
.7 & 1/2 & 1/4 & 1/8 \\
1/2 & .7 & 1/4 & 1/8 \\
1/4 & 1/2 & .7 & 1/4 \\
1/8 & 1/4 & 1/2 & .7 \\
\end{bmatrix} \quad \text{and} \quad C = \begin{bmatrix}
.7 & 1/2 & 1/4 & 1/8 \\
1/2 & .7 & 1/4 & 1/8 \\
1/4 & 1/2 & .7 & 1/4 \\
1/8 & 1/4 & 1/2 & .7 \\
\end{bmatrix}$

The smallest eigenvalue of $A$ is $3/40$, whereas $C$ has the eigenvalue $-1/20$. The new feature is that this $A$ extends to a positive definite $A_\infty$—it is the Kac-Murdock-Szegö matrix with diagonals $a_k = (1/2)^k$ and with $a_0$ changed to $.7$.

Note that our algorithm will recognize the indefiniteness of $C$ at the first step, when $Cz_0 = r_0$ is solved by the FFT. $C$ is diagonalized so its eigenvalues are made explicit, and the algorithm can adapt by making a different choice of the circulant.

We now prove that when $A_\infty$ is positive definite and $n$ is sufficiently large, the circulants $C_n$ are uniformly positive definite. Of course the finite sections $A_n$ are also positive definite. The point is that an indefinite $C_n$—the possibility illustrated in Example 2—cannot continue as $n$ increases.

**Theorem 1** Suppose $f(\theta) = \sum_{k=-\infty}^{\infty} a_k e^{ik\theta}$ is real and positive and in the Wiener class $(\sum |a_k| < \infty)$. Then the circulants $C_n$ and $C_n^{-1}$ are uniformly bounded and positive definite for large $n$. 
Proof. The first column of $C_n$ contains by construction the numbers $a_0, \ldots, a_m, \ldots, a_1$. (For simplicity we take $n$ even and $m = n/2$.) Because the matrix is a circulant, its $j$th eigenvalue is

$$\lambda_j = a_0 + a_1 w^j + \cdots + a_m w^{jm} + \cdots + a_1 w^{j(n-1)}. \quad (3)$$

Here $w = e^{2\pi i/n}$ is the primitive $n$th root of unity. The corresponding eigenvector of $C$ (and of every circulant) is $x_j = (1, w^j, \ldots, w^{j(n-1)})$; a direct multiplication gives $Cx_j = \lambda_j x_j$. Simplifying (3) by $w^n = 1$ yields

$$\lambda_j = a_0 + a_1 (w^j + w^{-j}) + \cdots + a_{m-1} (w^{j(m-1)} + w^{j(1-m)}) + a_m w^{jm}. \quad (3')$$

Thus the eigenvalue equals the partial sum from $k = 1-m$ to $m$ of the series $\sum a_k e^{ik\theta}$, evaluated at the point $\theta = 2\pi j/n$ (where $e^{i\theta} = w^j$). Since the infinite series is absolutely convergent and its sum satisfies $f(\theta) \geq \delta > 0$, the partial sums are uniformly positive for large $n$ and the proof is complete.

Example 3: Diagonally dominant matrices, with $a_0 > 2 \sum_{k \neq 0} |a_k|$, are positive definite and so are the circulants $C_n$.

Example 4: The function $f = \cosh \theta$ is even and positive. Therefore the matrices with $a_k = (1+k^2)^{-1} \cos kw$ lead to uniformly bounded $C_n$ and $C_n^{-1}$, although the original matrix $A$ is not diagonally dominant.
The Limits of the Eigenvalues

We come now to the central problem: To study the eigenvalues of $C_n^{-1}A_n$ for large $n$. In the next sections we transform that problem in order to carry out the analysis, and a Hankel matrix appears. At the end, when the limit is found, we transform back. The result was anticipated in [3], and it may be useful to separate its statement from the details of its proof.

**Theorem 2** As $n \to \infty$, the eigenvalues of $C_n^{-1}A_n$ approach the eigenvalues of the doubly infinite problem

$$
\begin{bmatrix}
\ddots & \ddots & \\
& a_1 & \\
& a_0 & a_1 \\
& a_1 & \\
\end{bmatrix}
\begin{bmatrix}
\ddots \\
& a_1 \\
& a_0 & a_1 \\
& a_1 \\
\end{bmatrix}
= \lambda
\begin{bmatrix}
\ddots \\
& a_1 \\
& a_0 & a_1 \\
& a_1 \\
\end{bmatrix}
$$

The matrix on the right is an infinite circulant. The matrix on the left contains two back-to-back copies of the singly infinite Toeplitz matrix $A_\infty$. Somehow the source of all the difficulty—the two boundaries that prevented the finite matrices $A_n$ from being directly invertible by Fourier analysis—has reappeared in a new form.
The later sections of the paper give an expression for the limiting eigenvalues $\lambda$, by connecting them to a Hankel matrix and thus to a problem in rational approximation. That problem is approximation on the unit circle of a function $\mathbb{V}(\theta)$ derived limits of the from $f(\theta)$, and it achieves our goal--to determine the eigenvalues from $f$. (The function has $|\mathbb{V}(\theta)| = 1$ and it appears as a "phase function" in systems theory [6-7] and apparently also in methods for numerical conformal mapping.)

At the end we return to the preconditioned conjugate gradient method, to prove superlinear convergence.

Orthogonal Similarity and Hankel Matrices

The key problem is $Ax = \lambda Cx$. There is a preliminary transformation which cuts this problem in half (for $n = 2m$), since all eigenvectors are odd or even:

$$x_- = \begin{bmatrix} y \\ -Jy \end{bmatrix} \quad \text{and} \quad x_+ = \begin{bmatrix} z \\ Jz \end{bmatrix} \quad \text{with} \quad J = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}. \quad (5)$$

This property comes from the "centrosymmetry" of $C$ and $A$, and it leads to the orthogonal transformation suggested by Cantoni and Butler [5]:

$$Q = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ -J & J \end{bmatrix}.$$
This combination of the identity \( I \) and the counteridentity \( J \) will produce two diagonal blocks in \( Q^{-1}AQ \) and \( Q^{-1}CQ \). Suppose we write

\[
A = \begin{bmatrix} T & R \\ R^T & T \end{bmatrix} \quad \text{and} \quad C = \begin{bmatrix} T & S \\ S^T & T \end{bmatrix},
\]

in which \( T, R, S \) are Toeplitz of order \( m \). \( T \) is symmetric around its main diagonal \( a_0 \), and \( S \) is symmetric around \( a_m \), while \( R \) has diagonals \( a_1, \ldots, a_{n-1} \) and is displayed below.

From \( JTJ = T \) and \( JRJ = R^T \) we reach

\[
Q^{-1}AQ = QT^AQ = \begin{bmatrix} T-RJ & 0 \\ 0 & T+RJ \end{bmatrix},
\]

\[
Q^{-1}CQ = QT^CQ = \begin{bmatrix} T-SJ & 0 \\ 0 & T+SJ \end{bmatrix}.
\]

Thus the eigenvalue problem \( Ax = \lambda Cx \) splits into

\[(T-RJ)y = \lambda_-(T-SJ)y \quad \text{and} \quad (T+RJ)z = \lambda_+(T+SJ)z.\]

(Note: Those equations also appear directly when (5) and (6) are substituted into \( Ax = \lambda Cx \).) There are \( m \) eigenvalues \( \lambda_+ \) and \( m \) eigenvalues \( \lambda_- \), which together represent the \( n = 2m \) eigenvalues \( \lambda \). The eigenvectors \( x_+ \) and \( x_- \) are \( C \)-orthogonal as required \( (x_+^T C x_- = 0) \), and they are also orthogonal.

We emphasize the effect of the counteridentity \( J \). The matrices \( RJ \) and \( SJ \) are no longer Toeplitz. Instead they are
Hankel matrices. Like J itself, they are constant down each counterdiagonal. The $i,j$ entry depends on the sum $i+j$ instead of the difference $i-j$:

$$R = \begin{bmatrix} a_m & a_{n-1} \\ & \ddots & \ddots \\ & & a_1 \end{bmatrix} \quad \text{and} \quad RJ = \begin{bmatrix} a_{n-1} & a_m \\ & \ddots & \ddots \\ & & a_1 \end{bmatrix}. $$

Thus equation (8) becomes a Toeplitz-Hankel eigenvalue problem. The northeast and southwest quarters of the original Toeplitz matrices have swung into blocks on the diagonal, and in the process they have become Hankel matrices.

A Hankel matrix is determined from its counterdiagonal entries $v_1, v_2, \ldots$ by $V_{ij} = v_{i+j-1}$. In the singly infinite case its operator norm comes from the associated function $v(\theta) = \sum v_j e^{ij\theta}$, by Nehari's theorem [8-9], but not quite in the same way that the norm of a Toeplitz matrix comes from $f(\theta)$: The anti-analytic terms are at our disposal.

$$||A|| = \sup |f(\theta)| \quad \text{and} \quad ||V|| = \sup |\overline{v}(\theta)| = \inf \sup_{v_0, v_1, \ldots} |\sum v_j e^{ij\theta}|. \quad (9)$$

We emphasize that the eigenvalues in the two cases are very different. The spectrum of $A$ is the interval $[f_{\min}, f_{\max}]$, while $V$ is a compact operator for $v$ in $\ell^1$ --and its eigenvalues are the errors in a rational approximation problem.
The Limiting Equation \( Hz = \sqrt{Tz} \)

This section begins our first approach to the eigenvalues of \( C_n^{-1}A_n \) as \( n \to \infty \). We return to equation (8), where the problem was split in half, and we look at either of the two problems of order \( m \):

\[
(T + RJ)z = \lambda_+ (T + SJ)z .
\]  

(10)

Our intention is to find the limits of the eigenvalues \( \lambda_+ \) as \( n \to \infty \). Numerical experiments indicate that those limits exist.

First a simplification. Recall that the Hankel matrices \( RJ \) and \( SJ \) are identical on and below the main counterdiagonal. Their difference is the Hankel matrix

\[
H = SJ - RJ = \begin{bmatrix}
h_1 & h_2 & \cdot & 0 \\
h_2 & 0 & \cdot & 0 \\
\cdot & 0 & \cdot & 0 \\
0 & 0 & 0 & 0
\end{bmatrix}, \text{ with } h_j = a_j - a_{n-j} \text{ for } j \leq m .
\]

Subtracting \( \lambda_+ (T+RJ)z \) from both sides of (10) and dividing through by \( \lambda_+ \), the result is

\[
\frac{1-\lambda_+}{\lambda_+} (T+RJ)z = Hz .
\]

(11)

We study that form of the problem, writing \( \nu \) for the eigenvalue:

\( Hz = \nu(T+RJ)z \) with

\[
\nu = \frac{1-\lambda_+}{\lambda_+} \text{ and } \lambda_+ = \frac{1}{1+\nu} .
\]
The clustering of $\lambda_+$ around 1 corresponds to the clustering of $\nu$ around 0.

We go directly to a statement of the limiting problem, and then consider its justification. As the order $n$ increases, $T$ and $H$ approach singly infinite Toeplitz and Hankel matrices

$$
\mathbf{T} = \begin{bmatrix}
a_0 & a_1 & a_2 & \cdots \\
a_1 & a_0 & a_1 & \cdots \\
a_2 & a_1 & a_0 & \cdots \\
\vdots & \vdots & \vdots & \ddots \\
\end{bmatrix}
\quad \text{and} \quad
\mathbf{H} = \begin{bmatrix}
a_1 & a_2 & a_3 & \cdots \\
a_2 & a_3 & \cdots & \ddots \\
a_3 & \cdots & \ddots & \ddots \\
\vdots & \vdots & \ddots & \ddots \\
\end{bmatrix}
$$

This is strong convergence of operators, if we think of all operators as acting on $\ell^1$ or $\ell^2$. The matrices $T$ and $H$ represent operators $T_m$ and $H_m$ which act as the zero operator after the first $m$ components:

$$
T_m = \begin{bmatrix} T & 0 \\
0 & 0 \\
\end{bmatrix} \quad \text{and} \quad
H_m = \begin{bmatrix} H & 0 \\
0 & 0 \\
\end{bmatrix}
$$

**Lemma 1** The Hankel sequence $H_m$ converges uniformly to $H$: $\|H-H_m\| \to 0$. The Toeplitz sequence $T_m$ converges strongly to $T$, and the sequence $(RJ)_m$ converges strongly to the zero operator:

$$
\|Tx-T_mx\| \to 0 \quad \text{and} \quad \|\begin{bmatrix} R & 0 \\
0 & 0 \\
\end{bmatrix}_m x\| \to 0 \quad \text{for each vector } x.
$$

Proof. For the Hankel matrices $H-H_m$ the norm is given by (9):
\[ \|H - H_m\| \leq \sup_{1 \leq n \leq m-1} | \sum_{j=1}^{\infty} a_{n-j} e^{ij\theta} + \sum_{j=m}^{\infty} a_j e^{ij\theta} | \]
\[ \leq 2 \sum_{j=m}^{\infty} |a_j| \to 0 \text{ as } m \to \infty. \] (12)

For the others, the estimate for \((RJ)_m\) is typical. With \(N\) fixed, the \(N\) by \(N\) submatrix in the upper left corner has operator norm going to zero. (Its entries are \(a_{n-j}\), displayed earlier.) The larger entry \(a_1\) is in the \((m,m)\) position and it moves out as \(m \to \infty\). There is convergence to zero for each fixed \(x\) but not uniform convergence--pointwise convergence but not norm convergence. We give here only the simplest consequence for the eigenvalues:

**Theorem 3** Each eigenvalue \(\nu\) of the infinite Hankel-Toeplitz problem

\[ H \bar{z} = \nu T \bar{z} \] (13)

is a limit of eigenvalues of the finite problems \(Hz = \nu(T+RJ)z\).

Proof. The eigenvector \(\bar{z}\) will be our fixed vector \(x\). Then the strong convergence noted above gives

\[ \|H_m \bar{z} - \nu(T+RJ)_m \bar{z}\| \to 0. \]

Looking only at the \(m\) by \(m\) submatrices, where the operators are nonzero, and at the vector \(z_m\) taken from the first \(m\) components of \(\bar{z}\), this is

\[ \|H_m z_m - \nu(T+RJ)z_m\| \to 0. \] (14)
Writing B for T + RJ, this means that $\|(H-\overline{\nu}B)^{-1}\| \to \infty$ as $m \to \infty$. In case $H - \overline{\nu}B$ is singular it means that $\overline{\nu}$ is an exact eigenvalue of the finite problem.

Because this is the generalized eigenvalue problem, with a matrix B on the right side instead of the identity, we need one extra step. Notice that these matrices $B = T + RJ$ are uniformly invertible. They are diagonal blocks in the original $Q^T AQ$ of equation (7). The matrices A are uniformly invertible because they are finite sections of a positive definite singly infinite Toeplitz matrix (it is $\bar{T}$!). Therefore we can convert to the ordinary eigenvalue problem for $B^{-1/2}HB^{-1/2}$, maintaining symmetry with the symmetric positive definite square root of $B$:

$$H - \overline{\nu}B = B^{1/2}(B^{-1/2}HB^{-1/2} - \overline{\nu}I)B^{1/2}.$$ 

In the $l^2$ matrix norm this yields

$$\|(H-\overline{\nu}B)^{-1}\| \leq \|B^{-1}\| \max \frac{1}{|\nu_i - \overline{\nu}|}.$$  (15)

Since $\|B^{-1}\|$ is bounded and the left side approaches infinity, we conclude that some eigenvalue $\nu_i$ of the finite problem converges to $\overline{\nu}$ as $m \to \infty$.

That completes an argument which could be made more precise. The fact that the eigenvectors belong to $l^1$ was established by Adamjan-Arov-Krein [10] and pointed out to us by Nick Trefethen—who also led us into the eigenvalue theory
of Hankel operators. In the Kac-Murdock-Szego example \( a_k = t^k \), where nearly half the eigenvalues are \( \lambda = 1/(1+t^m) \), there is geometric convergence to \( \bar{\lambda} = 1 \). In terms of \( \nu = (1-\lambda)/\lambda = t^m \), it is geometric convergence to the limit \( \bar{\nu} = 0 \). In that special example we will show that the only nonzero \( \bar{\nu} \) is the number \( t \), corresponding to the eigenvalue \( \lambda = 1/(1+t) \) that stays away from 1.

We now check on the number of eigenvalues of the finite problem that are outside an interval around 1, after a remark on the other half of our original problem—the \( \lambda^- \) eigenvalues of \( C^{-1}A \).

The Twin Problem \( (T-RJ)y = \lambda_-(T-SJ)y \)

The same simplification as in (11), now subtracting \( \lambda_-(T-RJ)y \) from both sides of the twin problem and dividing by \( \lambda_- \), yields

\[
\frac{(1-\lambda_-)}{\lambda_-} (T-RJ)y = -\mu y.
\]

(16)

In this case we set

\[
\mu = \frac{1-\lambda_-}{\lambda_-} \text{ and thus } \lambda_- = \frac{1}{1+\mu}.
\]

The finite problem is \( H y = -\mu(T-RJ)y \). Exactly as before, the strong convergence of \( H \) to \( \bar{H} \), \( T \) to \( \bar{T} \), and \( RJ \) to 0 leads to the limiting problem

\[
\bar{H} y = -\mu \bar{T} y.
\]
This is identical to $Hz = \sqrt{\lambda} z$ except for the sign change: $\mu = -\tilde{\nu}$. As before, each $\mu$ is the limit of eigenvalues $\lambda$ of the finite problem. There is an interesting corollary for the original eigenvalues $\lambda_-$ and $\lambda_+$: In the limit there are pairs $\lambda_-$ and $\lambda_+$ which satisfy

$$\frac{1}{\lambda_-} + \frac{1}{\lambda_+} = 2.$$  \hfill (17)

The left side approaches $(1+\mu) + (1+\tilde{\nu}) = 2$. This was first noticed by Alan Edelman in MATLAB experiments.

With this pair of limit problems, we have completed the proof of Theorem 2. The splitting into odd and even eigenvectors of that doubly infinite eigenvalue problem gives exactly (13) and its twin, with the same similarity $Q$ and change from $\lambda$ to $\sqrt{\lambda}$ and $\mu$ as in the finite case. A corresponding limit could be found for other constructions of the circulant $C$ -- and for multidimensional Toeplitz equations, in which our algorithm may be particularly useful. We concentrate here on understanding more clearly the asymptotic behavior for this choice of $C^{-1}A$.

**The Clustering of the Spectrum of $C^{-1}A$**

The limit problem gives us precise information about the asymptotic behavior of the algorithm. Even without that knowledge we can show that the eigenvalues of $C^{-1}A$ cluster at 1, by using the theory of collectively compact operators:
A family \( S \) of bounded operators on \( \ell^2 \) is collectively compact if the set \( \{Kx : K \in S, \|x\| \leq 1\} \) is relatively compact in \( \ell^2 \).

This applies to \( \{H_m\} \) and \( \{H-H_m\} \) when \( \Sigma |h_j| < \infty \). For every \( \epsilon \) we can choose \( n(\epsilon) \) so that outside the leading submatrix of that order, all these matrices have norm less than \( \epsilon \). Anselone [11] has established the following consequences for approximation of the spectrum:

**Lemma 2** [11, Thm. 4.8] If an open set contains the spectrum of \( \bar{H} \), it also contains the spectrum of \( H_m \) for all sufficiently large \( m \).

**Lemma 3** [11, Thm. 4.14] The number of eigenvalues of \( H_m \) in a small ball around a nonzero eigenvalue \( \mu \) of \( \bar{H} \) is, for large \( m \), no greater than the multiplicity of \( \mu \). Since \( \bar{H} \) is compact, those multiplicities are finite. (Our \( \bar{H} \) is the norm limit of the finite-dimensional \( H_m \), by (12). From Hartman's theorem [9] it remains compact if its associated function is only continuous, and not necessarily in the Wiener class.) It follows easily that the eigenvalues of \( H_m \) are clustered around zero.

**Theorem 4** There exist \( M(\epsilon) \) and \( N(\epsilon) \) such that for \( m > M \), at most \( N \) eigenvalues of \( H_m \) and of \( (T+RJ)^{-1}H \) have absolute value exceeding \( \epsilon \). Only a finite number of eigenvalues of \( \bar{H} \) have \( |\mu| > \epsilon \). Therefore Lemma 2 allows us to locate the eigenvalues of \( H_m \) for large \( m \), and Lemma 3 allows us to count them. The total cannot exceed the total for \( \bar{H} \).
It is this count that was not included in Theorem 3 on the limiting values. As there, we have to handle the matrices $B = T + RJ$ on the right side of the eigenvalue problem. They are diagonal blocks in $Q^T AQ$, and their eigenvalues are between $f_{\text{min}}$ and $f_{\text{max}}$ (where $f = \sum a_k e^{1k\theta}$). Therefore the eigenvalues of $(T+RJ)^{-1}H$ are also counted by Theorem 4.

Thus the eigenvalues $\lambda$ of $C^{-1}A$ cluster around 1. Only a fixed number, independent of $n$, can be outside $(1-\epsilon, 1+\epsilon)$. Now we go back to the equation $Hz = \sqrt{\nu T}z$ which identifies the asymptotic limits of the eigenvalues.

**The Eigenvalues of $T^{-1}H$**

It is this singly infinite problem that is attractive to work with, because all the information is in the function $f(\theta) = \sum a_k e^{1k\theta}$. The difficulty is to extract it. One preliminary difficulty is that we still have a generalized eigenvalue problem, with $T$ on the right hand side: $Hz = \sqrt{\nu T}z$. The inverse of $T$ is not Toeplitz, and $T^{-1}H$ is not Hankel, but there is a way to preserve those properties—by factoring $T$ and putting part of $T^{-1}$ on each side of $H$:

$$T = WW^T \text{ with } W = \text{upper triangular Toeplitz operator.}$$

This is equivalent to representing the positive function $f$ as a square:
From the Wiener theory the sequence \( \ldots, w_{-2}, w_{-1}, w_0 \) is in \( L^1 \).

The function \( w \), with no zeros on or outside the unit circle, corresponds to \( W \) in the same way that \( f \) corresponds to \( T \):

\[
W = \begin{bmatrix}
w_0 & w_{-1} & w_{-2} \\
w_1 & w_0 & w_{-1} \\
w_0 & & \\
\end{bmatrix}
\]

Note that \( w \) is anti-analytic, with negative \( k \). The same properties hold for the matrix \( U = W^{-1} \), associated with the reciprocal function \( u = w^{-1} \):

\[
u(\theta) = \sum_{-\infty}^{0} u_k e^{ik\theta} = \left( \sum_{-\infty}^{0} w_k e^{ik\theta} \right)^{-1}.
\]

These functions will take us from the Toeplitz-Hankel product \( T^{-1}H \) (the eigenvalue problem with \( T \) on the right side) to a single Hankel matrix \( V \). It has the same eigenvalues \( \overline{V} \) —and to study them we need to know its associated function.

**Theorem 5** The matrix \( T^{-1}H = W^{-T}W^{-1}H = U^TUH \) is similar to the Hankel matrix \( V = UHU^T \). The associated function is

\[
v(z) = \sum_{1}^{\infty} v_k z^k = \text{analytic part of} \frac{\sum_{-\infty}^{0} a_k z^k}{\left( \sum_{-\infty}^{0} w_k z^k \right)^2} = \text{analytic part of} \frac{w}{w}.
\]

\[(18)\]
Proof. Certainly $U^T U H$ is similar to $V = U H U^T$. To verify that this is a Hankel matrix, and to identify its function $v(z)$, we carry out an example with two nonzero coefficients $h_1, h_2$ and $u_0, u_1$. (The general rule is that upper triangular Toeplitz times Hankel is Hankel, and Hankel times lower triangular Toeplitz is Hankel; we hope the example will be convincing.)

The infinite matrices can be cut off after three rows and columns:

$$
\begin{bmatrix}
  u_0 & u_1 & 0 \\
  0 & u_0 & u_1 \\
  0 & 0 & u_0 \\
\end{bmatrix}
\begin{bmatrix}
  h_1 & h_2 & 0 \\
  0 & h_2 & 0 \\
  0 & 0 & 0 \\
\end{bmatrix}
\begin{bmatrix}
  u_0 & 0 & 0 \\
  u_1 & u_0 & 0 \\
  0 & u_1 & u_0 \\
\end{bmatrix}
= \begin{bmatrix}
  h_1 u_0^2 + 2 h_2 u_1 u_0 & h_2 u_0^2 & 0 \\
  h_2 u_0^2 & 0 & 0 \\
  0 & 0 & 0 \\
\end{bmatrix}.
$$

The corresponding multiplication of functions correctly gives

$$
\text{analytic part of } (h_1 z + h_2 z^2)(u_0 + u_1/z)^2 = h_1 u_0^2 z + 2 h_2 u_0 u_1 z + h_2 u_0^2 z^2 = v(z).
$$

This verifies the first line of (18), since $u = 1/w$, and we only mention a surprise in comparison with the Toeplitz case. If the matrix $H$ in the middle were Toeplitz, with associated function $h$, then the product $U H U^T$ would again be Toeplitz, but its associated function would be $h |u|^2$ -- where in the Hankel case it is the analytic part of $hu^2$. Note that the analytic part is taken to start at the linear term in $z$ -- because the Hankel matrix starts at $v_1$.

Now we look at the last step in (18)--the elegant form for $v$. It was noticed by Alan Edelman, and the second author
observed that it follows immediately from the line above: We can add the anti-analytic terms $\sum_{k=0}^{\infty} a_k z^k$ to the numerator to obtain $f$, without changing the analytic part. Therefore

$$v$$ is the analytic part of $$\frac{f_v}{w^2} = \frac{w \bar{w}}{w^2} = \frac{\bar{w}}{w}.$$  

Now the asymptotic problem is the spectrum of $V$. We recall that it is a compact operator, so its eigenvalues $\bar{v}$ cluster at zero. It is also a Hankel operator, and the eigenvalues are connected to a part of mathematics that looks entirely separate: approximation by rational functions on the unit circle.

**Hankel Eigenvalues and Rational Approximation**

We recall the main facts from [10] and [12]. The singular values of a Hankel matrix $V$ --which in our symmetric case means the absolute values $|\bar{v}_0| \geq |\bar{v}_1| \geq \cdots$ --are the errors in the best approximation of the function $v(z)$ from the class $\tilde{R}_n$. The error is measured in the sup norm on the unit circle:

$$|\bar{v}_n| = \inf_{\text{re} \tilde{R}_n} \|v(z) - \tilde{F}(z)\|_\infty.$$  \hspace{1cm} (19)

It is important that $\tilde{R}_n$ is larger than the class $R_n$ of rational functions (ratios of polynomials of degree $n$). The anti-analytic part of $\tilde{F}$ is arbitrary: $\tilde{R}_n = R_n + \text{anti-analytic}$. Thus the approximation problem for $v$ is the same as
for \( \bar{w}/w \) -- because the analytic parts are the same. The approximants have the form

\[
\hat{f} = \sum_{k=-\infty}^{n} d_k z^k / \sum_{k=0}^{n} e_k z^k,
\]

where the numerator is bounded in any bounded subset of \( \{ |z| \geq 1 \} \). The optimal error curve \( v - \hat{f}^* \) is a circle of radius \( |\nu_n| \) around the origin. Except in degenerate cases its winding number is \( 2n + 1 \) [12b].

Thus the estimation of the asymptotic eigenvalues \( \nu \) is equivalent to a problem in rational approximation. That may not be a simplification; most applications go the other way.

Of course there is a special (but important) case when \( f \) itself is rational. That corresponds to banded matrices times banded inverses, and the Hankel matrix \( V \) has only finitely many nonzero eigenvalues. That was the case for the Kac-Murdock-Szegö example, in which

\[
f = \frac{1-t^2}{|1-tz|^2}, \quad u = \frac{1-t/z}{(1-t^2)^{1/2}}, \quad v = \frac{(1-t^2)tz}{1-tz}.
\]  \hspace{1cm} (20)

Here \( v \) maps the unit circle to a circle with center at \( t^2 \) and radius \( t \). Therefore the best constant approximation to \( v \) is \( r = t^2 \), and the error is \( |\nu| = t \). The corresponding \( \lambda \)'s are \( 1/(1+t) \). That is the correct limit, from our explicit calculation, of the original eigenvalue problem \( Ax = \lambda Cx \) -- in which all other eigenvalues approached \( 1 \).
A second example will bring out the important point, which is the very rapid decrease of the errors $|\vec{v}|$ in rational approximation. Suppose the matrix is not tridiagonal but pentadiagonal. The function $f$ will be

$$f = |(1-te^i\theta)(1-se^i\theta)|^2.$$  

As $s$ approaches zero this goes back to the previous example (or more precisely to its inverse—it is a corollary of (18) that $f$ and $f^{-1}$ lead to the same results, and now it is $A$ instead of $A^{-1}$ that is banded). There are two nonzero errors $|\vec{v}_0|$ and $|\vec{v}_1|$, after which the approximation is exact and all eigenvalues approach $\vec{v} = 0$ (which is $\lambda = 1$). Those two errors are given by the quadratic equation

$$v^2 - v(t+s)(ts-1) - t^2s^2 = 0.$$  

(21)

For $s = t = \frac{1}{2}$ the limiting eigenvalues are $|\vec{v}_0| = .825$ and $|\vec{v}_1| = .076$. Thus it is not the case that the two limits $|\vec{v}|$ are near $t$ and $s$. For $t = s$ the leading terms in $|\vec{v}|$ are $2t$ and $t^{3/2}$, and it is this cube of $t$ which indicates rapid decrease. A similar phenomenon was noticed by Trefethen [12, Theorem 6.3].

A future paper will come back to the rational approximation problem. Here we turn to the consequences for the iterative algorithm when there is rapid decrease of the $|\vec{v}|$. 
The Rate of Convergence to $x = A^{-1}b$

The conjugate gradient method is a recursive calculation of a sequence of projections. After $q$ cycles, $x_q$ is as close as possible (in an appropriate norm) to the solution $x = A^{-1}b$, among all vectors in the Krylov subspace spanned by $C^{-1}b$, $C^{-1}AC^{-1}b$, $C^{-1}AC^{-1}AC^{-1}b, \ldots$ That makes possible an estimate of the error $e_q = x - x_q$:

$$\|e_q\| \leq \left[\min_{P_q} \max_{\lambda} |P_q(\lambda)|\right] \|e_0\|.$$  \hspace{1cm} (22)

The maximum is taken over the eigenvalues of $C^{-1}A$. The minimum is over polynomials of degree $q$ with constant term 1. The problem is to estimate that minimum.

One choice of $P_q$, if the eigenvalues are known to lie in the interval $[\alpha, \beta]$, is the Chebyshev choice: the best polynomial when the maximization is taken over all $\alpha \leq \lambda \leq \beta$. That has the drawback of using only $\alpha$ and $\beta$; it cannot take advantage of clustering of the eigenvalues. By scale invariance, the estimate depends only on the condition number $\beta/\alpha$.

At the other extreme, we can choose $P_q$ to annihilate the $q$ extreme eigenvalues. In our problem those eigenvalues come in pairs $\lambda_+$ and $\lambda_-$, on opposite sides of 1, and such a pair is annihilated by the factor

$$p(x) = (1 - \frac{x}{\lambda_+})(1 - \frac{x}{\lambda_-}).$$ \hspace{1cm} (23)
Between those roots the maximum of $|p|$ is attained at the average $x = \frac{1}{2}(\lambda_+ + \lambda_-)$, where $|p| = (\lambda_+ - \lambda_-)^2/4\lambda_+\lambda_-$. It is easy to find the asymptotic convergence rate of the conjugate gradient method, in the important case when the rational approximation errors decrease geometrically to zero:

$$|\bar{v}_j| = O(r^j) \text{ with } r < 1.$$  \hspace{1cm} (24)

**Theorem 6** Suppose (24) holds, which depends on the original $f$. Then the errors in the circulant-preconditioned conjugate gradient method decrease, asymptotically as $n \to \infty$, at the superlinear rate

$$||e_q|| \leq c q r^{q^2/4 + q/2} ||e_0||.$$  \hspace{1cm} (25)

The decisive factor is $r^{q^2/4}$. To find it we note that

$$\lambda_+ = 1/(1 + \nu)$$

$$|p_j| \leq \frac{(\lambda_+ - \lambda_-)^2}{4\lambda_+\lambda_-} = \frac{\nu_j^2}{1 - \nu_j^2} \leq c^2 r^2 j \text{ by (24)}.$$

The polynomial $p_{2q} = p_1 p_2 \cdots p_q$ annihilates the $q$ extreme pairs of eigenvalues and satisfies

$$|p_{2q}(\lambda)| \leq c q r^{q^2} \cdots r^2 q = c q r^{q^2 + q}.$$

This holds for all $\lambda$ in the inner interval between $\lambda_{+q}$ and $\lambda_{-q}$, where the remaining eigenvalues are. Therefore (25) comes directly from (22), after a change from $2q$ to $q$.

**Superlinear Convergence**

In a sense the convergence rate $r^{q^2/4}$ was the object of this paper. By connecting the eigenvalues of $C^{-1}A$ to the spectrum
of the Hankel matrix $V$, and by assuming (24), we were led to that unusual rate. Trefethen observed that (24) will hold if the original $f(z) = \sum a_j z^j$ is analytic in a neighborhood of $|z| = 1$ --and further, that this condition is far from necessary. We expect that rate for a wide class of applications, but we can also prove superlinear convergence in its absence.

For this we modify the polynomial $P_q$ as in [13-14]. It will annihilate $N$ pairs of extreme eigenvalues, by including $N$ of the quadratic factors (23). The remaining factor of degree $q - 2N$ will be the Chebyshev choice on the interval between $\lambda_{-N}$ and $\lambda_{+N}$, which contains the remaining eigenvalues. We know from Theorem 4 (which applied to all $f$ in the Wiener class) that for any fixed $\epsilon$, all but $N(\epsilon)$ eigenvalues are within $\epsilon$ of 1. The $N$ extreme eigenvalues are in the fixed interval

$$\frac{f_{\min}}{f_{\max}} - \epsilon < \lambda(C^{-1}A) < \frac{f_{\max}}{f_{\min}} + \epsilon$$

for large $n$, applying Theorem 1 to $C^{-1}$ and the elementary bounds $f_{\min} \leq \lambda(A) \leq f_{\max}$ to $A$. The quadratics $p$ that annihilate those extreme pairs are bounded by a fixed constant $K$ on $(1-\epsilon, 1+\epsilon)$. The other (Chebyshev) factor of degree $q-2N$ is of order $\epsilon^{q-2N}$ on that interval. Therefore the polynomial $P_q$ satisfies a crude bound

$$|P_q(\lambda)| \leq c \epsilon^{q-2N} K^N \leq C(\epsilon) \epsilon^q$$

for all eigenvalues $\lambda$ of $C^{-1}A$, when the order $n$ is sufficiently large.
It follows from (22) that \( \| e_q \| \leq C(\epsilon) e^{q\| e_0 \|} \). The number of iterations to achieve a fixed accuracy remains bounded as the matrix order \( n \) is increased. Each iteration requires \( O(n \log n) \) operations using the FFT. Therefore the work to obtain the solution \( x = A^{-1}b \) to given accuracy \( \delta \) is \( c(f,\delta)n \log n \). The real question is the efficiency in practice, and we will be glad to see this straightforward algorithm tried on genuine applications.

REFERENCES


