USC-SIPI REPORT #155

Design and Analysis of Toeplitz Preconditioners

by

Takang Ku and C.C. Jay Kuo

May 1990

Signal and Image Processing Institute
UNIVERSITY OF SOUTHERN CALIFORNIA
Department of Electrical Engineering-Systems
Powell Hall of Engineering
University Park/MC-0272
Los Angeles, CA 90089 U.S.A.

Design and Analysis of Toeplitz Preconditioners*

Takang Ku[†] and C.-C. Jay Kuo [†]

May, 1990

Abstract

The solution of symmetric positive definite Toeplitz systems $A\mathbf{x} = \mathbf{b}$ by the preconditioned conjugate gradient (PCG) method was recently proposed by Strang [21] and analyzed by Strang and R. Chan [7]. The convergence rate of the PCG method heavily depends on the choice of preconditioners for 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 $O(N \log N)$.

*This work was supported by USC Faculty Research and Innovation Fund.

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

1 Introduction

The solution of an $N \times N$ symmetric positive definite (SPD) Toeplitz system $A\mathbf{x} = \mathbf{b}$ by *direct* methods has been studied intensively in the past. Fast algorithms based on Levinson recursion formula [11] [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], [2], [3], [15]. More recently, Strang [21] 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 [21] 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 $A\mathbf{x} = \mathbf{b}$ can be derived by minimizing the quadratic functional $\frac{1}{2}\mathbf{x}^T A\mathbf{x} - \mathbf{b}^T\mathbf{x}$ 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 smaller condition number or clustered eigenvalues. In order to accelerate its convergence rate, a preconditioning step is often introduced at each CG iteration. A good preconditioner for A, is a matrix P that approximates A well (in the sense that the spectrum for the preconditioned matrix $P^{-1}A$ is clustered around 1 or has a small condition number), and for which the matrix-vector product $P^{-1}\mathbf{v}$ can be computed efficiently for a given vector $\mathbf{v}$. With such a preconditioner, one then solves in principle the preconditioned system $\tilde{A}\tilde{\mathbf{x}} = \tilde{\mathbf{b}}$, where $\tilde{A} = P^{-1/2}AP^{-1/2}$, $\tilde{\mathbf{x}} = P^{1/2}\mathbf{x}$ and $\tilde{\mathbf{b}} = P^{-1/2}\mathbf{b}$, by the CG method [14]. 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 [21] and analyzed by Strang and R. Chan [5] [6] [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 $S^{-1}\mathbf{v}$ can be conveniently computed via Fast Fourier Transform (FFT) with $O(N \log N)$ operations. It has been shown [5] [6] [7] that for a large class of matrices (called the Wiener class), the spectrum $S^{-1}A$ is clustered around 1 except a finite number of outsiders.

In constructing Strang's preconditioner S, half the information of A is lost. In order to use all information of A, T. 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.

This research was motivated by seeking another direction to generalize Strang's preconditioner so that all information 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 (FCT), or Fast Sine Transform (FST). 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 $2N \times 2N$.

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 2. Then, we propose a general framework to construct Toeplitz preconditioners by exploiting the relationship between linear and circular deconvolutions in Section 3. In particular, a class of new preconditioners K_i, $i = 1, 2, 3, 4$, which use all elements of A are described. In Section 4, 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 5, we give some numerical results and compare the performance of different preconditioners. The efficiency of new preconditioners K_i are demonstrated.

2 THE PCG METHOD FOR TOEPLITZ SYSTEMS

With the initialization

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

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

Step 1: Preconditioning. Solve

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

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

Step 2: CG iteration. Compute

$$\begin{aligned}
\beta_k &= (\mathbf{z}_{k-1}, \mathbf{r}_{k-1})/(\mathbf{z}_{k-2}, \mathbf{r}_{k-2}), \\
\mathbf{p}_k &= \mathbf{z}_{k-1} + \beta_k \mathbf{p}_{k-1}, \\
\alpha_k &= (\mathbf{z}_{k-1}, \mathbf{r}_{k-1})/(\mathbf{p}_k, A\mathbf{p}_k), \\
\mathbf{x}_k &= \mathbf{x}_{k-1} + \alpha_k \mathbf{p}_k, \\
\mathbf{r}_k &= \mathbf{r}_{k-1} - \alpha_k A\mathbf{p}_k.
\end{aligned}$$

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 $A\mathbf{p}_k$ and the preconditioning $P^{-1}\mathbf{r}_{k-1}$, requires $O(N)$ operations. Since we can view $A\mathbf{p}_k$ as a circular convolution between two extended periodic sequences, the Toepliz matrix-vector product can be computed via FFT with $O(N \log N)$ operations. We will show that the preconditioning $P^{-1}\mathbf{r}_{k-1}$ can also be achieved by various fast transform algorithms with $O(N \log N)$ operations in Section 3. 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 $\mathbf{x}_k - \mathbf{x}^*$, where $\mathbf{x}^*$ is the exact solution of $A\mathbf{x} = \mathbf{b}$, with

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

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

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

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 outsiders λ_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

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

The inequality (2) can be simplified to be

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

where

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

In deriving (4), we assume that α outsiders are annihilated and the error reduction of $R(\mathbf{x}_k)$ simply depends on eigenvalues clustered around one. It implies that, when $k > \alpha$, $R(\mathbf{x}_k)$ can be reduced at least by a factor ϵ^2 per iteration in average. Thus, the number of outsiders 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 5.

3 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 and 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 4.

3.1 Motivation: A Convolutional Interpretation

Let $\mathbf{u}_N = (u_0, u_1, \cdots, u_{N-1})^T$ and $\mathbf{v}_N = (v_0, v_1, \cdots, v_{N-1})^T$ 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 interpret the matrix-vector products $T_N\mathbf{u}_N$, $R_N\mathbf{u}_N$, $T_N^{-1}\mathbf{v}_N$ and $R_N^{-1}\mathbf{v}_N$, from a convolutional point of view, since our approach to the design of Toeplitz preconditioners can be well motivated by this viewpoint.

First, consider $\mathbf{v}_N = T_N\mathbf{u}_N$. 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, equation (6) with any integers i and j defines a linear convolution

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

where

$$\mathbf{t} = \cdots, 0, t_{-(N-1)}, \cdots, t_{-1}, t_0, t_1, \cdots, t_{N-1}, 0, \cdots \quad \text{and} \quad \mathbf{u} = \cdots, 0, u_0, u_1, \cdots, u_{N-1}, 0, \cdots.$$

Note that $\mathbf{v}$, $\mathbf{t}$ and $\mathbf{u}$ in (7) are infinite sequences of duration $3N-2$, $2N-1$ and N, respectively. In linear system theory, $\mathbf{u}$ and $\mathbf{v}$ are usually known as the input and output, and $\mathbf{t}$ the impulse response of the system [19]. Since the output $\mathbf{v}$ contains elements v_i of $\mathbf{v}_N$, Toeplitz matrix-vector product $\mathbf{v}_N = T_N\mathbf{u}_N$ is embedded in the linear convolution (7). For (7), we can define a linear deconvolution problem, namely, to determine the input $\mathbf{u}$ from the output $\mathbf{v}$ and the impulse response $\mathbf{t}$.

Next, consider $\mathbf{v}_N = R_N\mathbf{u}_N$. 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

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

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

$$\mathbf{v}_N^T = (v_0, \cdots, v_{N-1}), \quad \mathbf{u}_N^T = (u_0, \cdots, u_{N-1}), \quad \text{and} \quad (r_0, \cdots, r_{N-1}).$$

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

The circular convolution and deconvolution can be performed effectively by using FFT.

That is, by applying the discrete Fourier transform, defined as

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

to periodic sequences $\tilde{\mathbf{v}}$, $\tilde{\mathbf{r}}$ and $\tilde{\mathbf{u}}$ 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 $\mathbf{v}_N = R_N \mathbf{u}_N$ ($\mathbf{u}_N = R_N^{-1} \mathbf{v}_N$) 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 $\mathbf{v}$, $\mathbf{t}$ and $\mathbf{u}$ 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 $\mathbf{v}_N = T_N \mathbf{u}_N$ can be embedded in (7) and since we know all nontrivial $2N-1$ and N values of $\mathbf{t}$ and $\mathbf{u}$, we can compute $\mathbf{v}$ as well as $\mathbf{v}_N$ effectively. However, the computation of $\mathbf{u}_N = T_N^{-1} \mathbf{v}_N$ is not as easy. Since only N values (i.e. $\mathbf{v}_N$) of the output $\mathbf{v}$ 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

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

to define a periodic sequence $\tilde{\mathbf{r}}$ of period N. Although the N-point circular deconvolution of $\tilde{\mathbf{r}}$ and $\tilde{\mathbf{v}}$ does not embed the desired computation $T_N^{-1}\mathbf{v}_N$, it can be viewed as its approximation and used in the preconditioning step of the PCG method. This was originally suggested by Strang [21] and analyzed by Strang and R. Chan [5] [6] [7]. One shortcoming of Strang's idea is that half of the information contained in t_n's is lost. To preserve all information of t_n's, we may choose to extend $\mathbf{v}$ periodically with $\mathbf{v}_N$ as the basic unit, which will be detailed below.

3.2 Construction of Toeplitz preconditioners

Let A be an $N \times N$ symmetric positive definite (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 $T_{N,1}$,

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

with a constant c. If the behavior of the sequence t_n is known, we choose c to be t_N. Otherwise, any $0 \leq |c| \leq |t_{N-1}|$ can be used.

Now, let us consider the following argumented system,

$$\begin{bmatrix} T_{N,1} & T_{N,2} \\ T_{N,2} & T_{N,1} \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \mathbf{x} \end{bmatrix} = \begin{bmatrix} \mathbf{b} \\ \mathbf{b} \end{bmatrix}. \tag{15}$$

From the discussion in Section 3.1, we know that (15) can be embedded by a circular convolution between two $2N$-periodic sequences, whose periods are

$$t_0, t_1, \cdots t_{N-2}, t_{N-1}, c, t_{N-1}, t_{N-2}, \cdots, t_1 \tag{16}$$

and

$$x_1, x_2, \cdots x_{N-1}, x_N, x_1, x_2, \cdots, x_{N-1}, x_N. \tag{17}$$

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

$$b_1, b_2, \cdots b_{N-1}, b_N, b_1, b_2, \cdots, b_{N-1}, b_N. \tag{18}$$

This is illustrated in Figure 1(a), where the case $N = 3$ is given. The solution of (15) for $\mathbf{x}$ 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

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

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

$$P_1 = T_{N,1} + T_{N,2}, \tag{19}$$

as a preconditioner for A.

Various preconditioners can be constructed in a similar way by assuming different periodicities for $\mathbf{x}$ and $\mathbf{b}$. The negative periodicity, even periodicity, and odd periodicity. are illustrated in Figures 1(b), 1(c) and 1(d), respectively. The corresponding argumented systems and preconditioners can be written as follows:

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

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

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

where J is the $N \times N$ symmetric elementary matrix [4] which has, by definition, ones along the secondary diagonal and zeros elsewhere (equivalently, $J_{i,j} = 1$ if $i + j = N + 1$ and $J_{i,j} = 0$ if $i + j \neq N + 1$).

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's will be detailed in Section 3.4. The subscript N of matrices is omitted hereinafter whenever there is no confusion.

3.3 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 3.2.

Example 1. (Strang's preconditioner)

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

$$T_1 = \begin{bmatrix} 32 & 16 & 8 & 0 & 0 \\ 16 & 32 & 16 & 8 & 0 \\ 8 & 16 & 32 & 16 & 8 \\ 0 & 8 & 16 & 32 & 16 \\ 0 & 0 & 8 & 16 & 32 \end{bmatrix}, \quad T_2 = \begin{bmatrix} 0 & 0 & 0 & 8 & 16 \\ 0 & 0 & 0 & 0 & 8 \\ 0 & 0 & 0 & 0 & 0 \\ 8 & 0 & 0 & 0 & 0 \\ 16 & 8 & 0 & 0 & 0 \end{bmatrix}.$$

Strang's preconditioner S is constructed by

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

which is a special case of P_1.

Example 2. (Chan's preconditioner)

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

$$c_i = \frac{1}{N}(i \times a_{(N-i)} + (N-i) \times a_i), \qquad i = 0, 1, 2, \cdots, N-1. \tag{23}$$

By choosing

$$T_1 = \begin{bmatrix} 32 & 13.2 & 6.4 & 0 & 0 \\ 13.2 & 32 & 13.2 & 6.4 & 0 \\ 6.4 & 13.2 & 32 & 13.2 & 6.4 \\ 0 & 6.4 & 13.2 & 32 & 13.2 \\ 0 & 0 & 6.4 & 13.2 & 32 \end{bmatrix}, \qquad T_2 = \begin{bmatrix} 0 & 0 & 0 & 6.4 & 13.2 \\ 0 & 0 & 0 & 0 & 6.4 \\ 0 & 0 & 0 & 0 & 0 \\ 6.4 & 0 & 0 & 0 & 0 \\ 13.2 & 6.4 & 0 & 0 & 0 \end{bmatrix},$$

where $c = 0$, we find that preconditioner C is also a special case of P_1, i.e.

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

Example 3. (Preconditioners K_i)

We use equations (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

$$K_1 = \begin{bmatrix} 33 & 18 & 12 & 12 & 18 \\ 18 & 33 & 18 & 12 & 12 \\ 12 & 18 & 33 & 18 & 12 \\ 12 & 12 & 18 & 33 & 18 \\ 18 & 12 & 12 & 18 & 33 \end{bmatrix}, \quad K_2 = \begin{bmatrix} 31 & 14 & 4 & -4 & -14 \\ 14 & 31 & 14 & 4 & -4 \\ 4 & 14 & 31 & 14 & 4 \\ -4 & 4 & 14 & 31 & 14 \\ -14 & -4 & 4 & 14 & 31 \end{bmatrix},$$

$$K_3 = \begin{bmatrix} 48 & 24 & 12 & 6 & 3 \\ 24 & 36 & 18 & 9 & 6 \\ 12 & 18 & 33 & 18 & 12 \\ 6 & 9 & 18 & 36 & 24 \\ 3 & 6 & 12 & 24 & 48 \end{bmatrix}, \qquad K_4 = \begin{bmatrix} 16 & 8 & 4 & 2 & 1 \\ 8 & 28 & 14 & 7 & 2 \\ 4 & 14 & 31 & 14 & 4 \\ 2 & 7 & 14 & 28 & 8 \\ 1 & 2 & 4 & 8 & 16 \end{bmatrix}.$$

Note 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

$$(a_0, a_1, \cdots, a_k, -a_k, \cdots, -a_1) \text{ for odd } N \quad \text{and} \quad k = (N-1)/2,$$

or

$$(a_0, a_1, \cdots, a_{k-1}, 0, -a_{k-1}, \cdots, -a_1) \text{ for even } N \quad \text{and} \quad k = N/2,$$

we say that B is *skew-circulant* [10]. 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.

3.4 Comparison of Computational Cost

We compare the computational cost for the preconditioning step $P^{-1}\mathbf{r}$ 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 $N \log N$ real multiplications and $3N \log N$ real additions [20]. 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 [10]. Consequently, the implementation of $K_2^{-1}\mathbf{r}$ is almost as easy as that of $K_1^{-1}\mathbf{r}$. Although preconditioners K_3 and K_4 are noncirculant, $K_3^{-1}\mathbf{r}$ and $K_4^{-1}\mathbf{r}$ 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 proportional constants [18] [22]. Therefore, they are as competitive as C, S, and K_i, $i = 1, 2$.

4 Spectral Properties of the Preconditioned Toeplitz Matrix

In this section, we consider the case $T_1 = A$. The corresponding T_2 is denoted by ΔA. Based on (19)-(22), four preconditioners can be derived, namely,

$$\begin{aligned} K_1 &= A + \Delta A, \quad K_2 = A - \Delta A, \\ K_3 &= A + J\Delta A, \quad K_4 = A - J\Delta A. \end{aligned} \tag{24}$$

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) If A belongs to the Wiener class ($\sum_{-\infty}^{\infty} |a_i| < \infty$) [7], all eigenvalues of $K_i^{-1}A$ except a finite number are clustered around 1.

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

$$A = A^T, \qquad \text{and} \qquad (JA)^T = (JA). \tag{25}$$

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

Lemma 1 *If T is a symmetric Toeplitz matrix, T and JT are doubly symmetric.*

Proof. This can be verified directly with definitions. □

A consequence of this lemma is that 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 (25) 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 2 *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

$$[\lambda(K_i^{-1}A)]^{-1} = \lambda(A^{-1}(A + K_i - A)) = \lambda(I + A^{-1}(K_i - A)) = 1 + \lambda(A^{-1}(K_i - A)). \quad (26)$$

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.*

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

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 $|\lambda|$. To illustrate this theorem, an example is given in Figure 2, where A is a 32×32 symmetric Toeplitz matrix with $a_i = 1/(1 + i)$. Eigenvalues of $A^{-1}(P - A)$ and $P^{-1}A$ are plotted for preconditioners C, S, and K_i in Figures 2(a) and 2(b), respectively. 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.

If preconditioners K_i are positive definite, the preconditioned matrices $K_i^{-1/2}AK_i^{-1/2}$ are symmetric positive definite and the CG method can be conveniently applied. To show the positive definiteness of K_i, we consider a sequence of $n \times n$ symmetric Toeplitz matrices $\{A_n\}_{n=1}^{\infty}$ and study the asymptotical behavior. The first row of A_N are elements from the infinite sequence $\{a_n\}_{n=0}^{\infty}$ up to element a_{N-1}, where $\{a_n\}_{n=0}^{\infty}$ is known as the generating sequence of A_n. We assume that the sequence a_n satisfy the following two conditions:

$$\sum_{-\infty}^{\infty} a_n e^{-in\theta} \geq \delta > 0, \quad \forall\theta, \tag{27}$$

$$\sum_{-\infty}^{\infty} |a_n| < \infty. \tag{28}$$

Since $f(\theta) = \sum_{-\infty}^{\infty} a_n e^{-in\theta}$ describes the asymptotic eigenvalue distribution of A_n, the above conditions assume that eigenvalues of A_n are bounded and uniformly positive definite asymptotically.

Theorem 2 *Preconditioners K_i, $i = 1,2,3,4$, for symmetric positive definite Toeplitz matrices with the generating sequence satisfying (28) and (29) are uniformly positive definite 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 R. Chan in [5].

Theorem 3 *Let A be the $N \times N$ leading matrix of a sequence symmetric positive definite Toeplitz matrices A_n with the generating sequence satisfying (28) and (29). The spectrum of* *the matrix $A^{-1}(K_i - A)$ are clustered between $(-\varepsilon, +\varepsilon)$ except a finite number of outsiders for sufficiently large $N(\varepsilon)$.*

Proof. See Appendix D. □

5 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, $i = 1, 2, 3, 4$ as well as the convergence history of the PCG method.

For a sequence of Toeplitz matrices A_n 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

$$A_+(z) = \frac{\sum_{n=0}^{p} c_n z^{-n}}{\sum_{n=0}^{q} d_n z^{-n}},$$

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 AR (Auto-Regressive), MA (Moving Average), 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 $\mathbf{b} = (1, \cdots, 1)^T$ and initial condition $\mathbf{x}_0 = 0$. We can roughly classify the eigenvalues of $P^{-1}A$ into two categories: the outsiders 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 outsiders 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 outsiders. The spectra of $P^{-1}A$ are plotted in Figure 3(a). For $K_i^{-1}A$, $i = 1, 2, 3, 4$, there are 6 ($p = 3$) outsiders and $N - 6$ eigenvalues repeated at one. For $K_i^{-1}A$, $i = 3, 4$, each pair of outsiders 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.

According to the discussion in Section 2, 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 that the PCG method with K_i converges in $p + 1$ iterations for banded Toeplitz matrices with different values of p. In Figure 3(b), we plot the 2-norm of the residual $\mathbf{b} - A\mathbf{x}$ 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. The PCG method with C converges slowlier.

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

For this generating sequence, it has been observed by Strang that the spectrum of $S^{-1}A$ has two outsiders 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 1 and plot the spectra of $P^{-1}A$ with $t = 0.9$ in Figure 4(a). For preconditioners K_i, the two outsiders are located at $(1+t)^{-1}$ or $(1-t)^{-1}$ and other $N-2$ clustered eigenvalues are repeated at $(1-t^N)^{-1}$ or $(1+t^N)^{-1}$. The outsiders of $K_i^{-1}A$, $i = 3,4$ are repeated with multiplicity 2.

Table 1. 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	$(1-t^N)^{-1}$	$(1+t^N)^{-1}$	$(1-t^N)^{-1}$	$(1-t^N)^{-1}$

The convergence history of the PCG method with $t = 0.9$ is given in Figure 4(b). Since $K_i^{-1}A$ has 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 Figures 5(a). The spectra of $P^{-1}A$ consists of 4 outsiders and $N-4$ clustered eigenvalues between $(1-\epsilon, 1+\epsilon)$. Similar to Problem 1, each pair of outsiders of $K_i^{-1}A$, $i = 3, 4$, are closely located so that only two distinct dots appear. The corresponding convergence history are plotted in Figure 5(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.

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 outsiders is equal to $2 \times \max(p, q)$.

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

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

Table 2.

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

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

Table 3.

t	$\epsilon(S^{-1}A)$	$\frac{a_{16}}{a_0}$	$\epsilon(K_i^{-1}A)$	$\frac{a_{32}}{a_0}$
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}

We still do not understand Rules *R1* and *R2* theoretically. Rule *R2* nevertheless explains why our preconditioners K_i behave better than Strang's preconditioner S. From *R2*, we have

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

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

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 Figure 6(a). There are 6 ($\max(p,q) = 3$) outsiders for $K_i^{-1}A$ and $S^{-1}A$. The outsiders of $K_i^{-1}A$, $i = 3,4$, are clustered into three distinct dots. The clustering radii ϵ, $\frac{a_{16}}{a_0}$ and $\frac{a_{32}}{a_0}$ with different t_0 and t_1 are given in Table 4 to verify Rule *R2*.

Table 4.

t_0	t_1	$\epsilon(S^{-1}A)$	$\frac{a_{16}}{a_0}$	$\epsilon(K_i^{-1}A)$	$\frac{a_{32}}{a_0}$
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}

The convergence history of the PCG method with $t_0 = 0.3$ and $t_1 = 0.8$ is given in Figure 6(b). Preconditioners K_i behave better than C and S. It takes approximately 7 (or 10) iterations for K_i (or S) to converge.

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 $t_1 = 0.6$ are plotted in Figure 7(a). There are 8 ($\max(p, q) = 4$) outsiders for K_i and S. The outsiders 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 Figure 7(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 exists direct methods which solve the system with $O(pN)$ operations [12]. Additionally, if $A_+(z)$ is of order (p, q), $q > 0$, Dickinson proposed a method to transform $A\mathrm{x} = \mathrm{b}$ into an equivalent symmetric banded system $\tilde{A}\tilde{\mathrm{x}} = \tilde{\mathrm{b}}$

with upper bandwidth $\max(p,q)$, whose solution can be obtained with $\max(p,q) \times O(N)$ operations [13]. However, this transformation requires the knowledge of the exact form of $A(z)$ and takes $O(N \log N)$ operations. Thus, the total computational complexity is $O(N \log N)$, which is the same as that of the PCG method.

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.

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 slowlier than $|a_n|$ in Problems 1-5 asymptotically. The numbers of iterations required to achieve $||\mathbf{b} - A\mathbf{x}||_2 \leq 10^{-15}$ are summarized in Table 5 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.

Table 5.

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

In order to understand the asymptotical behavior, we consider a typical case $P = K_1$ and perform experiments for problems with sizes 32, 64, and 128. We plot the spectra of $K_1^{-1}A$ and the corresponding convegence history for Problems 6-8 in Figures 8(a) and 8(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 $O(N \log N)$. This is lower than that of fast or superfast direct methods.

6 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 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 information 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. This is under our current investigation. We also found from numerical experiments that, for Toeplitz matrices A with rational generating functions, there exist strong regularities in the number α of outsiders and the clustering radius ϵ for the spectra of $P^{-1}A$, where P is either Strang's preconditioner S or our preconditioners K_i. How to explain these regularities analytically is still open.

Appendices

Appendix A: Proof of Lemma 2.

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

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

or

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

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 $B_2^T = JB_2J$, b is a column vector of length $\lfloor N/2 \rfloor$, and c_b is a constant. By defining the orthonormal matrix

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

or

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

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

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

or

$$Q^{-1}BQ = Q^TBQ = \begin{bmatrix} B_1 - JB_2 & 0 & 0 \\ 0 & c_b & \sqrt{2}\mathrm{b}^T \\ 0 & \sqrt{2}\mathrm{b} & B_1 + JB_2 \end{bmatrix}, \qquad \text{for odd } N.$$

For the generalized eigenvalue problem

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

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

$$\tilde{A}\mathbf{y} = \lambda\tilde{B}\mathbf{y},$$

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

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

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

and

$$\begin{bmatrix} \mathbf{y}_1 \\ 0 \\ \mathbf{0} \end{bmatrix} \quad \text{or} \quad \begin{bmatrix} \mathbf{0} \\ \alpha \\ \mathbf{y}_3 \end{bmatrix} \quad \text{for odd } N,$$

where $\mathbf{y}_1$, $\mathbf{y}_2$, $(\alpha, \mathbf{y}_3^T)^T$ are eigenvectors of the following generalized eigenvalue problems:

$$(A_1 - JA_2)\mathbf{y}_1 = \lambda_1(B_1 - JB_2)\mathbf{y}_1,$$

$$(A_1 + JA_2)\mathbf{y}_2 = \lambda_2(B_1 + JB_2)\mathbf{y}_2,$$

and

$$\begin{bmatrix} c_a & \sqrt{2}\mathbf{a}^T \\ \sqrt{2}\mathbf{a} & A_1 + JA_2 \end{bmatrix} \begin{bmatrix} \alpha \\ \mathbf{y}_3 \end{bmatrix} = \lambda_3 \begin{bmatrix} c_b & \sqrt{2}\mathbf{b}^T \\ \sqrt{2}\mathbf{b} & B_1 + JB_2 \end{bmatrix} \begin{bmatrix} \alpha \\ \mathbf{y}_3 \end{bmatrix}.$$

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

$$\frac{1}{\sqrt{2}}\begin{bmatrix} \mathbf{y}_1 \\ -J\mathbf{y}_1 \end{bmatrix} \text{ or } \frac{1}{\sqrt{2}}\begin{bmatrix} \mathbf{y}_2 \\ J\mathbf{y}_2 \end{bmatrix} \quad \text{for even } N,$$

and

$$\frac{1}{\sqrt{2}}\begin{bmatrix} \mathbf{y}_1 \\ 0 \\ -J\mathbf{y}_1 \end{bmatrix} \text{ or } \frac{1}{\sqrt{2}}\begin{bmatrix} \mathbf{y}_3 \\ \sqrt{2}\alpha \\ J\mathbf{y}_3 \end{bmatrix} \quad \text{for odd } N,$$

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.

If λ is an eigenvalue of $A^{-1}(K_1 - A)$, it satisfies

$$\triangle A\mathbf{x} = \lambda A\mathbf{x}. \tag{29}$$

Now, since

$$(K_2 - A)\mathbf{x} = -\triangle A\mathbf{x} = -\lambda A\mathbf{x}, \tag{30}$$

we conclude that $-\lambda$ is an eigenvalue of $A^{-1}(K_2 - A)$ and $Q_1 = Q_2$. Similarly, we can show that if λ is an eigenvalue of $A^{-1}(K_3 - A)$, $-\lambda$ is an eigenvalue of $A^{-1}(K_4 - A)$ and, therefore, $Q_3 = Q_4$.

Since A, $\triangle A$ and $J\triangle A$ are all doubly symmetric by Lemma 1 and A is positive definite, we can find a set of eigenvectors of $A^{-1}\triangle A$ and $A^{-1}J\triangle A$ which are either symmetric or skew-symmetric by Lemma 2. Let x be a symmetric eigenvector of $A^{-1}(K_1 - A)$ with eigenvalue λ. By substituting $\mathbf{x} = J\mathbf{x}$ into the L.H.S. of (30), we obtain

$$\triangle AJ\mathbf{x} = \lambda A\mathbf{x}. \tag{31}$$

In addition, we have

$$(K_3 - A)\mathbf{x} = J\Delta A\mathbf{x} = \Delta AJ\mathbf{x} \tag{32}$$

where the last equality is due to the commutability of J and ΔA. From (32) and (33), we know that the eigenvalue λ of $A^{-1}(K_1 - A)$ associated with a symmetric eigenvector is also an eigenvalue of $A^{-1}(K_3 - A)$. On the other hand, if x is a skew-symmetric eigenvector of $A^{-1}(K_1 - A)$ with eigenvalue λ, we can similarly show that λ is an eigenvalue of $A^{-1}(K_4 - A)$. This implies that $Q_1 \subset Q_3(= Q_3 \cup Q_4)$. With the same arguments, we can also derive $Q_3 \subset Q_1(= Q_1 \cup Q_2)$. Hence, $Q_1 = Q_3$ 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

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

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

$$\lambda = \sum_{n=-(N-1)}^{N} a_n e^{\frac{-i2\pi kn}{2N}} = a_0 + \sum_{n=1}^{N-1} a_n 2\cos(\frac{2\pi kn}{2N}) + (-1)^k a_N, \tag{33}$$

which is real and equal to a partial sum of the infinite series $\sum_{-\infty}^{\infty} a_n e^{-in\theta}$ from $n = 1 - N$ to N. With conditions (28) and (29), 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 $\triangle A_q$ be the leading $q \times q$ submatrix of $\triangle A$.

$$||\triangle A_q||_1 = \max_j \sum_{i=1}^{q} |(\triangle A_q)_{i,j}| \leq \sum_{n=N+1-q}^{N} |a_n| \leq \gamma. \tag{34}$$

Since $\triangle A_q$ is symmetric, we have $||\triangle A_q||_\infty = ||\triangle A_q||_1$ [14]. Thus

$$||\triangle A_q||_2 \leq (||\triangle A_q||_1 ||\triangle A_q||_\infty)^{1/2} \leq \gamma. \tag{35}$$

Hence, the spectrum of $\triangle A_q$ clustered between $(-\gamma\ ,\ \gamma)$, and there are at most $2(N-q)$ eigenvalues of $K_1 - A$ outside the range ($-\gamma$, γ) by the interlacing theorem [9]. With the assumption that A is uniformly positive definite, A^{-1} is bounded by a constant c. Thus, for given ε, we can choose sufficiently large N such that

$$||A^{-1}(K_1 - A)||_2 \leq ||A^{-1}||_2 ||K_1 - A||_2 \leq c\gamma = \varepsilon.$$

The same arguments can also be applied to preconditioners K_2, K_3 and K_4. This completes the proof. □

References

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

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

[3] R. Brent, F. Gustavson, and D. 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. Chan, "Circulant preconditioners for Hermitian Toeplitz system," *SIAM J. Matrix Anal. Appl.*, Vol. 10, pp. 542–550, Oct. 1989.

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

[7] R. 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. Chan, "An optimal circulant preconditioner for Toeplitz systems," *SIAM J. Sci. Stat. Comput.*, Vol. 9, pp. 766–771, July 1988.

[9] J. K. Cullum and R. A. Willoughby, *Lanczos Algorithms for Large Symmetric Eigenvalue Computations*, vol. I. Theory, Birkhauser, 1985.

[10] P. Davis, *Circulant matrices*, New York: Wiley, 1979.

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

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

[13] B. Dickinson, "Solution of linear equations with rational Toeplitz matrices," *Math. Comp.*, Vol. 34, pp. 227–233, Jan. 1980.

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

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

[16] N. Levnison, "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*, Addison-Wesley, 1984.

[18] H. 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.

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

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

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

[22] 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.

Figure Captions

Figure 1: Circular convolutions for preconditioners K_i.

Figure 2: Eigenvalue distribution of (a) $A^{-1}(P-A)$ and (b) $P^{-1}A$ for different preconditioners with $a_n = (n+1)^{-1}$.

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

Figure 4: (a) Eigenvalue distribution of $P^{-1}A$ and (b) convergence history for Problem 2.

Figure 5: (a) Eigenvalue distribution of $P^{-1}A$ and (b) convergence history for Problem 3.

Figure 6: (a) Eigenvalue distribution of $P^{-1}A$ and (b) convergence history for Problem 4.

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

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

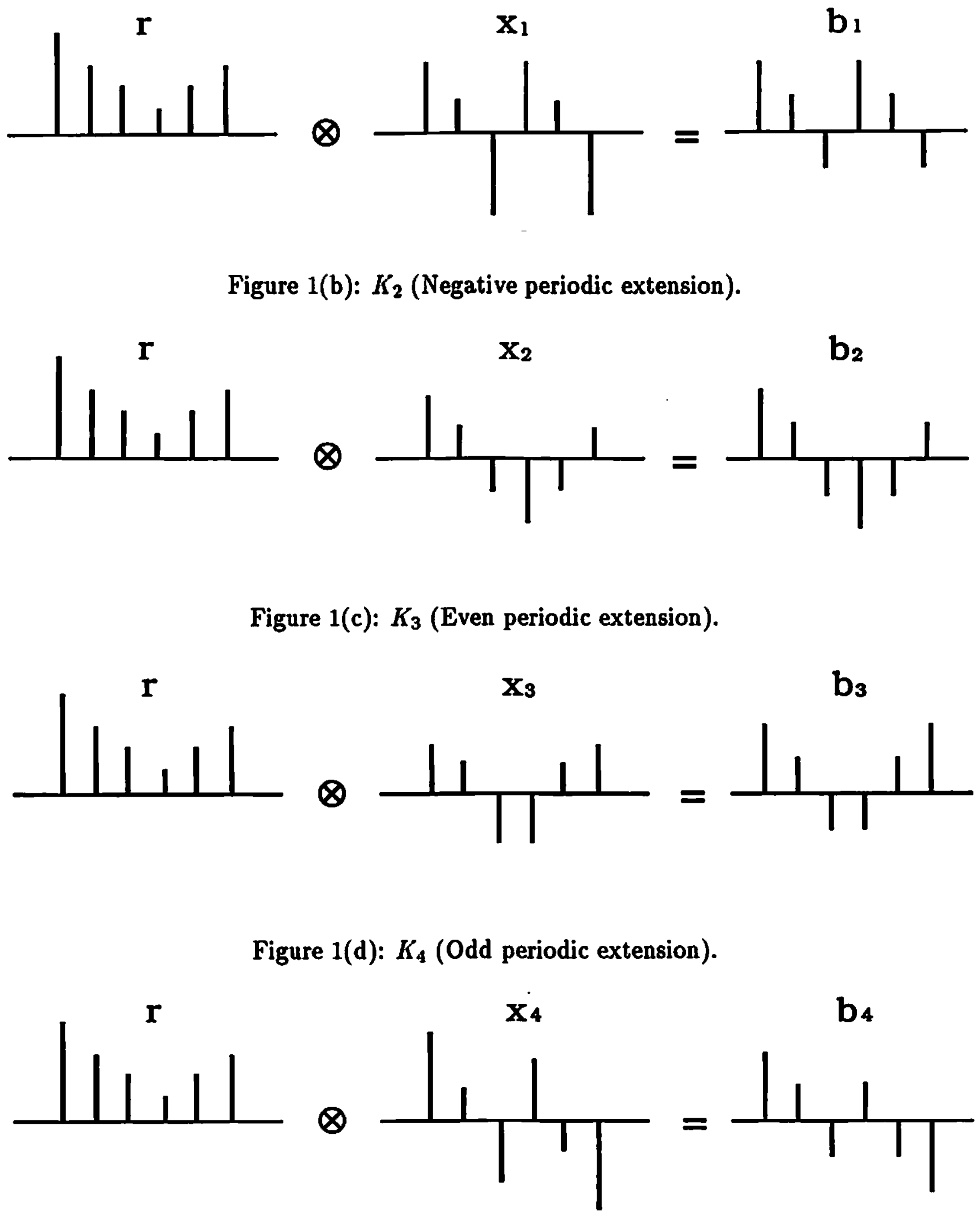

Figure 1(a): K_1 (Periodic extension).

Figure 1(b): K_2 (Negative periodic extension).

Figure 1(c): K_3 (Even periodic extension).

Figure 1(d): K_4 (Odd periodic extension).

Figure 1: Circular convolutions for preconditioners K_i.

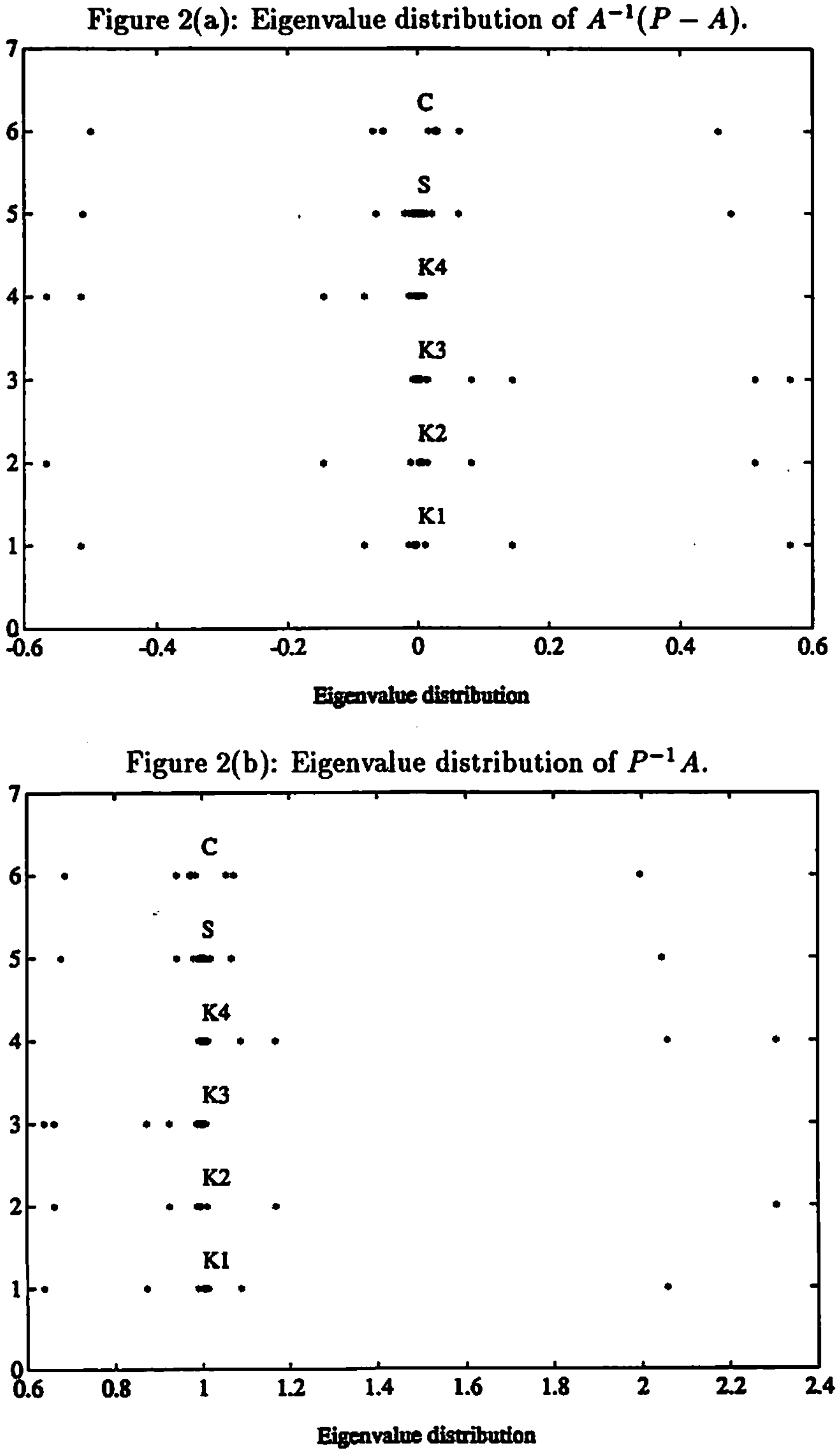
Figure 2(a): Eigenvalue distribution of $A^{-1}(P - A)$.

Figure 2(b): Eigenvalue distribution of $P^{-1}A$.

Figure 2: Eigenvalue distribution of (a) $A^{-1}(P - A)$ and (b) $P^{-1}A$ for different preconditioners with $a_n = (n+1)^{-1}$.

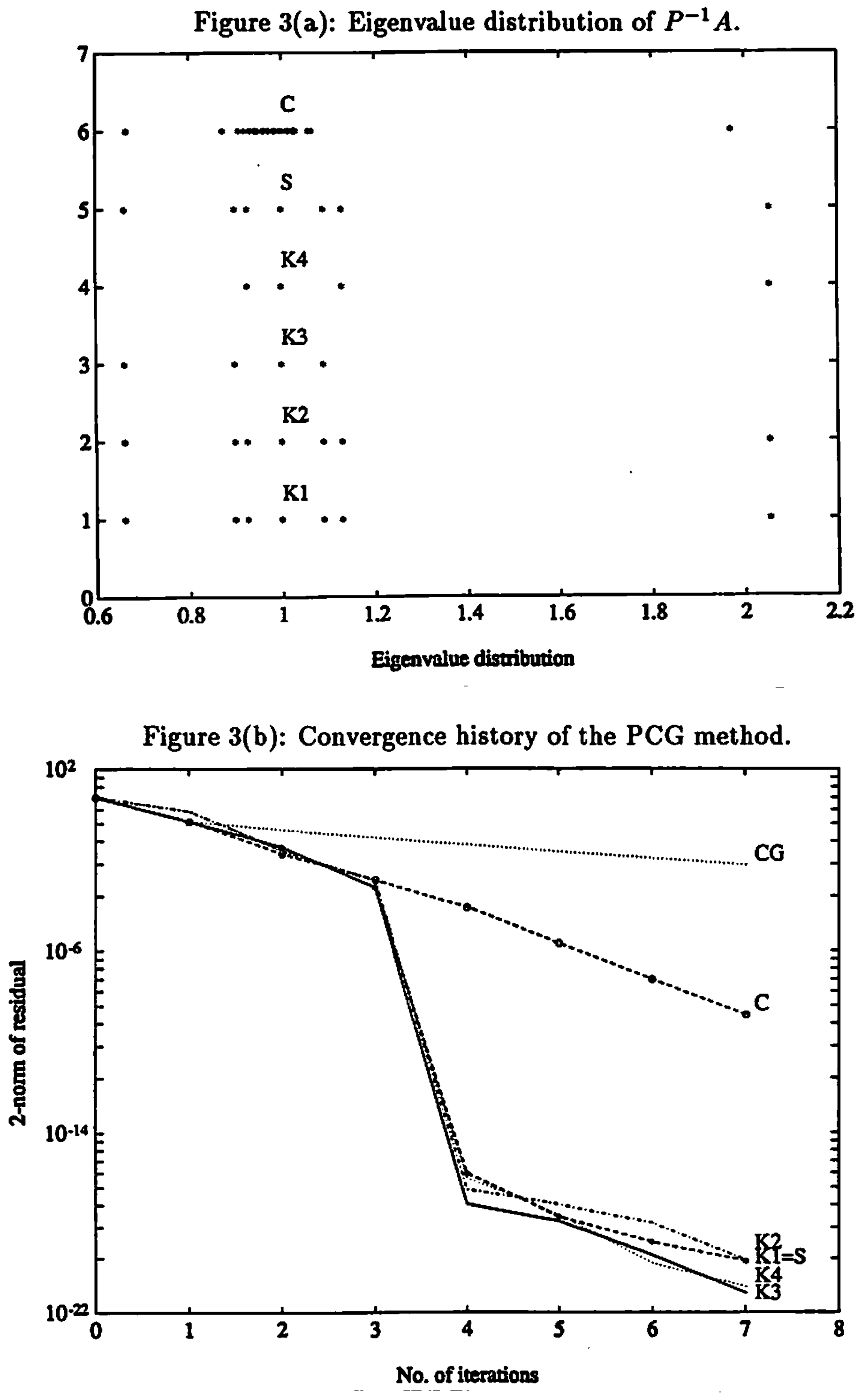

Figure 3(a): Eigenvalue distribution of $P^{-1}A$.

Figure 3(b): Convergence history of the PCG method.

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

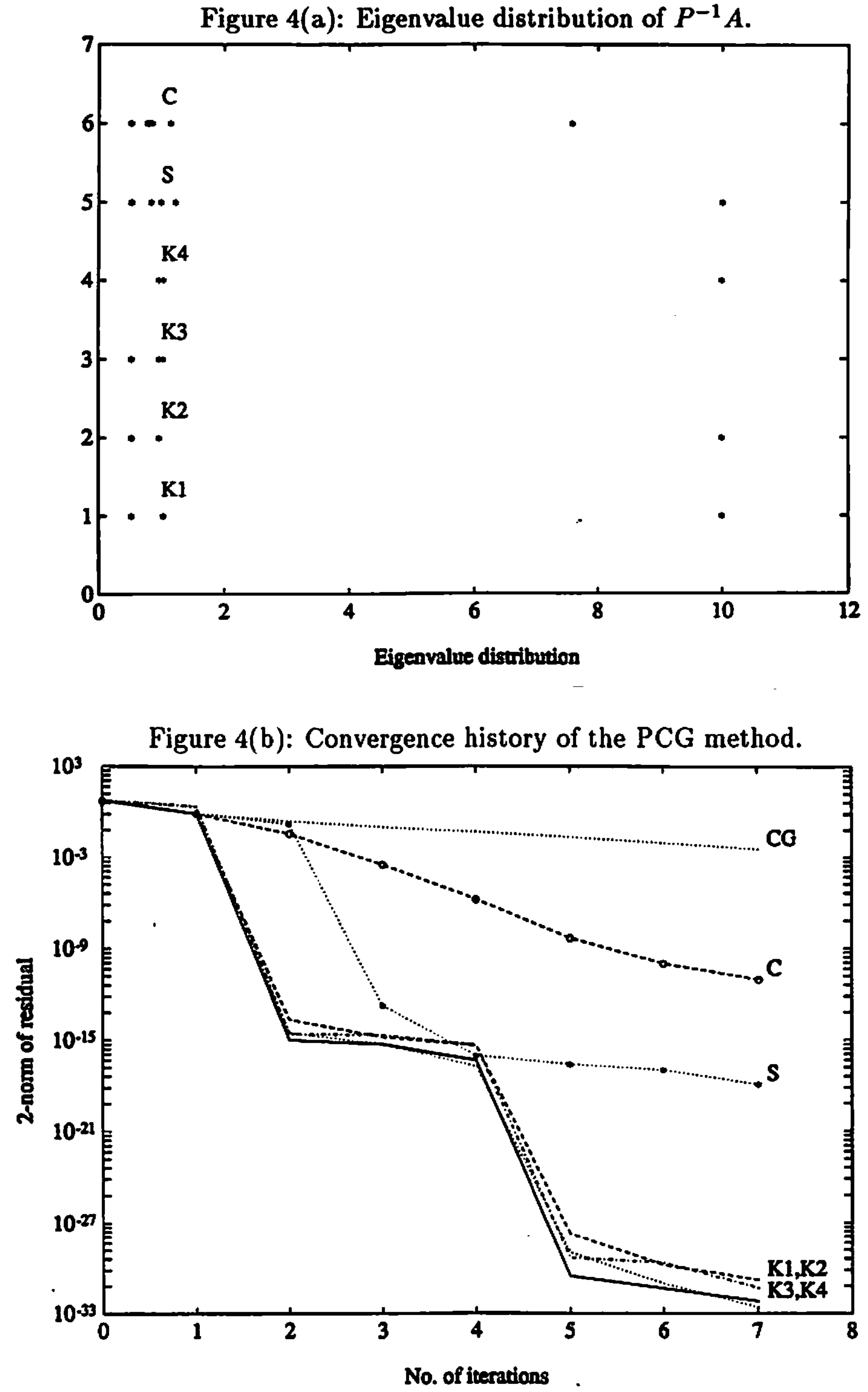

Figure 4(a): Eigenvalue distribution of $P^{-1}A$.

Figure 4(b): Convergence history of the PCG method.

Figure 4: (a) Eigenvalue distribution of $P^{-1}A$ and (b) convergence history for Problem 2.

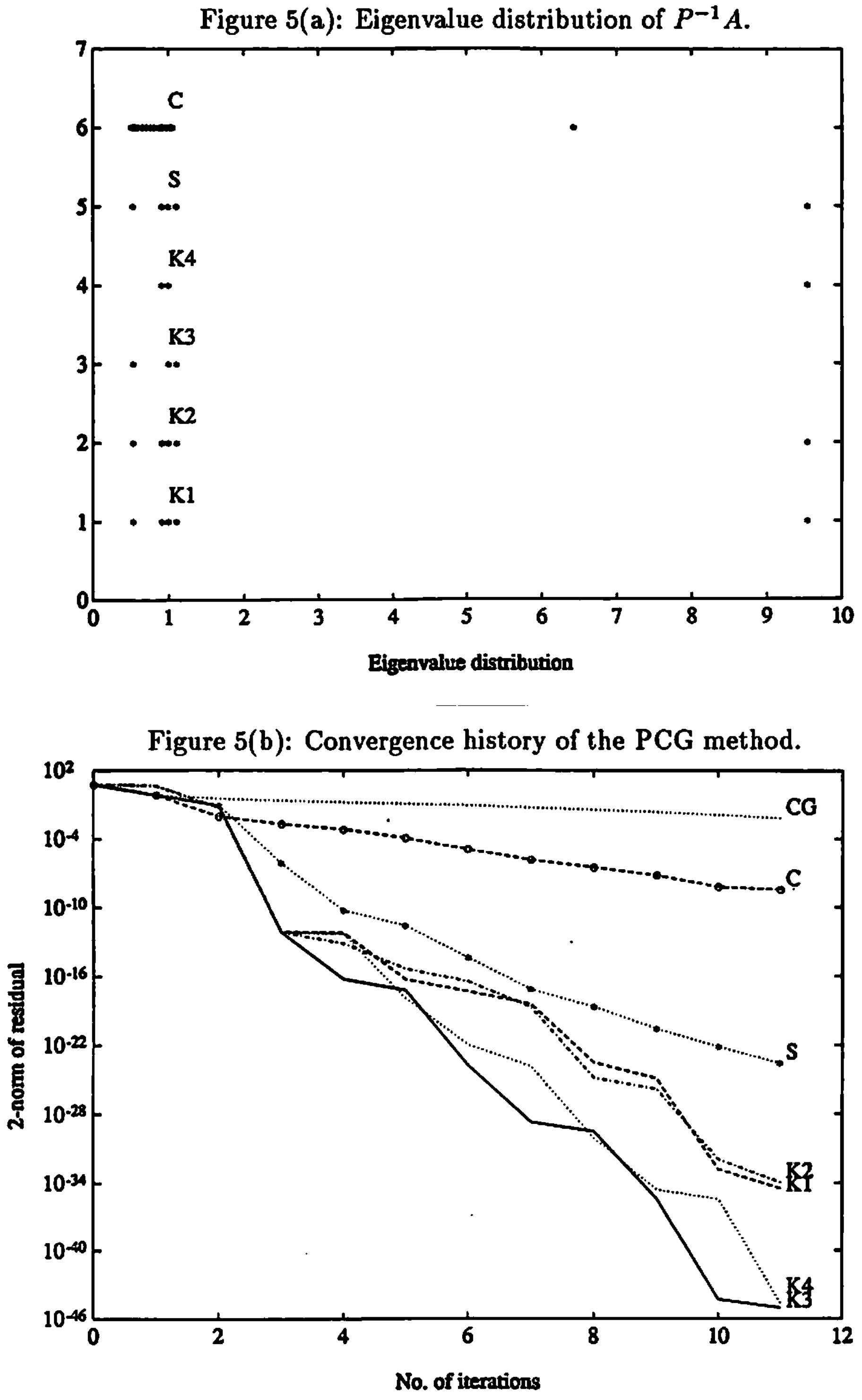

Figure 5(a): Eigenvalue distribution of $P^{-1}A$.

Figure 5(b): Convergence history of the PCG method.

Figure 5: (a) Eigenvalue distribution of $P^{-1}A$ and (b) convergence history for Problem 3.

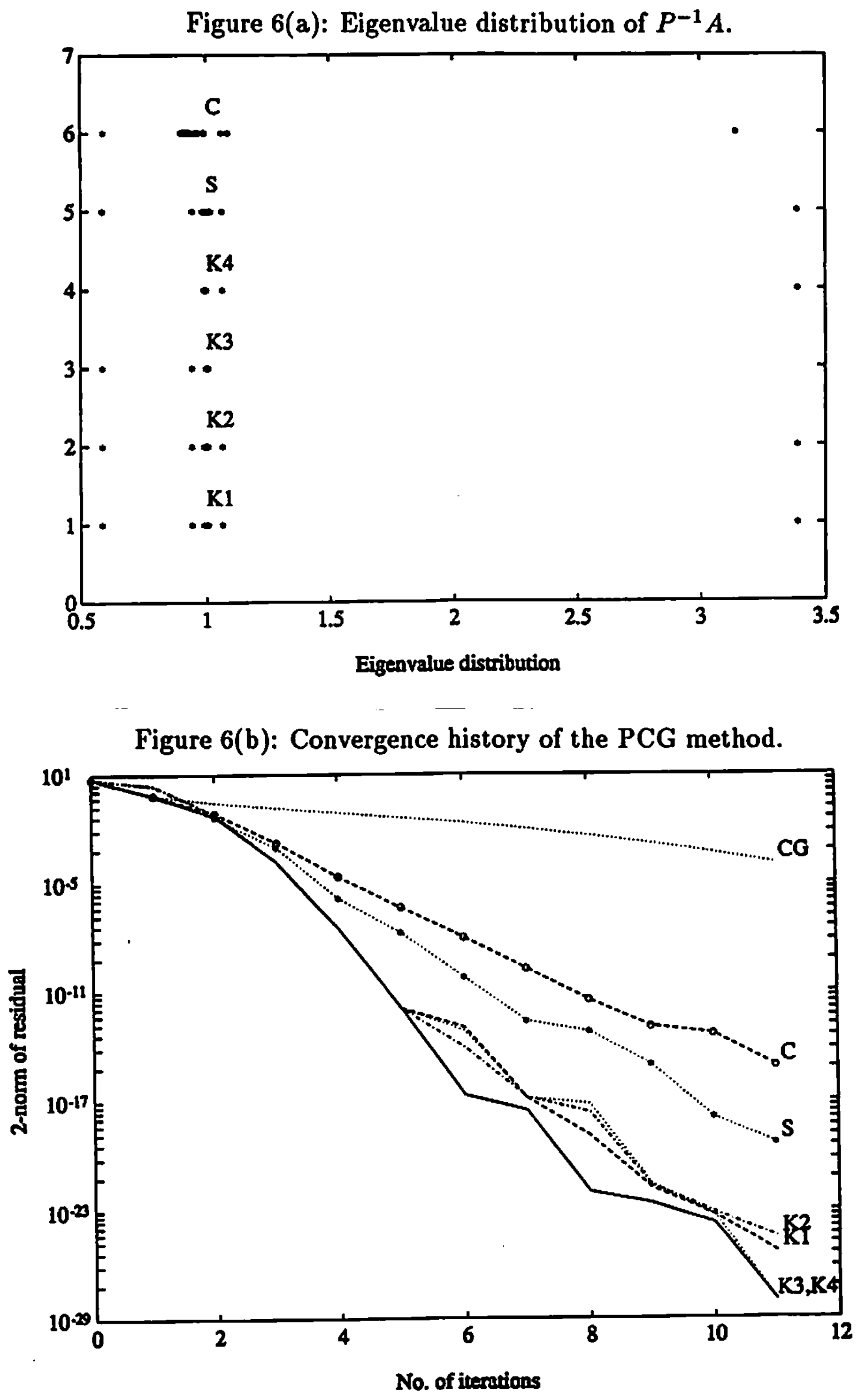

Figure 6(a): Eigenvalue distribution of $P^{-1}A$.

Figure 6(b): Convergence history of the PCG method.

Figure 6: (a) Eigenvalue distribution of $P^{-1}A$ and (b) convergence history for Problem 4.

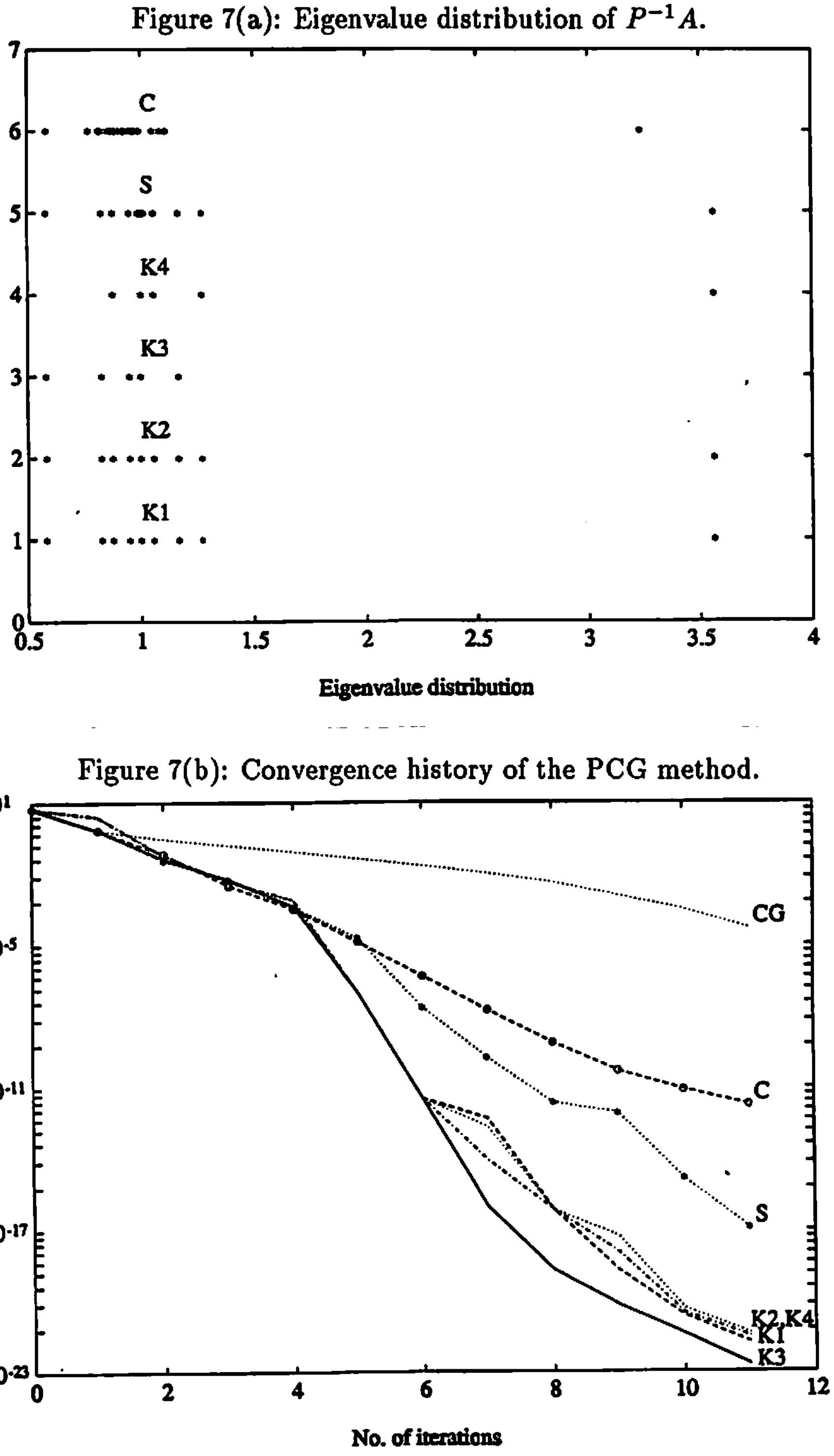

Figure 7: (a) Eigenvalue distribution of $P^{-1}A$ and (b) convergence history for Problem 5.

Figure 8(a): Eigenvalue distribution of $K_1^{-1}A$ with different N.

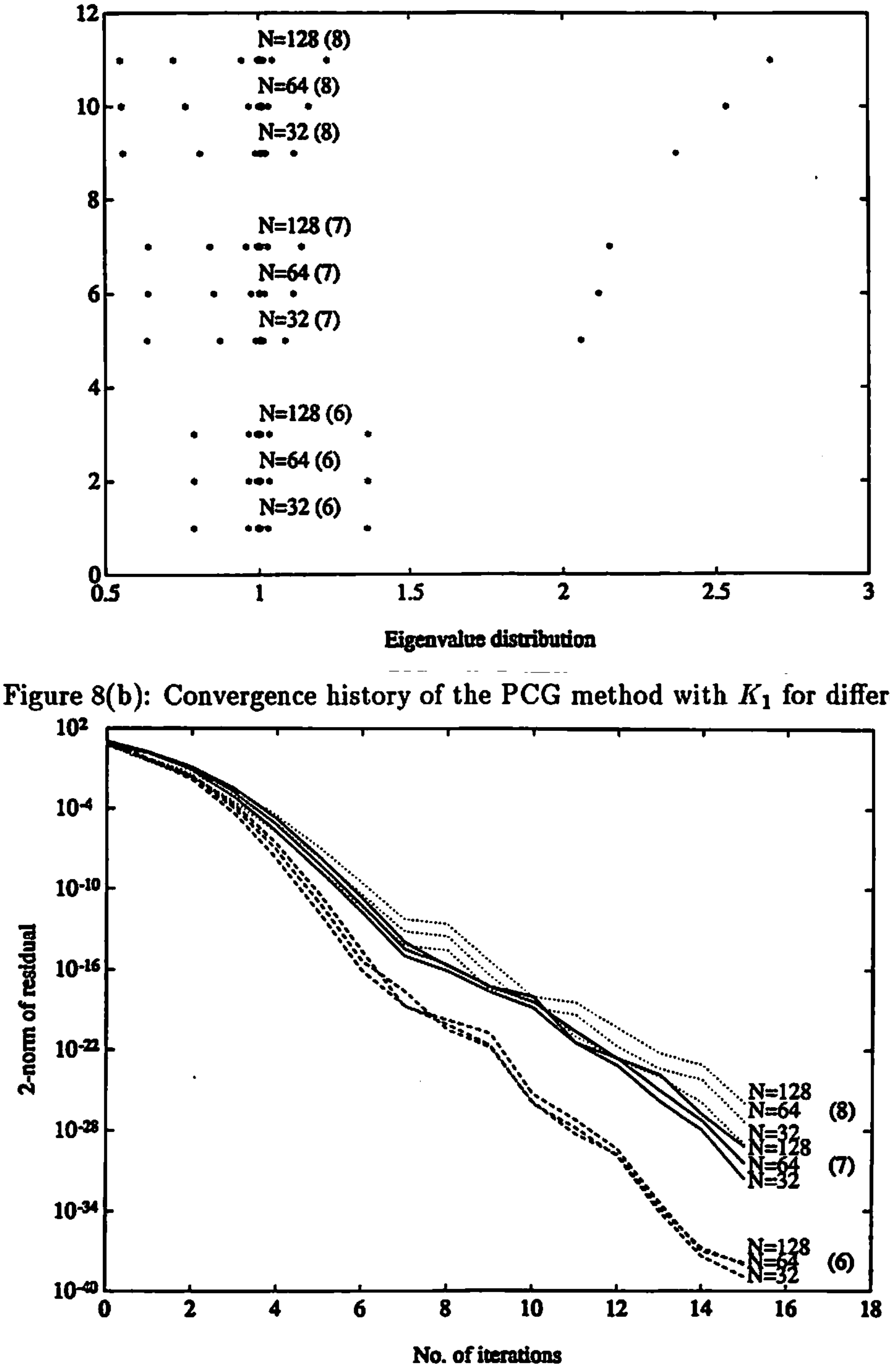

Figure 8(b): Convergence history of the PCG method with K_1 for different N.

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