Design and Analysis of Toeplitz Preconditioners

Ta-Kang Ku and C.-C. Jay Kuo, *Member, IEEE*

***Abstract*—The solution of symmetric positive definite Toeplitz systems $Ax = b$ by the preconditioned conjugate gradient (PCG) method was recently proposed by Strang and analyzed by R. Chan and Strang. The convergence rate of the PCG method depends heavily on the choice of preconditioners for the given Toeplitz matrices. In this paper, we present a general approach to the design of Toeplitz preconditioners based on the idea to approximate a partially characterized linear deconvolution with circular deconvolutions. All resulting preconditioners can therefore be inverted via various fast transform algorithms with $O(N \log N)$ operations. For a wide class of problems, the PCG method converges in a finite number of iterations independent of N so that the computational complexity for solving these Toeplitz systems is $O(N \log N)$.**

I. Introduction

THE solution of an $N \times N$ symmetric positive definite (SPD) Toeplitz system $Ax = b$ arises in many digital signal processing applications. Direct methods based on Levinson recursion formula [10], [16] with $O(N^2)$ complexity are well known. Superfast algorithms with $O(N \log^2 N)$ complexity have also been investigated by researchers [1]–[3], [14]. More recently, Strang [22] proposed to use an iterative method, i.e., the preconditioned conjugate gradient (PCG) method, to solve the SPD Toeplitz system. The PCG method has a computational complexity proportional to $O(N \log N)$ for a large class of problems [22], and is therefore competitive with any direct method. Another advantage with the PCG method is that it is highly parallelizable whereas most direct methods cannot be parallelized as easily.

An iterative method for solving the SPD system $Ax = b$ can be derived by minimizing the quadratic functional $\frac{1}{2}x^TAx - b^Tx$ with the conjugate gradient (CG) method, and the unique minimum gives the desired solution. The convergence rate of the CG method depends on the spectrum of A. Generally speaking, the CG method converges faster if A has a small condition number of clustered eigenvalues. In order to accelerate its convergence rate, a preconditioning step is often introduced at each CG iteration, which leads to the PCG method. A good preconditioner for A is a matrix P that approximates A well (in the sense that the spectrum of the preconditioned matrix $P^{-1}A$ is clustered around 1 or has a small condition number), and for which the matrix–vector product $P^{-1}v$ can be computed efficiently for a given vector v. With such a preconditioner, one then solves in principle the preconditioned system $\tilde{A}\tilde{x} = \hat{b}$, where $\tilde{A} = P^{1/2}AP^{-1/2}$, $\tilde{x} = P^{1/2}x$ and $\hat{b} = P^{-1/2}b$, by the CG method [13]. The idea of preconditioning is a simple one but is now recognized as critical to the effectiveness of the PCG method.

A Toeplitz preconditioner has been proposed by Strang [22], and analyzed by Chan and Strang [5], [7]. Strang's preconditioner S is obtained by preserving the central half diagonals of A and using them to form a circulant matrix. Since S is circulant, the matrix–vector product $S^{-1}v$ can be conveniently computed via fast Fourier transform (FFT) with $O(N \log N)$ operations. It has been shown [5]–[7] that for a large class of matrices (called the Wiener class), the spectrum of $S^{-1}A$ is clustered around 1 except a finite number of outliers.

In constructing Strang's preconditioner S, only half the elements of A is used. In order to use all elements of A, Chan [8] proposed another Toeplitz preconditioner C. It is, by definition, the circulant matrix which minimizes the Frobenius norm $\|R - A\|_F$ over all circulant matrices R. This turns out to be a simple optimization problem, for which a closed-form solution exists. The elements of C can be computed directly from the elements of A by a simple formula. However, Chan's preconditioner C does not necessarily improve the convergence performance of the PCG method in comparison with Strang's preconditioner S [6].

This research was motivated by seeking another direction to generalize Strang's preconditioner so that all elements of A can be effectively used. Our study leads to a general approach for constructing Toeplitz preconditioners. Strang's and Chan's preconditioners can be viewed as special cases under this framework. We also obtain new preconditioners with better performance for Toeplitz matrices generated by rational functions. Our idea can be simply stated as follows. We formulate the inverse Toeplitz matrix–vector product as a partially characterized linear deconvolution problem, which can be approximated by a certain circular deconvolution. The preconditioning step corresponds to the implementation of the approximating circular deconvolution. Thus, all resulting preconditioners can be inverted with $O(N \log N)$ operations via various fast transform algorithms such as FFT, fast cosine transform, or fast sine transform. One interesting consequence of our approach is that it allows even noncirculant preconditioning matrix P, which is nevertheless related to a circulant matrix of size $2N \times 2N$.

Manuscript received May 8, 1990; revised December 5, 1990. This work was supported by the USC Faculty Research and Innovation Fund and a National Science Foundation Research Initiation Award.

The authors are with the Signal and Image Processing Institute and the Department of Electrical Engineering-Systems, University of Southern California, Los Angeles, CA 90089-0272.

IEEE Log Number 9104035.

The outline of this paper is as follows. The PCG algorithm for solving a symmetric positive definite system of equations is briefly reviewed in Section II. Then, we propose a general framework to construct Toeplitz preconditioners by exploiting the relationship between linear and circular deconvolutions in Section III. In particular, a class of new preconditioners K_i, $i = 1, 2, 3, 4$, which use all elements of A are described. In Section IV, we show the relationship among K_i's and prove the positive definite property of K_i's and the clustering effect of the spectrum of $K_i^{-1}A$. In Section V, we give some numerical results and compare the performance of different preconditioners. The efficiency of new preconditioners K_i is demonstrated.

II. The PCG Method for Toeplitz Systems

With the initialization

$$\text{arbitrary } \boldsymbol{x}_0, \qquad \boldsymbol{r}_0 = \boldsymbol{p}_0 = \boldsymbol{b} - A\boldsymbol{x}_0, \qquad \text{and} \quad \beta_1 = 0$$

the kth iteration ($k = 1, 2, \cdots$) of the PCG algorithm [13] consists of the following two steps:

Step 1: Preconditioning. Solve

$$P\boldsymbol{z}_{k-1} = \boldsymbol{r}_{k-1}$$

for $\boldsymbol{z}_{k-1}$.

Step 2: CG iteration. Compute

$$\beta_k = (\boldsymbol{z}_{k-1}, \boldsymbol{r}_{k-1})/(\boldsymbol{z}_{k-2}, \boldsymbol{r}_{k-2})$$

$$\boldsymbol{p}_k = \boldsymbol{z}_{k-1} + \beta_k \boldsymbol{p}_{k-1}$$

$$\alpha_k = (\boldsymbol{z}_{k-1}, \boldsymbol{r}_{k-1})/(\boldsymbol{p}_k, A\boldsymbol{p}_k)$$

$$\boldsymbol{x}_k = \boldsymbol{x}_{k-1} + \alpha_k \boldsymbol{p}_k$$

$$\boldsymbol{r}_k = \boldsymbol{r}_{k-1} - \alpha_k A\boldsymbol{p}_k.$$

It is easy to see that each computational unit above (the scalar–vector and vector–vector products and vector addition), except the Toeplitz matrix–vector product $A\boldsymbol{p}_k$ and the preconditioning $P^{-1}\boldsymbol{r}_{k-1}$, requires $O(N)$ operations. Since we can view $A\boldsymbol{p}_k$ as a circular convolution between two extended periodic sequences, the Toeplitz matrix–vector product can be computed via FFT with $O(N \log N)$ operations. We will show that the preconditioning $P^{-1}\boldsymbol{r}_{k-1}$ can also be achieved by various fast transform algorithms with $O(N \log N)$ operations in Section III. Consequently, each PCG iteration requires $O(N \log N)$ operations. Since fast transform algorithms are highly parallelizable, the above PCG method can be parallelized in a straightforward way. The parallel time complexity can be reduced to $O(\log N)$ when $O(N)$ processors are used.

For the PCG method to be attractive, it must converge fast. The convergence rate of the PCG method depends on the eigenvalue distribution of the preconditioned matrix $P^{-1}A$. Suppose that we measure the error $\boldsymbol{x}_k - \boldsymbol{x}^*$, where $\boldsymbol{x}^*$ is the exact solution of $A\boldsymbol{x} = \boldsymbol{b}$, with

$$R(\boldsymbol{x}_k) = (\boldsymbol{x}_k - \boldsymbol{x}^*)^T P^{-1} A (\boldsymbol{x}_k - \boldsymbol{x}^*) \tag{1}$$

which is the square of a matrix norm. It can be shown that the reduction of $R(\boldsymbol{x}_k)$ [17] by the PCG method is

$$R(\boldsymbol{x}_{k+1}) \le \min_{G_k} \max_{\lambda_i} (1 + \lambda_i G_k(\lambda_i))^2 R(\boldsymbol{x}_0) \tag{2}$$

where the minimum is taken over any polynominal of degree k, and the maximum is taken over all eigenvalue λ_i of $P^{-1}A$.

It is typical that the eigenvalues of the preconditioned Toeplitz matrices are clustered in a small interval $(1 - \epsilon, 1 + \epsilon)$, where ϵ is called the clustering radius, except α outliers $\lambda_1, \lambda_2, \cdots, \lambda_\alpha$. For such a case, we are able to characterize the convergence rate more precisely. Let us choose $G_{\alpha+\beta}(\lambda)$ such that

$$1 + \lambda G_{\alpha+\beta}(\lambda) = (1 - \lambda)^{\beta+1}\left(1 - \frac{\lambda}{\lambda_1}\right)\left(1 - \frac{\lambda}{\lambda_2}\right)\cdots\left(1 - \frac{\lambda}{\lambda_\alpha}\right). \tag{3}$$

The inequality (2) can be simplified as

$$R(\boldsymbol{x}_k) \le C\epsilon^{2(k-\alpha)} R(\boldsymbol{x}_0), \qquad \text{for } k > \alpha \tag{4}$$

where

$$c \approx \left(1 - \frac{1}{\lambda_1}\right)^2\left(1 - \frac{1}{\lambda_2}\right)^2\cdots\left(1 - \frac{1}{\lambda_\alpha}\right)^2. \tag{5}$$

In deriving (4), we assume that α outliers are annihilated by the first α iterations and the reduction of $R(\boldsymbol{x}_k)$ simply depends on eigenvalues clustered around one. It implies that, when $k > \alpha$, $R(\boldsymbol{x}_k)$ can be reduced at least by a factor ϵ^2 per iteration in average. Thus, the number of outliers α and the clustering radius ϵ provide some characterization for the convergence rate of the PCG method. For rationally generated Toeplitz matrices, we find that there exist strong regularities on the values of α and ϵ so that they can be predicted quite accurately. These will be detailed in Section V.

III. Design of Toeplitz Preconditioners

A good preconditioner P for an $N \times N$ symmetric Toeplitz matrix A should satisfy the following two criteria: i) P can be inverted effectively; and ii) P approximates A well in the sense that $P^{-1}A$ has a small condition number or that the spectrum of $P^{-1}A$ has a certain clustering feature. In this section, we present a systematic approach to the design of a class of preconditioners P, which can be inverted directly via various fast transform algorithms with $O(N \log N)$ operations. The spectral property of $P^{-1}A$ will then be discussed in Section IV.

A. Motivation: A Convolutional Interpretation

Let $\boldsymbol{u}_N = (u_0, u_1, \cdots, u_{N-1})^T$ and $\boldsymbol{v}_N = (v_0, v_1, \cdots, v_{N-1})^T$ be arbitrary N-dimensional vectors, and T_N and R_N be $N \times N$ Toeplitz and circulant matrices, respectively. By definition, the i, j entry of T_N is t_{i-j} and the i, j entry of R_N is r_{i-j}, where $r_n = r_{n \bmod N}$. We will inter-

pret the matrix–vector products $T_N u_N$, $R_N u_N$, $T_N^{-1} v_N$ and $R_N^{-1} v_N$, from a convolutional point of view, since our approach to the design of Toeplitz preconditioners can be well motivated by this viewpoint.

First, consider $v_N = T_N u_N$. The element v_i, $0 \le i \le N-1$, can be written as

$$v_i = \sum_{j=0}^{N-1} t_{i-j} u_j. \tag{6}$$

More generally, (6) with any integers i and j defines a linear convolution

$$v = t * u \tag{7}$$

where

$$t = \cdots, 0, t_{-(N-1)}, \cdots, t_{-1}, t_0, t_1, \cdots, t_{N-1}, 0, \cdots, \text{ and}$$

$$u = \cdots, 0, u_0, u_1, \cdots, u_{N-1}, 0, \cdots.$$

Note that v, t, and u in (7) are infinite sequences of duration $3N-2$, $2N-1$, and N, respectively. In linear system theory, u and v are usually known as the input and output, and t the impulse response of the system [20]. Since the output v contains elements v_i of v_N, the Toeplitz matrix–vector product $v_N = T_N u_N$ is embedded in the linear convolution (7). For (7), we can define a linear deconvolution problem, namely, to determine the input u from the output v and the impulse response t.

Next, consider $v_N = R_N u_N$. The element v_i, $0 \le i \le N-1$, can be written as

$$v_i = \sum_{j=0}^{N-1} r_{i-j} u_j, \qquad i = 0, 1, \cdots, N-1. \tag{8}$$

Equation (8) with any integers i and j defines a circular convolution

$$\tilde{v} = \tilde{r} \otimes \tilde{u} \tag{9}$$

where the output $\tilde{v}$, input $\tilde{u}$ and impulse response $\tilde{r}$ are all N-periodic sequences with periods

$$v_N^T = (v_0, \cdots, v_{N-1}),$$

$$u_N^T = (u_0, \cdots, u_{N-1}), \quad \text{and} \quad (r_0, \cdots, r_{N-1}).$$

Hence, we can embed the circulant matrix–vector product $v_N = R_N u_N$ in the circular convolution (9). The circular deconvolution problem is to determine the input $\tilde{u}$ based on the output $\hat{v}$ and the impulse response $\tilde{r}$.

The circular convolution and deconvolution can be performed effectively by using FFT. That is, by applying the discrete Fourier transform, defined as

$$\hat{u}_k = \sum_{n=0}^{N-1} u_n e^{-(i2\pi kn/N)}$$

to periodic sequences $\tilde{v}$, $\tilde{r}$, and $\tilde{u}$ in (9), we obtain

$$\hat{v}_k = \hat{r}_k \hat{u}_k \quad \text{or} \quad \hat{u}_k = \hat{v}_k / \hat{r}_k \tag{10}$$

in the transform domain. Thus, the circular convolution (deconvolution) or the embedded $v_N = R_N u_N$ ($u_N = R_N^{-1} v_N$) can be obtained with $O(N \log N)$ operations.

It is also possible to compute the linear convolution (7) and the corresponding linear deconvolution with FFT. For example, we may view v, t, and u of (7) as if they were all $(3N-2)$-periodic sequences, and treat the linear convolution (deconvolution) problem as a $(3N-2)$-point circular convolution (deconvolution) problem. Since $v_N = T_N u_N$ can be embedded in (7) and since we know all nontrivial $2N-1$ and N values of t and u, we can compute v as well as v_N effectively. However, the computation of $u_N = T_N^{-1} v_N$ is not as easy. Since only N values (i.e., v_N) of the output v are given, we do not have sufficient information to perform the linear deconvolution (but sufficient for solving the Toeplitz system). Thus, the inverse Toeplitz matrix–vector product only partially characterizes a linear deconvolution problem.

In order to exploit the low computational complexity provided by FFT, we seek some circular deconvolution to approximate the partially characterized linear deconvolution problem. For example, we can cut the length of t_n's and use

$$r_N = \begin{cases} (t_{-(N-1)/2}, \cdots, t_{-1}, t_0, t_1, \cdots, t_{(N-1)/2}) \\ \quad \text{for odd} \quad N \\ (t_{-N/2}, \cdots, t_{-1}, t_0, t_1, \cdots, t_{N/2-1}) \\ \quad \text{for even} \quad N \end{cases} \tag{11}$$

to define a periodic sequence $\tilde{r}$ of period N. Although the N-point circular deconvolution of $\tilde{r}$ and $\tilde{v}$ does not embed the desired computation $T_N^{-1} v_N$, it can be viewed as its approximation, and used in the preconditioning step of the PCG method. This was originally suggested by Strang [22]. One shortcoming of Strang's idea is that half of the elements contained in T_N is lost. To use all elements of T_N, we may choose to extend v periodically with v_N as the basic unit, which will be detailed below.

B. Construction of Toeplitz Preconditioners

Let A be an $N \times N$ SPD Toeplitz matrix, and $T_{N,1}$ be an $N \times N$ symmetric Toeplitz matrix approximating A. For example, we can choose $T_{N,1} = A$ or $T_{N,1}$ which minimizes the difference $T_{N,1} - A$ with respect to a certain norm. We define a $2N \times 2N$ symmetric circulant matrix as

$$R_{2N} = \begin{bmatrix} T_{N,1} & T_{N,2} \\ T_{N,2} & T_{N,1} \end{bmatrix} \tag{12}$$

where

$$T_{N,1} = \begin{bmatrix} t_0 & t_1 & \cdot & t_{N-2} & t_{N-1} \\ t_1 & t_0 & t_1 & \cdot & t_{N-2} \\ \cdot & t_1 & t_0 & \cdot & \cdot \\ T_{N-2} & \cdot & \cdot & \cdot & t_1 \\ t_{N-1} & t_{N-2} & \cdot & t_1 & t_0 \end{bmatrix} \tag{13}$$

and where $T_{N,2}$ is determined by elements of $T_{N,1}$

$$T_{N,2} = \begin{bmatrix} c & t_{N-1} & \cdot & t_2 & t_1 \\ t_{N-1} & c & t_{N-1} & \cdot & t_2 \\ \cdot & t_{N-1} & c & \cdot & \cdot \\ t_2 & \cdot & \cdot & \cdot & t_{N-1} \\ t_1 & t_2 & \cdot & t_{N-1} & c \end{bmatrix} \quad (14)$$

with a constant c.

Now, let us consider the following augmented system:

$$\begin{bmatrix} T_{N,1} & T_{N,2} \\ T_{N,2} & T_{N,1} \end{bmatrix} \begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{x} \end{bmatrix} = \begin{bmatrix} \boldsymbol{b} \\ \boldsymbol{b} \end{bmatrix}. \quad (15)$$

From the discussion in Section III-A, we know that (15) can be embedded by a circular convolution between two $2N$-periodic sequences, whose periods are

$$t_0, t_1, \cdots t_{N-2}, t_{N-1}, c, t_{N-1}, t_{N-2}, \cdots, t_1 \quad (16)$$

and

$$x_1, x_2, \cdots x_{N-1}, x_N, x_1, x_2, \cdots, x_{N-1}, x_N. \quad (17)$$

The output sequence is also $2N$-periodic, whose period is

$$b_1, b_2, \cdots, b_{N-1}, b_N, b_1, b_2, \cdots, b_{N-1}, b_N. \quad (18)$$

The solution of (15) for $\boldsymbol{x}$ corresponds to a circular deconvolution problem and can be computed via FFT with $O(N \log N)$ operations. Since the system (15) is equivalent to

$$(T_{N,1} + T_{N,2})\boldsymbol{x} = \boldsymbol{b}$$

we can compute $(T_{N,1} + T_{N,2})^{-1}\boldsymbol{b}$ efficiently and use

$$P_1 = T_{N,1} + T_{N,2} \quad (19)$$

as a preconditioner for A.

Various preconditioners can be constructed in a similar way by assuming different periodicities for $\boldsymbol{x}$ and $\boldsymbol{b}$, such as negative periodicity, even periodicity, and odd periodicity. The corresponding augmented systems and preconditioners can be written as follows:

$$\begin{bmatrix} T_{N,1} & T_{N,2} \\ T_{N,2} & T_{N,1} \end{bmatrix} \begin{bmatrix} \boldsymbol{x} \\ -\boldsymbol{x} \end{bmatrix} = \begin{bmatrix} \boldsymbol{b} \\ -\boldsymbol{b} \end{bmatrix} \quad \text{and } P_2 = T_{N,1} - T_{N,2} \quad (20)$$

$$\begin{bmatrix} T_{N,1} & T_{N,2} \\ T_{N,2} & T_{N,1} \end{bmatrix} \begin{bmatrix} \boldsymbol{x} \\ J\boldsymbol{x} \end{bmatrix} = \begin{bmatrix} \boldsymbol{b} \\ J\boldsymbol{b} \end{bmatrix} \quad \text{and } P_3 = T_{N,1} + JT_{N,2} \quad (21)$$

$$\begin{bmatrix} T_{N,1} & T_{N,2} \\ T_{N,2} & T_{N,1} \end{bmatrix} \begin{bmatrix} \boldsymbol{x} \\ -J\boldsymbol{x} \end{bmatrix} = \begin{bmatrix} \boldsymbol{b} \\ -J\boldsymbol{b} \end{bmatrix} \quad \text{and } P_4 = T_{N,1} - JT_{N,2} \quad (22)$$

where J is the $N \times N$ symmetric elementary matrix which has, by definition, ones along the secondary diagonal and zeros elsewhere (equivalently, $J_{i,j} = 1$ if $i + j = N + 1$ and $J_{ij} = 0$ if $i + j \neq N + 1$).

To choose the appropriate constant c, several factors should be considered. First, we know from the above derivation that if λ is an eigenvalue of preconditioner P_i, $i = 1, 2, 3, 4$, it is also an eigenvalue of the matrix R_{2N}. To guarantee the positive definiteness of P_i, we require that

$$\sum_{n=-(N-1)}^{N-1} t_n e^{-i\pi kn/N} + (-1)^k c > 0. \quad (23)$$

Second, since we want the norm of $P_i - A$ to be as small as possible, c should be a small number. For sufficiently large N, we can adopt the simple rule of thumb, namely, if the behavior of the sequence t_n is known, $c = t_N$. Otherwise, $c = 0$.

Since preconditioners P_i, $i = 1, 2, 3, 4$, correspond to $2N$-circulant systems, they can be inverted via fast transform algorithms with $O(N \log N)$ operations. The implementation of P_i^{-1} will be detailed in Section III-D. The subscript N of matrices is omitted hereinafter whenever there is no confusion.

C. Examples of Toeplitz Preconditioners

We describe various preconditioners for the Toeplitz matrix

$$A = \begin{bmatrix} 32 & 16 & 8 & 4 & 2 \\ 16 & 32 & 16 & 8 & 4 \\ 8 & 16 & 32 & 16 & 8 \\ 4 & 8 & 16 & 32 & 16 \\ 2 & 4 & 8 & 16 & 32 \end{bmatrix}$$

to illustrate the construction procedure given in Section III-B.

Example 1: (Strang's preconditioner) [22]

By choosing T_1 to be the central half-band of A and $c = 0$, we obtain

$$T_1 = \begin{bmatrix} 32 & 16 & 8 & 0 & 0 \\ 16 & 32 & 16 & 8 & 0 \\ 8 & 16 & 32 & 16 & 8 \\ 0 & 8 & 16 & 32 & 16 \\ 0 & 0 & 8 & 16 & 32 \end{bmatrix}$$

$$T_2 = \begin{bmatrix} 0 & 0 & 0 & 8 & 16 \\ 0 & 0 & 0 & 0 & 8 \\ 0 & 0 & 0 & 0 & 0 \\ 8 & 0 & 0 & 0 & 0 \\ 16 & 8 & 0 & 0 & 0 \end{bmatrix}.$$

The resulting preconditioner

$$P_1 = T_1 + T_2 = \begin{bmatrix} 32 & 16 & 8 & 8 & 16 \\ 16 & 32 & 16 & 8 & 8 \\ 8 & 16 & 32 & 16 & 8 \\ 8 & 8 & 16 & 32 & 16 \\ 16 & 8 & 8 & 16 & 32 \end{bmatrix}$$

is the same as Strang's preconditioner S.

Example 2: (Chan's preconditioner) [8]: Chan's preconditioner C is the circulant matrix which minimizes the Frobenius norm of $A - R$ over all circulant matrices R. It turns out that the elements of C can be computed as

$$c_i = \frac{1}{N}(i \times a_{(N-i)} + (N - i) \times a_i)$$

$$i = 0, 1, 2, \cdots, N - 1.$$

By choosing

$$T_1 = \begin{bmatrix} 32 & 13.2 & 6.4 & 0 & 0 \\ 13.2 & 32 & 13.2 & 6.4 & 0 \\ 6.4 & 13.2 & 32 & 13.2 & 6.4 \\ 0 & 6.4 & 13.2 & 32 & 13.2 \\ 0 & 0 & 6.4 & 13.2 & 32 \end{bmatrix}$$

$$T_2 = \begin{bmatrix} 0 & 0 & 0 & 6.4 & 13.2 \\ 0 & 0 & 0 & 0 & 6.4 \\ 0 & 0 & 0 & 0 & 0 \\ 6.4 & 0 & 0 & 0 & 0 \\ 13.2 & 6.4 & 0 & 0 & 0 \end{bmatrix}.$$

The resulting preconditioner

$$P_1 = T_1 + T_2 = \begin{bmatrix} 32 & 13.2 & 6.4 & 6.4 & 13.2 \\ 13.2 & 32 & 13.2 & 6.4 & 6.4 \\ 6.4 & 13.2 & 32 & 13.2 & 6.4 \\ 6.4 & 6.4 & 13.2 & 32 & 13.2 \\ 13.2 & 6.4 & 6.4 & 13.2 & 32 \end{bmatrix}$$

is the same as Chan's preconditioner C. It is straightforward to generalize the above examples to prove that, for any give SPD Toeplitz matrix A, Strang's preconditioner S and Chan's preconditioner C are special cases of P_1 with appropriately chosen $T_{N,1}$ and $c = 0$.

Example 3: (Preconditioners K_i): We use (19)–(22) to construct preconditioners. Although there exist many choices to select T_1 for the design of preconditioners P_i, the choice $T_1 = A$ seems natural. For this choice, all elements of A are used in a straightforward way, and we call the resulting preconditioners K_i. The corresponding T_2 becomes

$$T_2 = \begin{bmatrix} 1 & 2 & 4 & 8 & 16 \\ 2 & 1 & 2 & 4 & 8 \\ 4 & 2 & 1 & 2 & 4 \\ 8 & 4 & 2 & 1 & 2 \\ 16 & 8 & 4 & 2 & 1 \end{bmatrix}$$

where $c = 1$. From (19)–(22), we have

$$K_1 = \begin{bmatrix} 33 & 18 & 12 & 12 & 18 \\ 18 & 33 & 18 & 12 & 12 \\ 12 & 18 & 33 & 18 & 12 \\ 12 & 12 & 18 & 33 & 18 \\ 18 & 12 & 12 & 18 & 33 \end{bmatrix}$$

$$K_2 = \begin{bmatrix} 31 & 14 & 4 & -4 & -14 \\ 14 & 31 & 14 & 4 & -4 \\ 4 & 14 & 31 & 14 & 4 \\ -4 & 4 & 14 & 31 & 14 \\ -14 & -4 & 1 & 14 & 31 \end{bmatrix}$$

$$K_3 = \begin{bmatrix} 48 & 24 & 12 & 6 & 3 \\ 24 & 36 & 18 & 9 & 6 \\ 12 & 18 & 33 & 18 & 12 \\ 6 & 9 & 18 & 36 & 24 \\ 3 & 6 & 12 & 24 & 48 \end{bmatrix}$$

$$K_4 = \begin{bmatrix} 16 & 8 & 4 & 2 & 1 \\ 8 & 28 & 14 & 7 & 2 \\ 4 & 14 & 31 & 14 & 4 \\ 2 & 7 & 14 & 28 & 8 \\ 1 & 2 & 4 & 8 & 16 \end{bmatrix}.$$

We want to point out that a preconditioner which is very similar to K_1 except $c = 0$ was described in [5]. Note also that preconditioners S, C, and K_1, which are special cases of P_1, are all circulant. If B is a symmetric Toeplitz matrix with the first row

$$(a_0, a_1, \cdots, a_k, -a_k, \cdots, -a_1)$$

for odd N and $k = (N - 1)/2$

or

$$(a_0, a_1, \cdots, a_{k-1}, 0, -a_{k-1}, \cdots, -a_1)$$

for even N and $k = N/2$

we say that B is *skew circulant* [9]. It is clear that K_2 is skew-circulant. In fact, one can verify that the circulant

and skew-circulant properties hold for general P_1 and P_2 given by (19) and (20), respectively. However, new preconditioners K_3 and K_4 are neither circulant nor Toeplitz.

D. Comparison of Computational Cost

We compare the computational cost for the preconditioning step $P^{-1}\boldsymbol{r}$ with different preconditioners at each PCG iteration as follows. Preconditioners C, S, and K_1 are all $N \times N$ circulant matrices and the preconditioning can be done via N-point FFT with approximately $1.5N \log N$ real multiplications and $4.5N \log N$ real additions when $N = 2^l$ [21]. Preconditioner K_2 is skew circulant and can be transformed into a circulant matrix through $D^H K_2 D$, where D is a diagonal matrix [9]. Consequently, the implementation of $K_2^{-1}\boldsymbol{r}$ is almost as easy as that of $K_1^{-1}\boldsymbol{r}$. Although preconditioners K_3 and K_4 are noncirculant, $K_3^{-1}\boldsymbol{r}$ and $K_4^{-1}\boldsymbol{r}$ can be performed via N-point fast cosine and sine transforms, respectively. The operation counts for N-point fast cosine (or sine) transform are approximately equal to that of N-point FFT in both the order and the proportional constant [19], [26]. Therefore, they are as competitive as C, S, and K_i, $i = 1, 2$.

IV. Spectral Properties of the Preconditioned Toeplitz Matrix

We let $T_1 = A$ and denote the corresponding T_2 with $c = a_N$ by ΔA so that preconditioners K_i can be expressed as

$$K_1 = A + \Delta A, \qquad K_2 = A - \Delta A$$

$$K_3 = A + J\Delta A, \qquad K_4 = A - J\Delta A. \tag{24}$$

To study the spectral properties of $K_i^{-1}A$, we view the matrix A to be a member in a sequence of $m \times m$ symmetric Toeplitz matrices $\{A_m\}_{m=1}^{\infty}$, where the first row of A_m are elements from the infinite sequence $\{a_n\}_{n=0}^{\infty}$ up to element a_{m-1} and $\{a_n\}_{n=0}^{\infty}$ is known as the generating sequence of A_m. We assume that the sequence a_n satisfies the following two conditions:

$$f(\theta) = \sum_{-\infty}^{\infty} a_n e^{-in\theta} \geq \delta > 0, \ \forall\theta \tag{25}$$

$$\sum_{-\infty}^{\infty} |a_n| < \infty \tag{26}$$

and the resulting Toeplitz matrices are said to be generated by a positive function in the Wiener class [7]. Since $f(\theta)$ describes the asymptotic eigenvalue distribution of A_m, the above conditions imply that eigenvalues of A_m are bounded and uniformly positive asymptotically. With conditions (25) and (26), we will establish three main results for the spectra of $K_i^{-1}A$. 1) There exists a simple relationship between eigenvalues of $K_i^{-1}A$, $i = 1, 2, 3, 4$. 2) The eigenvalues are all real and positive for sufficiently large N. 3) The eigenvalues of $K_i^{-1}A$ are clustered around 1 except a finite number of outliers.

To relate the eigenvalues of $K_i^{-1}A$, we introduce some definitions and related concepts. An N-dimensional vector $\boldsymbol{v}$ is called symmetric if $J\boldsymbol{v} = \boldsymbol{v}$ or skew-symmetric if $J\boldsymbol{v} = -\boldsymbol{v}$, where J is the symmetric elementary matrix. An $N \times N$ matrix Q is called doubly symmetric (or symmetric centrosymmetric) if

$$Q = Q^T \quad \text{and} \quad (JQ)^T = (JQ). \tag{27}$$

Note that if Q is doubly symmetric, matrices Q and J commute.

When A and ΔA are symmetric Toeplitz matrices, A, ΔA, and $J\Delta A$ are all doubly symmetric. Since any linear combination of doubly symmetric matrices results in a doubly symmetric matrix, preconditioners K_i given by (24) are doubly symmetric. The eigenvectors of $K_i^{-1}A$ can be characterized by the following lemma, which will be needed in proving Theorem 1.

Lemma 1: If matrices A and B are both doubly symmetric, there exists a set of $\lceil N/2 \rceil$ symmetric eigenvectors and $\lfloor N/2 \rfloor$ skew-symmetric eigenvectors for $B^{-1}A$.

Proof: See Appendix A. □

Let us rewrite the spectra of $K_i^{-1}A$, $1 \leq i \leq 4$, as

$$[\lambda(K_i^{-1}A)]^{-1} = \lambda(A^{-1}(A + K_i - A))$$

$$= \lambda(I + A^{-1}(K_i - A)) = 1 + \lambda(A^{-1}(K_i - A)). \tag{28}$$

The following theorem characterizes the relation between the eigenvalues of $A^{-1}(K_i - A)$.

Theorem 1: Let Q_i be the set of the absolute values of the eigenvalues of $A^{-1}(K_i - A)$, i.e.,

$$Q_i = \{|\lambda|: A^{-1}(K_i - A)\boldsymbol{x} = \lambda\boldsymbol{x}\}, \qquad i = 1, 2, 3, 4.$$

Then, $Q_1 = Q_2 = Q_3 = Q_4$.

Proof: See Appendix B. □

The above theorem can be stated alternatively as follows. For an arbitrary eigenvalue λ of $A^{-1}(K_i - A)$, there exists an eigenvalue of $A^{-1}(K_j - A)$, $j \neq i$, with magnitude $|\lambda|$. Note that the spectra of $A^{-1}(K_i - A)$ clustered around zero is equivalent to those of $K_i^{-1}A$ clustered around unity. Since the spectra of $A^{-1}(K_i - A)$ are clustered in a very similar pattern, so are those of $K_i^{-1}A$. This theorem implies that the PCG method with preconditioners K_i, $i = 1, 2, 3, 4$, should converge in a similar rate.

When preconditioners K_i are positive definite, the preconditioned matrices $K_i^{-1/2}AK_i^{-1/2}$ are symmetric positive definite. Therefore, the PCG method can be conveniently applied and it converges to the unique solution. The positive definiteness of K_i is given in the following theorem. Its proof is similar to that of Theorem 1 in [7].

Theorem 2: Preconditioners K_i, $i = 1, 2, 3, 4$, for symmetric positive definite Toeplitz matrices with the generating sequence satisfying (25) and (26) are uniformly positive and bounded for sufficiently large N.

Proof: See Appendix C. □

In the next theorem, we describe the clustering feature of the spectra of $A^{-1}(K_i - A)$ and, hence, that of $K_i^{-1}A$. The proof is similar to that given by Chan in [5].

Theorem 3: Let A be the $N \times N$ matrix of a sequence of symmetric positive definite Toeplitz matrices A_m with

the generating sequence satisfying (25) and (26). The eigenvalues of matrices $A^{-1}(K_i - A)$ are clustered between $(-\epsilon, +\epsilon)$ except a finite number of outliers for sufficiently large $N(\epsilon)$.

Proof: See Appendix D. □

V. Numerical Results

We compare Strang's preconditioner S, Chan's preconditioner C, and our preconditioners K_i for different numerical test problems in this section. We will show the clustering properties of the spectra of $P^{-1}A$ with $P = C$, S, K_i as well as the convergence history of the PCG method.

For a sequence of Toeplitz matrices A_m generated by sequence a_n, we can define their generating function as the Z transform of a_n

$$A(z) = \sum_{n=-\infty}^{\infty} a_n z^{-n}.$$

If A is symmetric, $A(z)$ can be decomposed into

$$A(z) = A_+(z) + A_+(z^{-1})$$

where

$$A_+(z) = \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n z^{-n}$$

is the Z transform of a causal sequence. Thus, $A(z)$ is completely characterized by $A_+(z)$. If

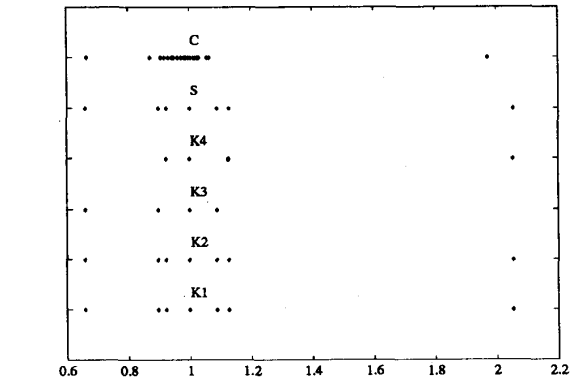
$$A_+(z) = \frac{\sum_{n=0}^{p} c_n z^{-n}}{\sum_{n=0}^{q} d_n z^{-n}}, \qquad \text{where } c_p d_0 d_q \neq 0$$

we call $A_+(z)$ a rational function of order (p, q). In the digital signal processing context, Toeplitz matrices with rational generating functions are particularly of interest, since the covariance matrices of stationary autoregressive (AR), moving average (MA), and ARMA random processes can be expressed in this form.

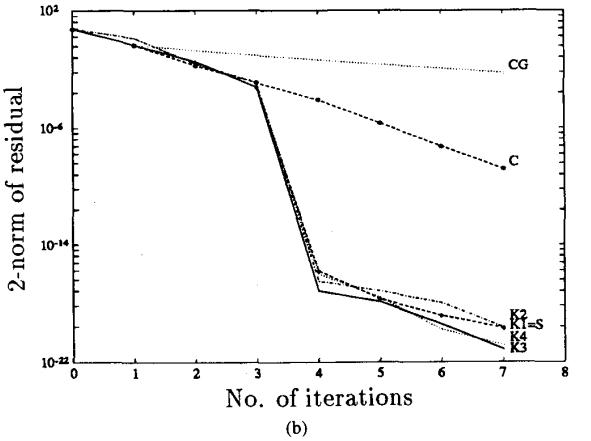
We choose $A_+(z)$ to be rational for problems 1–5 and nonrational for problems 6–8. All numerical experiments are performed with respect to 32×32 Toeplitz matrices A with right-hand-side $\boldsymbol{b} = (1, \cdots, 1)^T$ and initial condition $\boldsymbol{x}_0 = 0$. We can roughly classify the eigenvalues of $P^{-1}A$ into two categories: the outliers and the clustered eigenvalues between $(1 - \epsilon, 1 + \epsilon)$ for general $A(z)$. However, a more precise distinction can be made for rational $A(z)$. That is, the clustered eigenvalues are those contained in the interval $(1 - \epsilon, 1 + \epsilon)$, where the clustering radius ϵ converges to zero when N goes to infinity, and the outliers are the eigenvalues not converging to one.

Problem 1: $a_n = 0.5^n$ for $n \leq 3$ and $a_n = 0$ for $n > 3$ (banded Toeplitz matrix with $p = 3$, $q = 0$).

For a banded Toeplitz matrix with bandwidth $p \leq \lfloor N/2 \rfloor$, K_1 and S are the same. Since K_i's and A have $N - 2p$ identical rows, $K_i^{-1}(K_i - A) = I - K_i^{-1}A$ has a null space of dimension $N - 2p$. This implies that $K_i^{-1}A$ has the eigenvalue one with multiplicity $N - 2p$, which correspond to the clustered eigenvalues defined above. The other $2p$ eigenvalues are outliers. The spectra of $P^{-1}A$ are plotted in Fig. 1(a). For $K_i^{-1}A$, $i = 1, 2, 3, 4$, there are 6 ($p = 3$) outliers and $N - 6$ eigenvalues repeated at one. For $K_i^{-1}A$, $i = 3, 4$, each pair of outliers are closely located so that only three distinct dots appear in the figure. The eigenvalues of $C^{-1}A$ are not clustered as well for this problem.



(a)



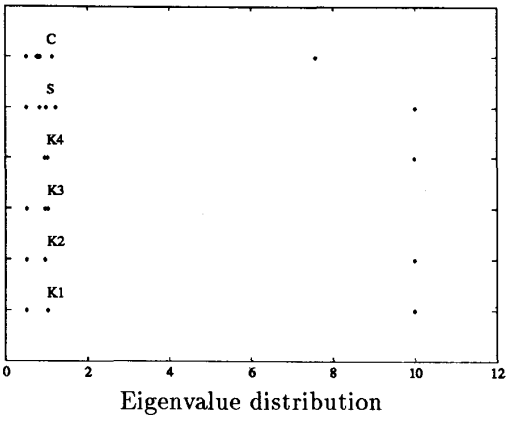
(b)

Fig. 1. (a) Eigenvalue distribution of $P^{-1}A$ and (b) convergence history for problem 1.

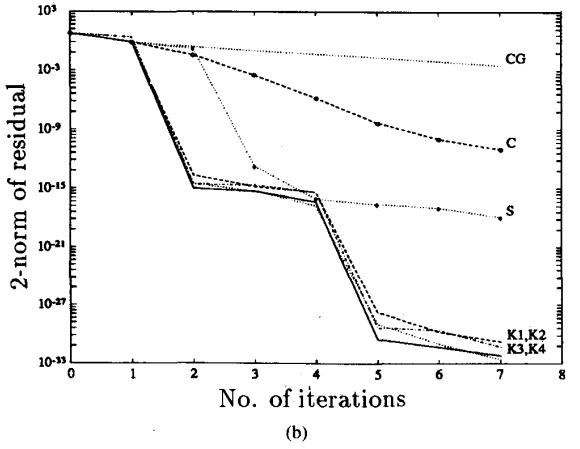
According to the discussion in Section II, the PCG method with K_i should converge in at most $2p + 1$ iterations with exact arithmetic since $K_i^{-1}A$ has $2p + 1$ distinct eigenvalues. However, it is worthwhile to point out that (4) only provides an upper bound estimate of the convergence rate. From our experience, this estimate seems pessimistic. We observe the PCG method with K_i converges in $p + 1$ iterations for banded Toeplitz matrices with different values of p. In Fig. 1(b), we plot the 2-norm of the residual $\boldsymbol{b} - A\boldsymbol{x}$ as a function of the number of PCG iterations. It is clear from the figure that the PCG method converges in 4 ($= p + 1$) iterations for all K_i's.

TABLE I
EIGENVALUES OF $K_i^{-1}A$

	$K_1^{-1}A$	$K_2^{-1}A$	$K_3^{-1}A$	$K_4^{-1}A$
λ_1	$(1+t)^{-1}$	$(1+t)^{-1}$	$(1+t)^{-1}$	$(1-t)^{-1}$
λ_2	$(1-t)^{-1}$	$(1-t)^{-1}$	$(1+t^N)^{-1}$	$(1+t^N)^{-1}$
λ_3	$(1-t^N)^{-1}$	$(1+t^N)^{-1}$	$(1-t^N)^{-1}$	$(1-t^N)^{-1}$



(a)

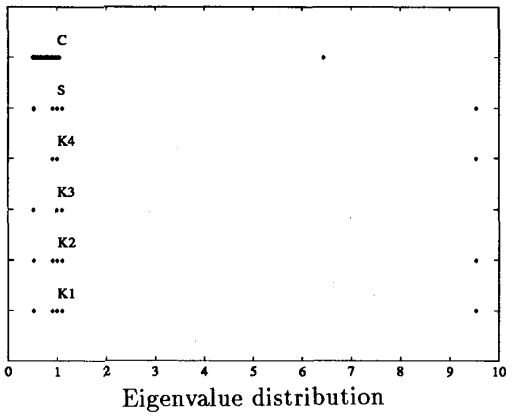


(b)

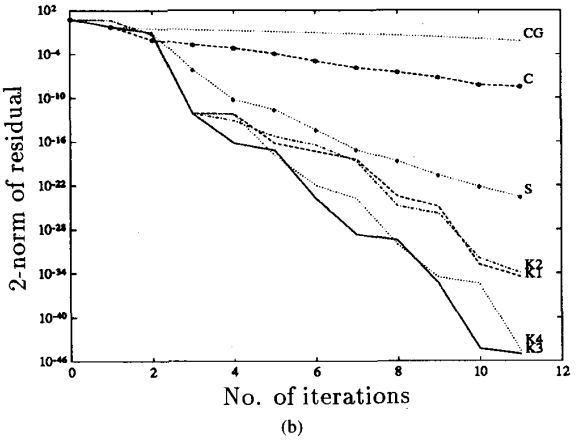
Fig. 2. (a) Eigenvalue distribution of $P^{-1}A$ and (b) convergence history for problem 2.

Problem 2: $a_n = t^n$, ($q = 1$, a single pole at t).

For this generating sequence, it has been observed by Strang [24], that the spectrum of $S^{-1}A$ has two outliers at $(1+t)^{-1}$ and $(1-t)^{-1}$, two eigenvalues repeated at 1, and other eigenvalues at $(1+t^{N/2})^{-1}$ and $(1-t^{N/2})^{-1}$ with multiplicity $(N-4)/2$ when N is even. Nevertheless, the same regularity does not hold for odd N. For the same generating sequence, the spectra of $K_i^{-1}A$, $i = 1, 2, 3, 4$, have only three distinct eigenvalues for both even and odd N. We summarize these values in Table I and plot the spectra of $P^{-1}A$ with $t = 0.9$ in Fig. 2(a). For preconditioners K_i, the two outliers are located at $(1+t)^{-1}$ or $(1-t)^{-1}$ and other $N-2$ clustered eigenvalues



(a)



(b)

Fig. 3. (a) Eigenvalue distribution of $P^{-1}A$ and (b) convergence history for problem 3.

are repeated at $(1-t^N)^{-1}$ or $(1+t^N)^{-1}$. The outliers of $K_i^{-1}A$, $i = 3, 4$ are repeated with multiplicity 2.

The convergence history of the PCG method with $t = 0.9$ is given in Fig. 2(b). Since $K_i^{-1}A$ has only 3 distinct eigenvalues, the PCG method converges in at most 3 iterations independent of N. From this figure, we see that the PCG method converges with 2 (or 3) iterations with preconditioners K_i (or S).

Problem 3: $a_n = (n+1)t^n$, ($q = 2$, a double pole at t).

We plot the eigenvalues of $P^{-1}A$ with $t = 0.4$ in Fig. 3(a). The spectra of $P^{-1}A$ consists of 4 outliers and $N-4$ clustered eigenvalues between $(1-\epsilon, 1+\epsilon)$. Similar to problem 1, each pair of outliers of $K_i^{-1}A$, $i = 3, 4$, are closely located so that only two distinct dots appear. The corresponding convergence history is plotted in Fig. 3(b). We see that preconditioners K_i converge faster in comparison with S and C. It takes approximately 5 (or 7) iterations for preconditioners K_i (or S) to converge. Note that K_3 and K_4 behave better than K_1 and K_2, when the number of iteration becomes large.

TABLE II

	α	max (p, q)	$\epsilon(S^{-1}A)$	$\lvert a_{N/2}/a_0\rvert$	$\epsilon(K_i^{-1}A)$	$\lvert a_N/a_0\rvert$
Problem 1	6	3	0	0	0	0
Problem 2	2	1	$\lvert t^{N/2}\rvert + O(t^N)$	$\lvert t^{N/2}\rvert$	$\lvert t^N\rvert + O(t^{2N})$	$\lvert t^N\rvert$

TABLE III

t	$\epsilon(S^{-1}A)$	$\lvert a_{16}/a_0\rvert$	$\epsilon(K_i^{-1}A)$	$\lvert a_{32}/a_0\rvert$
0.3	2.0×10^{-7}	7.3×10^{-8}	3.6×10^{-15}	6.1×10^{-16}
0.4	1.3×10^{-4}	7.3×10^{-6}	1.2×10^{-10}	6.1×10^{-12}
0.5	4.6×10^{-3}	2.6×10^{-4}	4.4×10^{-7}	7.7×10^{-9}

When $A_+(z)$ is a rational function of order (p, q), we observe two important regularities for the spectra of $K_i^{-1}A$ and $S^{-1}A$:

R1) The number α of outliers is equal to $2 \times \max(p, q)$.

R2) The order of ϵ is proportional to $|a_N/a_0|$ (or $|a_{N/2}/a_0|$) for K_i's (or S).

The values of α, max (p, q), $\epsilon(S^{-1}A)$, $|a_{N/2}/a_0|$, $\epsilon(K_i^{-1}A)$ and $|a_N/a_0|$ for problems 1 and 2 are listed in Table II. We can clearly see that they are consistent with the above two rules.

For problem 3, $\alpha = 4$ and max $(p, q) = q = 2$ and rule R1 holds. We list $\epsilon(S^{-1}A)$, $|a_{16}/a_0|$, $\epsilon(K_i^{-1}A)$ and $|a_{32}/a_0|$ for $t = 0.3, 0.4, 0.5$ in Table III to verify rule R2.

Rule R2 explains why our preconditioners K_i behave better than Strang's preconditioner S. From R2, we have

$$\frac{\epsilon(K_i^{-1}A)}{\epsilon(S^{-1}A)} = O\left(\left|\frac{a_N}{a_{N/2}}\right|\right).$$

Recall that the construction of preconditioner S uses only half the elements of A (up to the element $a_{N/2}$) whereas the construction of K_i uses all elements in A (up to the element a_N). Thus, to use more elements of A by our approach does improve the clustering radius ϵ by a factor of $O(|a_N/a_{N/2}|)$.

Based on rule R2, the clustering radius ϵ converges to 0 as N goes to infinity for rational generating sequence a_n in the Wiener class. There are at most $\alpha + 1$ distinct eigenvalues asymptotically. Therefore, the PCG method converges in a finite number of iterations for large N, and the total computational complexity is $O(N \log N)$.

Problem 4: $a_n = (n + 1)t_0^n + t_1^n$, ($q = 3$, a double pole at t_0 and a single pole at t_1).

The spectra of $P^{-1}A$ with $t_0 = 0.3$ and $t_1 = 0.8$ are plotted in Fig. 4(a). There are 6 (max $(p, q) = 3$) outliers for $K_i^{-1}A$ and $S^{-1}A$. The outliers of $K_i^{-1}A$, $i = 3, 4$, are clustered into three distinct dots. The clustering radii ϵ, a_{16}/a_0 and a_{32}/a_0 with different t_0 and t_1 are given in Table IV to verify rule R2:

The convergence history of the PCG method with $t_0 = 0.3$ and $t_1 = 0.8$ is given in Fig. 4(b). Preconditioners K_i behave better than C and S. It takes approximately 7 (or 10) iterations for K_i or (or S) to converge.

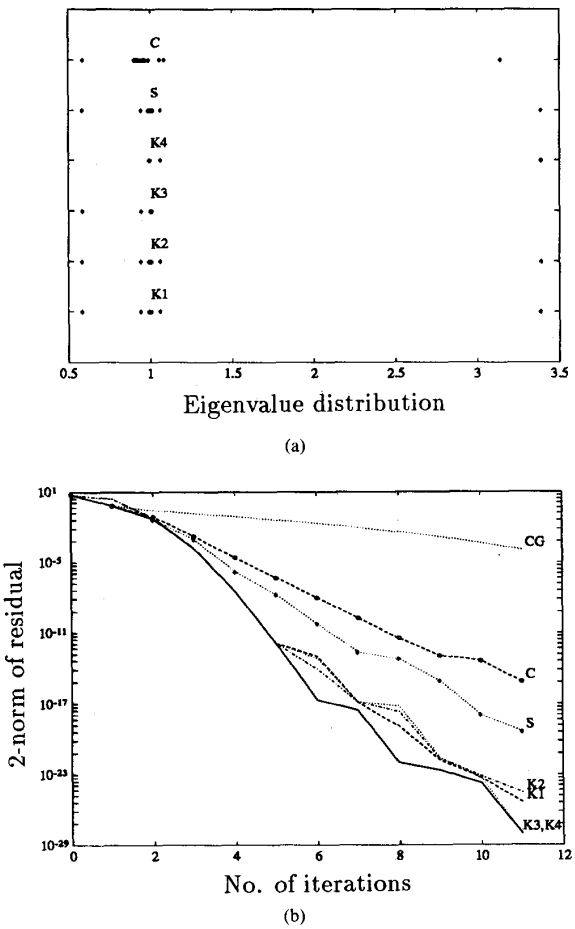


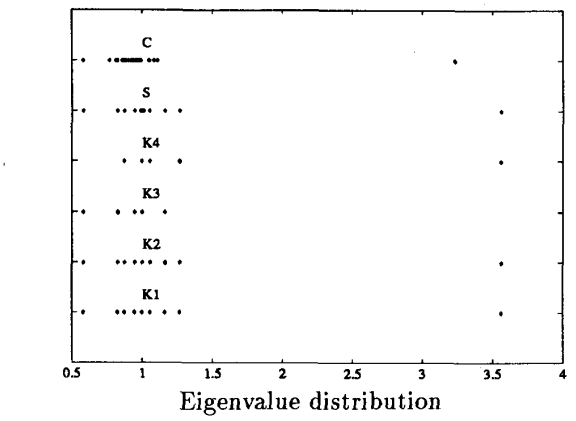
Fig. 4. (a) Eigenvalue distribution of $P^{-1}A$ and (b) convergence history for problem 4.

TABLE IV

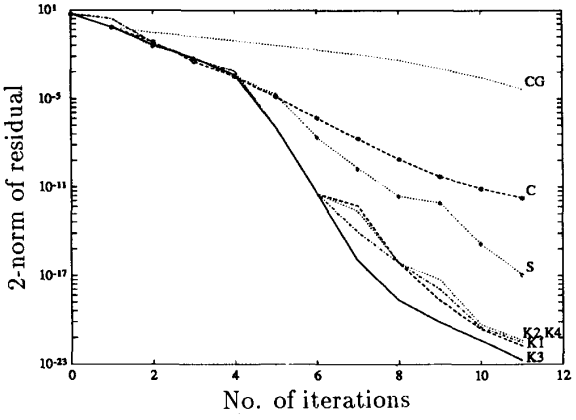
t_0	t_1	$\epsilon(S^{-1}A)$	a_{16}/a_0	$\epsilon(K_i^{-1}A)$	a_{32}/a_0
0.3	0.5	9.9×10^{-6}	7.7×10^{-6}	1.5×10^{-10}	1.2×10^{-10}
0.5	0.3	1.9×10^{-4}	1.3×10^{-4}	5.8×10^{-9}	3.8×10^{-9}
0.3	0.8	1.1×10^{-2}	1.4×10^{-2}	5.2×10^{-4}	4.0×10^{-4}

Problem 5: $a_n = t_0^n + t_1^n$ for $n \leq 3$ and $a_n = t_0^n$ for $n > 3$ ($p = 4$, $q = 1$).

In this example, the rational function $A_+(z)$ has the order (4, 1). The spectra of $P^{-1}A$ with $t_0 = 0.8$ and $t_1 =$



(a)



(b)

Fig. 5. (a) Eigenvalue distribution of $P^{-1}A$ and (b) convergence history for problem 5.

0.6 are plotted in Fig. 5(a). There are 8 (max $(p, q) = 4$) outliers for K_i and S. The outliers of $K_i^{-1}A$, $i = 3, 4$ are clustered into 4 distinguishable pairs. Rule R2 also holds for this problem. To avoid unnecessary repetition, we do not give a table to illustrate it.

The convergence history of the PCG method with $t_0 = 0.8$ and $t_1 = 0.6$ is plotted in Fig. 5(b). It takes approximately 8 (or 11) iterations for K_i (or S) to converge.

As discussed in problem 1, the PCG method converges in at most $2p + 1$ iterations (or $p + 1$ empirically) for p-banded Toeplitz matrices with $O(pN \log N)$ operations. It is worthwhile to point out that there exist direct methods which solve the system with $O(pN)$ operations [11]. Additionally, if $A_+(z)$ is or order (p, q), $q > 0$, Dickinson proposed a method to transform $\boldsymbol{Ax} = \boldsymbol{b}$ into an equivalent symmetric banded system $\tilde{\boldsymbol{A}}\tilde{\boldsymbol{x}} = \tilde{\boldsymbol{b}}$ with upper bandwidth max (p, q), whose solution can be obtained with max $(p, q) \times O(N)$ operations [12]. However, this transformation requires the knowledge of the exact form of $A(z)$.

The PCG method has three advantages in comparison with Dickinson's method. First, to implement the PCG algorithm, we only need a finite segment of the generating sequence a_n, $n = 0, 1, \cdots, N - 1$, rather than the precise formula of $A(z)$. Second, the PCG method can be easily parallelized due to the parallelism provided by FFT, and it is possible to reduce the time complexity to $O(\log N)$. In contrast, Dickinson's method is a sequential algorithm, and the time complexity can only be reduced to $O(N)$. Third, the PCG method is more widely applicable. For example, it can also be applied to Toeplitz matrices with nonrational generating functions.

TABLE V

a_i	C	S	K_i
$(n + 1)^{-2}$	8	7	6
$\cos(n\pi)/(n + 1)$	8	9	8
$(\log(n + 2))^{-1}$	8	10	9

Numerical results for Toeplitz matrices with nonrational generating functions are presented below. We consider 3 test problems, i.e.,

Problem 6: $a_n = (n + 1)^{-2}$.

Problem 7: $a_n = \cos(n\pi)/(n + 1)$.

Problem 8: $a_n = (\log(n + 2))^{-1}$.

Note that $|a_n|$ in problems 6–8 decay more slowly than $|a_n|$ in problems 1–5 asymptotically. The numbers of iterations required to achieve $\|\boldsymbol{b} - \boldsymbol{Ax}\|_2 \leq 10^{-15}$ are summarized in Table V for problems 6–8. Since all K_i's give the same performance, they are not distinguished. It turns out that all preconditioners have similar performances.

In order to understand their asymptotic behaviors, we consider a typical case $P = K_1$ and perform experiments for problems with sizes 32, 64, and 128. We plot the spectra of $K_i^{-1}A$ and the corresponding convergence history for problems 6–8 in Figs. 6(a) and (b). As seen in the figures, the change of the spectra and the convergence rates is not sensitive to the size of the problem. We conclude that the PCG method converges in a finite number of iterations independent of N for problems 6–8 and the total computational complexity is $O(N \log N)$.

Although the complexity of the PCG method is lower than that of fast or superfast direct methods for problems 6–8, it is worthwhile to point out that other factors have to be taken into account in comparing different methods, such as the constant in $O(N \log N)$, the necessity of $N = 2^l$ for using FFT efficiently, and the convergence rate of the PCG method for any matrix A. Besides, the Levinson algorithm with complexity $O(N^2)$ not only solves the Yule–Walker equations, which is a special case of $\boldsymbol{Ax} = \boldsymbol{b}$, but also gives A^{-1}.

VI. Conclusions and Extensions

In this paper, we have presented a systematic approach to the design of Toeplitz preconditioners by approximating a partially characterized linear deconvolution problem (the inverse Toeplitz-vector product) with some circular deconvolution problems. In particular, we show the design of four new preconditioners K_i, $i = 1, 2, 3, 4$, and

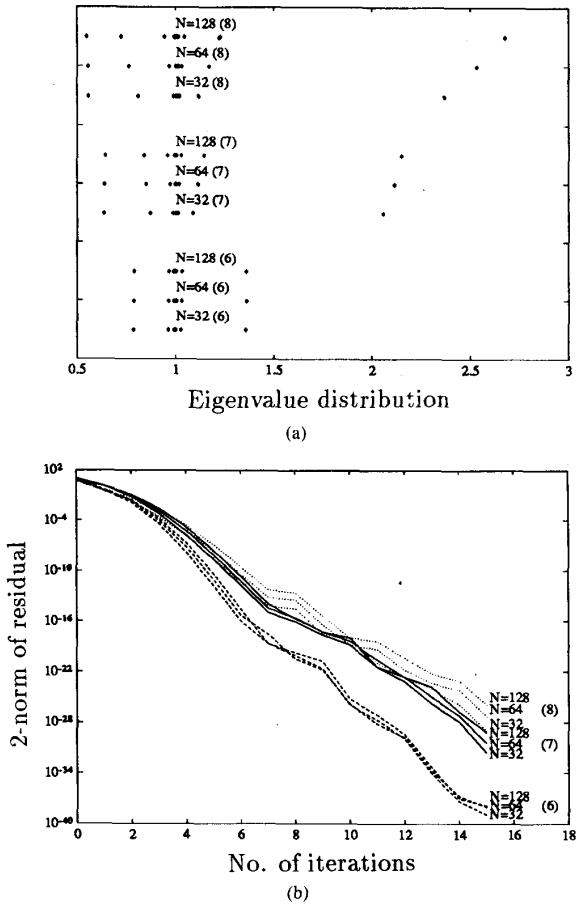


Fig. 6. (a) Eigenvalue distribution of $K_1^{-1}A$ with $N = 32$, 64, and 128, and (b) their corresponding convergence history for problems 6–8.

analyze their spectral properties. This new class of preconditioners are very attractive for Toeplitz matrices with rational generating functions.

The convolutional viewpoint not only provides ways to use all elements given by Toeplitz matrices so that preconditioned matrices may have better spectral properties. It also suggests naturally how to generalize the preconditioning technique to block Toeplitz matrices, which arise in many 2-D signal processing and estimation problems. This subject is under our current investigation. It appears that, in comparison with direct methods, the reduction of the computational complexity of the PCG method can be even more significant for 2-D problems than for 1-D problems. More research along this direction is expected in the near future.

We also found from numerical experiments that, for Toeplitz matrices A with rational generating functions, there exist strong regularities in the number of outliers and the clustering radius for the spectra of preconditioned Toeplitz matrices with Strang's preconditioner S or the proposed preconditioners K_i. These regularities have recently been analyzed and reported in [15]. One potential application of these regularities is to estimate the order of an ARMA model by examining the convergence history of the PCG method.

Appendix A
Proof of Lemma 1

For an $N \times N$ doubly symmetric matrix B, we can express it in form [4]

$$B = \begin{bmatrix} B_1 & JB_2J \\ B_2 & JB_1J \end{bmatrix}, \quad \text{for even } N$$

or

$$B = \begin{bmatrix} B_1 & \boldsymbol{b} & JB_2J \\ \boldsymbol{b}^T & c_b & \boldsymbol{b}^TJ \\ B_2 & J\boldsymbol{b} & JB_1J \end{bmatrix}, \quad \text{for odd } N$$

where B_1, B_2, and J are $\lfloor N/2 \rfloor \times \lfloor N/2 \rfloor$ matrices with $B_1^T = B_1$ and $B_2^T = JB_2J$, $\boldsymbol{b}$ is a column vector of length $\lfloor N/2 \rfloor$, and c_b is a constant. By defining the orthonormal matrix

$$Q = \frac{1}{\sqrt{2}} \begin{bmatrix} I & I \\ -J & J \end{bmatrix}, \quad \text{for even } N$$

or

$$Q = \frac{1}{\sqrt{2}} \begin{bmatrix} I & 0 & I \\ 0 & \sqrt{2} & 0 \\ -J & 0 & J \end{bmatrix}, \quad \text{for odd } N$$

we can decouple the eigenproblem of B into two separated subproblems, i.e.,

$$Q^{-1}BQ = Q^TBQ = \begin{bmatrix} B_1 - JB_2 & 0 \\ 0 & B_1 + JB_2 \end{bmatrix},$$

for even N

or

$$Q^{-1}BQ = Q^TBQ = \begin{bmatrix} B_1 - JB_2 & 0 & 0 \\ 0 & c_b & \sqrt{2}\boldsymbol{b}^T \\ 0 & \sqrt{2}\boldsymbol{b} & B_1 + JB_2 \end{bmatrix},$$

for odd N.

For the generalized eigenvalue problem

$$A\boldsymbol{x} = \lambda B\boldsymbol{x}$$

with doubly symmetric A and B, we can transform it to another generalized eigenvalue problem,

$$\tilde{A}\boldsymbol{y} = \lambda \tilde{B}\boldsymbol{y}$$

where $\tilde{A} = Q^{-1}AQ$, $\tilde{B} = Q^{-1}BQ$ and $\boldsymbol{x} = Q\boldsymbol{y}$.

Now, $\tilde{A}$ and $\tilde{B}$ are block diagonal matrices and the eigenvectors of $\tilde{B}^{-1}\tilde{A}$ can be written as

$$\begin{bmatrix} \boldsymbol{y}_1 \\ \boldsymbol{0} \end{bmatrix} \text{ or } \begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{y}_2 \end{bmatrix} \quad \text{for even } N$$

and

$$\begin{bmatrix} y_1 \\ 0 \\ \mathbf{0} \end{bmatrix} \quad \text{or} \quad \begin{bmatrix} \mathbf{0} \\ \alpha \\ y_3 \end{bmatrix} \qquad \text{for odd } N$$

where y_1, y_2, $(\alpha, y_3^T)^T$ are eigenvectors of the following generalized eigenvalue problems:

$$(A_1 - JA_2)y_1 = \lambda_1(B_1 - JB_2)y_1$$

$$(A_1 + JA_2)y_2 = \lambda_2(B_1 + JB_2)y_2$$

and

$$\begin{bmatrix} c_a & \sqrt{2}\boldsymbol{a}^T \\ \sqrt{2}\boldsymbol{a} & A_1 + JA_2 \end{bmatrix}\begin{bmatrix} \alpha \\ y_3 \end{bmatrix} = \lambda_3 \begin{bmatrix} c_b & \sqrt{2}\boldsymbol{b}^T \\ \sqrt{2}\boldsymbol{b} & B_1 + JB_2 \end{bmatrix}\begin{bmatrix} \alpha \\ y_3 \end{bmatrix}.$$

Through the transformation $\boldsymbol{x} = Q\boldsymbol{y}$, the eigenvector of $B^{-1}A$ can be written as

$$\frac{1}{\sqrt{2}}\begin{bmatrix} y_1 \\ -Jy_1 \end{bmatrix} \quad \text{or} \quad \frac{1}{\sqrt{2}}\begin{bmatrix} y_2 \\ Jy_2 \end{bmatrix} \qquad \text{for even } N$$

and

$$\frac{1}{\sqrt{2}}\begin{bmatrix} y_1 \\ 0 \\ -Jy_1 \end{bmatrix} \quad \text{or} \quad \frac{1}{\sqrt{2}}\begin{bmatrix} y_3 \\ \sqrt{2}\alpha \\ Jy_3 \end{bmatrix} \qquad \text{for odd } N$$

which are skew-symmetric and symmetric, respectively. It is clear that there are $\lceil N/2 \rceil$ symmetric and $\lfloor N/2 \rfloor$ skew-symmetric eigenvectors for $B^{-1}A$. □

Appendix B
Proof of Theorem 1

Let us define $E_i = A^{-1}(K_i - A)$, $i = 1, 2, 3, 4$. By using the property that the inverse of a doubly symmetric matrix is still doubly symmetric [18], we know that A^{-1} commutes with J. Therefore, we have

$$E_2 = -E_1, \qquad E_3 = JE_1, \qquad E_4 = -JE_1.$$

From above, it is clear that $Q_1 = Q_2$ and $Q_3 = Q_4$. Due to Lemma 1, E_i, $i = 1, 2, 3, 4$, have a set of $\lceil N/2 \rceil$ symmetric eigenvectors and $\lfloor N/2 \rfloor$ skew-symmetric eigenvectors. Let $\boldsymbol{x}$ be a symmetric eigenvector of E_1 with eigenvalue λ. Since

$$E_3\boldsymbol{x} = JE_1\boldsymbol{x} = E_1J\boldsymbol{x} = E_1\boldsymbol{x} = \lambda\boldsymbol{x}$$

the vector $\boldsymbol{x}$ is also an eigenvector of E_3 with the same eigenvalue λ. Similar arguments apply to the skew-symmetric eigenvector of E_1. Thus, $Q_1 = Q_3$ and the proof is completed. □

Appendix C
Proof of Theorem 2

Let R_{2N} be the $2N \times 2N$ circulant matrix

$$\begin{bmatrix} A & \Delta A \\ \Delta A & A \end{bmatrix}$$

whose first row is specified by $(a_0, a_1, \cdots, a_{N-1}, a_N, a_{N-1}, \cdots, a_1)$. It is clear that

$$\begin{bmatrix} A & \Delta A \\ \Delta A & A \end{bmatrix}\begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{x} \end{bmatrix} = \lambda \begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{x} \end{bmatrix} \Leftrightarrow (A + \Delta A)\boldsymbol{x} = \lambda\boldsymbol{x}.$$

Therefore, if λ is an eigenvalue of K_1 with eigenvector $\boldsymbol{x}$, λ is also an eigenvalue of R_{2N} with eigenvector $(\boldsymbol{x}^T, \boldsymbol{x}^T)^T$. Since R_{2N} is symmetric circulant, the eigenvalue λ can be written as

$$\lambda = \sum_{n=-(N-1)}^{N} a_n e^{-i2\pi kn/2N}$$

$$= a_0 + \sum_{n=1}^{N-1} a_n 2\cos\left(\frac{2\pi kn}{2N}\right) + (-1)^k a_N$$

which is real and equal to a partial sum of the infinite series $\Sigma_{-\infty}^{\infty} a_n e^{-in\theta}$ from $n = 1 - N$ to N. With conditions (25) and (26), we conclude that eigenvalues of K_1 are uniformly positive and bounded for large N. Similarly, we can show that eigenvalues of K_i, $i = 2, 3, 4$, are uniformly positive and bounded for large N. □

Appendix D
Proof of Theorem 3

Let A_M and ΔA_M be the leading $M \times M$ submatrices of A and ΔA, respectively. For a constant M

$$\|\Delta A_M\|_1 = \max_j \sum_{i=1}^{M} |(\Delta A_M)_{i,j}| \le \sum_{n=N+1-M}^{N} 2|a_n| \le \gamma. \tag{30}$$

When ΔA_M is symmetric, we have $\|\Delta A_M\|_\infty = \|\Delta A_M\|_1$ [13]. Thus

$$\|\Delta A_M\|_2 \le (\|\Delta A_M\|_1 \|\Delta A_M\|_\infty)^{1/2} \le \gamma. \tag{31}$$

Since eigenvalues of A_M are bounded by the maximum and the minimum eigenvalue of A [13]. With the assumption that A is bounded and uniformly positive definite, 2-norm of A is bounded by $c = 1/\delta$ and

$$\|A_M^{-1}\|_2 \le \|A^{-1}\|_2 \le c.$$

By the minimax theorem (or the Courant-Fisher theorem) of eigenvalues [23], [25], $A^{-1}\Delta A$ has at most $2(N - M)$ eigenvalues with magnitude larger than $\|A_M^{-1}\Delta A_M\|_2$. Since

$$\|A_M^{-1}\Delta A_M\|_2 \le \|A_M^{-1}\|_2 \|\Delta A_M\|_2 \le c\gamma = \epsilon$$

$A^{-1}(K_1 - A)$ has at most $2(N - M)$ eigenvalues with magnitude larger than ϵ. The same arguments can also be applied to preconditioners K_2, K_3 and K_4. This completes the proof. □

The above proof relies on arguments from matrix analysis. However, we want to point out that Chan and Strang [7] used the theory of collectively compact operators to prove a clustering result under a weaker assumption where $f(\theta)$ in (25) is continuous but not necessarily in the Wiener class.

ACKNOWLEDGMENT

The authors were grateful to Dr. B. Levy for helpful discussions. Some comments from the reviewers were also highly appreciated.

REFERENCES

[1] G. S. Ammar and W. B. Gragg, "Superfast solution of real positive definite Toeplitz systems," *SIAM J. Matrix Anal. Appl.*, vol. 9, pp. 61–76, 1988.

[2] R. R. Bitmead and B. D. Anderson, "Asymptotically fast solution of Toeplitz and related systems of equations," *Lin. Algeb. Appl.*, vol. 34, pp. 103–116, 1980.

[3] R. P. Brent, F. G. Gustavson, and D. Y. Yun, "Fast solution of Toeplitz systems of equations and computations of Padé approximations," *J. Algorithms*, vol. 1, pp. 259–295, 1980.

[4] A. Cantoni and P. Butler, "Eigenvalues and eigenvectors of symmetric centrosymmetric matrices," *Lin. Algeb. Appl.*, vol. 13, pp. 275–288, 1976.

[5] R. H. Chan, "Circulant preconditioners for Hermitian Toeplitz system," *SIAM J. Matrix Anal. Appl.*, vol. 10, pp. 542–550, Oct. 1989.

[6] R. H. Chan, "The spectrum of a family of circulant preconditioned Toeplitz systems," *SIAM J. Numer. Anal.*, vol. 26, pp. 503–506, Apr. 1989.

[7] R. H. Chan and G. Strang, "Toeplitz equations by conjugate gradients with circulant preconditioner," *SIAM J. Sci. Stat. Comput.*, vol. 10, pp. 104–119, Jan. 1989.

[8] T. F. Chan, "An optimal circulant preconditioner for Toeplitz systems," *SIAM J. Sci. Stat. Comput.*, vol. 9, pp. 766–771, July 1988.

[9] P. Davis, *Circulant Matrices*. New York: Wiley, 1979.

[10] P. Delsarte and Y. V. Genin, "The split Levinson algorithm," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-34, pp. 470–478, June 1986.

[11] B. W. Dickinson, "Efficient solution of linear equations with banded Toeplitz matrices," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-27, pp. 421–422, Aug. 1979.

[12] B. W. Dickinson, "Solution of linear equations with rational Toeplitz matrices," *Math. Comput.*, vol. 34, pp. 227–233, Jan. 1980.

[13] G. H. Golub and C. F. Van Loan, *Matrix Computations*. Baltimore, MD: John Hopkins University Press, 1983.

[14] F. D. Hoog, "A new algorithm for solving Toeplitz systems of equations," *Lin. Algeb. Appl.*, vol. 88/89, pp. 123–138, 1987.

[15] T. K. Ku and C. J. Kuo, "Spectral properties of preconditioned rational Toeplitz matrices," Tech. Rep. 163, Signal and Image Processing Inst., Univ. Southern California, Sept. 1990.

[16] N. Levinson, "The Wiener RMS error criterion in filter design and prediction," *J. Math. Phys.*, vol. 25, pp. 261–278, 1947.

[17] D. G. Luenberger, *Linear and Nonlinear Programming*. Reading, MA: Addison-Wesley, 1984.

[18] J. Makhoul, "On the eigenvectors of symmetric Toeplitz matrices," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-29, pp. 868–872, Aug. 1981.

[19] H. S. Malvar, "Fast computation of the discrete cosine transform and the discrete Hartley transform," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-35, pp. 1484–1485, Oct. 1987.

[20] A. V. Oppenheim and R. W. Schafer, *Digital Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1975.

[21] H. V. Sorensen, D. L. Jones, M. T. Heideman, and C. S. Burrus, "Real-value fast Fourier transform algorithms," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-35, pp. 849–863, June 1987.

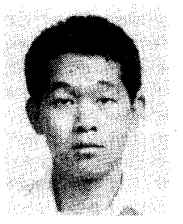
[22] G. Strang, "A proposal for Toeplitz matrix calculations," *Stud. Appl. Math.*, vol. 74, pp. 171–176, 1986.

[23] G. Strang, *Linear Algebra and Its Applications*, third ed. Orlando, FL: Harcourt, Brace Jovanovich, 1988.

[24] G. Strang and A. Edelman, "The Toeplitz-circulant eigenvalue problem $Ax = \lambda Cx$," in *Proc. Oakland Conf. PDE's*, 1987.

[25] J. H. Wilkinson, *The Algebraic Eigenvalue Problem*. New York: Oxford, 1965.

[26] P. Yip and K. R. Rao, "A fast computational algorithm for the discrete sine transform," *IEEE Trans. Commun.*, vol. COM-28, pp. 304–307, 1980.



Ta-Kang Ku was born in Taipei, Taiwan, in 1962. He received the B.S. degree in electrical engineering from the National Taiwan University, Taipei, Taiwan, in 1984 and the M.S. degree in electrical engineering from the University of Southern California, Los Angeles, in 1988. He is currently pursuing the Ph.D. degree in electrical engineering at the Signal and Image Processing Institute at the University of Southern California.

Since 1989 he has been employed by the Signal and Image Processing Institute, University of Southern California, as a Research Assistant. His current research interests include signal and image processing, numerical analysis, and parallel computation.

C.-C. Jay Kuo (S'83–M'86) was born in Hsinchu, Taiwan, in 1957. He received the B.S. degree from the National Taiwan University, Taipei, Taiwan, in 1980 and the M.S. and Ph.D. degrees from the Massachusetts Institute of Technology, Cambridge, in 1985 and 1987, respectively, all in electrical engineering.

From October 1987 to December 1988, he was an Assistant Professor of computational and applied mathematics (CAM) in the Department of Mathematics at the University of California, Los Angeles. Since January 1989, he has been with the Department of Electrical Engineering-Systems and the Signal and Image Processing Institute at the University of Southern California, where he is currently Assistant Professor. His research interests are in the areas of digital signal processing, numerical analysis, and parallel computation.

Dr. Kuo is a member of Sigma Xi, SIAM, and ACM.