FAST BAND-TOEPLITZ PRECONDITIONERS FOR HERMITIAN TOEPLITZ SYSTEMS

RAYMOND H. CHAN* AND PING TAK PETER TANG†

Abstract. We consider the solutions of Hermitian Toeplitz systems where the Toeplitz matrices are generated by nonnegative functions f. The preconditioned conjugate gradient method with well-known circulant preconditioners fails in the case when f has zeros. In this paper, we employ Toeplitz matrices of fixed band-width as preconditioners. Their generating functions g are trigonometric polynomials of fixed degree and are determined by minimizing the maximum relative error $||(f-g)/f||_\infty$. We show that the condition number of systems preconditioned by the band-Toeplitz matrices are $O(1)$ for f with or without zeros. When f is positive, our preconditioned systems converge at the same rate as other well-known circulant preconditioned systems. We also give an a priori bound of the number of iterations required for convergence.

Key words. Toeplitz matrix, generating function, preconditioned conjugate gradient method, Chebyshev approximation, Remez algorithm

AMS subject classifications. 65F10, 65F15

1. Introduction.

In this paper, we consider solutions of n-by-n Hermitian Toeplitz systems $A_n x = b$ by the preconditioned conjugate gradient method. The Toeplitz matrices A_n are assumed to be generated by 2π-periodic continuous real-valued functions f defined on $[-\pi, \pi]$, i.e. the entries of A_n are given by the Fourier coefficients of f:

$$[A_n]_{j,k} = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x) e^{-i(j-k)x} dx \quad \forall 0 \le j,k < n.$$

We emphasize that the generating function f is given in some applications of Toeplitz systems. Typical examples are the kernels of the Wiener-Hopf equations, see Gohberg and Fel'dman [12, p.82], the spectral density functions in stationary stochastic process, see Grenander and Szegö [14, p.171] and the point-spread functions in image deblurring, see Oppenheim [16, p.200].

If the generating function f is positive, the preconditioned conjugate gradient method with circulant preconditioners has proved to be a successful method – the preconditioned systems converge superlinearly when f is smooth, see for instance Chan and Strang [3] and Chan and Yeung [5]. However, these circulant preconditioners do not work in general when f has zeros. A specific example is the 1-dimensional discrete Laplacian given by the tridiagonal matrix trid$[-1, 2, -1]$. Its generating function is $f(x) = 4\sin^2 x$, which has a zero at $x = 0$. The corresponding Strang's circulant preconditioner, see [17], is actually singular. (See also the numerical results in §4 for the performance of the T. Chan [8] circulant preconditioner in the case where f has zeros.)

Recently, Chan [4] proposed using band-Toeplitz matrices $B_{n,\ell}$ as preconditioners for f that has zeros. These preconditioners are constructed by matching their generating function g with f at those zeros of f. It is proved that if the order of the zero of f is 2ℓ, then the condition number $\kappa(A_n)$ of A_n is $O(n^{2\ell})$ whereas $\kappa(B_{n,\ell}^{-1} A_n)$ is $O(1)$.

*Department of Mathematics, University of Hong Kong, Hong Kong. Research supported in part by ONR contract no. N00014-90-J-1695 and DOE grant no. DE-FG03-87ER25037.

†Mathematics and Computer Science Division, Argonne National Laboratory, Argonne, IL 60439, USA.

However, when f is positive, the band-Toeplitz preconditioned systems converge much slower than those preconditioned by circulant preconditioners.

Our main aim in this paper is to design band-Toeplitz preconditioners that work when f has zeros and yet their preconditioned systems converge at the same rate as the circulant preconditioned systems even when f is positive. Our idea is to increase the band-width of the band-Toeplitz preconditioner to get extra degrees of freedom which enable us not only to match the zeros in f but also to minimize the relative error $||(f-g)/f||_\infty$. The minimizer g is found by a version of the Remez algorithm proposed by Tang [18]. The algorithm also computes the minimum relative error which ultimately gives an a priori bound on the number of iterations required for convergence.

We note that the band-Toeplitz preconditioner we proposed has band-width ℓ that depends only on the order of the zeros of f and is independent of n, the size of the matrix. Hence for any vector x, $B_{n,\ell}^{-1}x$ can be obtained by band-solver in $O(\ell^2 n)$ operations. In contrast, solution of circulant systems requires $O(n \log n)$ operations. We remark that in [15], Ku and Kuo have considered using products of lower- and upper-triangular band-Toeplitz matrices as preconditioners. Their resulting preconditioners are in general non-Toeplitz and hence are different from ours.

The outline of the paper is as follows. In §2, we analyze the convergence rate of our preconditioned systems $B_{n,\ell}^{-1}A_n$ in terms of the generating functions g of B_n and f of A_n. In §3, we describe the Remez algorithm and how it is applied to construct the generating function g and hence the preconditioner $B_{n,\ell}$. In §4, we present numerical results that confirm our analysis in §2. In §5, we discuss the use of regularization, a technique that is relevant in computations corresponding to f having zeros and especially when n is large. Finally, concluding remarks are given in §6.

2. Convergence Analysis. In this section, we analyze the convergence rate of the preconditioned conjugate gradient method in terms of the generating functions f and g.

We first note that if f is nonnegative, then A_n is always positive definite.

LEMMA 2.1. *Let $f_{\min}$ and $f_{\max}$ be the minimum and maximum of f in $[-\pi, \pi]$. If $f_{\min} < f_{\max}$, then for all $n > 0$,*

$$f_{\min} < \lambda_i(A_n) < f_{\max}, \quad i = 1, \cdots, n,$$

where $\lambda_i(A_n)$ is the ith eigenvalue of A_n. In particular, if $f \geq 0$, then A_n are positive definite for all n.

The proof of the Lemma can be found in Chan [4]. Next we give a bound on the condition number of the preconditioned systems.

THEOREM 2.2. *Let f be the generating function of A_n and g be the generating function of a band-Toeplitz matrix $B_{n,\ell}$:*

$$g(x) = \sum_{j=-(\ell-1)}^{\ell-1} b_j e^{ijx}, \quad b_j = \bar{b}_{-j}$$

Then, if

$$||\frac{f-g}{f}||_\infty = h < 1,$$

then $B_{n,\ell}$ is positive definite and

$$\kappa(B_{n,\ell}^{-1}A_n) \leq \frac{1+h}{1-h}, \quad n = 1, 2, 3, \ldots.$$

Proof. By assumption, we have

$$f(x)(1-h) \le g(x) \le f(x)(1+h) \quad \forall x \in [-\pi, \pi].$$

Clearly, $g(x)$ is nonnegative. In particular, by Lemma 1, $B_{n,\ell}$ is positive definite for $n = 1, 2, 3, \ldots$. Since A_n and $B_{n,\ell}$ are Toeplitz matrices, we have

$$u^* A_n u = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) |\sum_{j=0}^{n-1} u_j e^{ijx}|^2 dx$$

and

$$u^* B_n u = \frac{1}{2\pi} \int_{-\pi}^{\pi} g(x) |\sum_{j=0}^{n-1} u_j e^{ijx}|^2 dx$$

for an arbitrary (complex) n-vector $u^* = [u_0, u_1, \cdots, u_{n-1}]$, see Grenander and Szegö [14]. Hence, we get

$$(1-h) u^* A_n u \le u^* B_n u \le (1+h) u^* A_n u. \tag{1}$$

Since $B_{n,\ell}$ is also positive definite, we finally have

$$\kappa(B_{n,\ell}^{-1} A_n) \le \frac{1+h}{1-h}. \qquad \square$$

By standard error analysis of the conjugate gradient method, see, for instance, Axelsson and Barker [1, p.14], we conclude that the number of iterations required for convergence is bounded by

$$\frac{1}{2} \left(\frac{1+h}{1-h} \right) \log \left(\frac{1}{\tau} \right) + 1,$$

where τ is the tolerance. Since h is given explicitly in the Remez algorithm, we have an a priori bound on the number of iterations required for convergence.

3. Construction of Preconditioner by the Remez Algorithm. In this section, we describe how we construct our band-Toeplitz preconditioner. For the ease of presentation, we assume the matrix A_n is real symmetric. Handling the complex Hermitian case will become clear once the symmetric case is explained. Since A_n is symmetric, its generating function f is even, in additional to being 2π-periodic and continuous. It suffices to consider f on $[0, \pi]$. The generating function of a band-Toeplitz matrix $B_{n,\ell}$,

$$B_{n,\ell} = \begin{pmatrix} b_0 & b_1 & \cdots & b_{\ell-1} & & & & \\ b_1 & \ddots & \ddots & & \ddots & & & \\ \vdots & \ddots & \ddots & \ddots & & \ddots & & \\ b_{\ell-1} & & \ddots & \ddots & \ddots & & \ddots & \\ & \ddots & & \ddots & \ddots & \ddots & & b_{\ell-1} \\ & & \ddots & & \ddots & \ddots & \ddots & \vdots \\ & & & \ddots & & \ddots & \ddots & b_1 \\ & & & & b_{\ell-1} & \cdots & b_1 & b_0 \end{pmatrix}$$

is given by

$$g(x) = b_0 + b_1(2\cos(x)) + b_2(2\cos(2x)) + \ldots + b_{\ell-1}(2\cos((\ell-1)x)).$$

Thus, in the case where $f > 0$ on $[0, \pi]$, determining the optimal P is a standard linear minimax approximation problem:

$$\text{minimize}_{p_0,p_1,\ldots,p_{\ell-1}} \left\| 1 - P(x) \right\|_\infty ,$$

where

$$P(x) = \sum_{j=0}^{\ell-1} p_j \phi_j(x), \quad \phi_0 = 1/f(x) \text{ and } \phi_j(x) = 2\cos(jx)/f(x) \quad \text{for } j > 0.$$

Note that $P = g/f$. This optimal P (and hence g) can be obtained by a standard Remez algorithm (see Cheney [10], for example). We, however, use the version proposed by Tang [18] which can be extended to handle the case when $f(x_0) = 0$ for some $x_0 \in [0, \pi]$. We now describe this version of the Remez algorithm briefly; after that, the extension will also be explained.

Given $\phi_0(x) = 0$, and $\phi_j(x) = 2\cos(jx)/f(x)$, $j = 1, \ldots, \ell - 1$, we are to solve

Problem $\mathcal{P}$:

$$\text{Minimize } h$$

subject to

$$h \geq s \left(1 - \sum_{j=0}^{\ell-1} p_j \phi_j(x) \right), \quad (s, x) \in \{-1, 1\} \times [0, \pi].$$

One can think of Problem $\mathcal{P}$ as a linear programming problem (by, say, replacing $[0, \pi]$ by a finite set of points). The dual of this problem is given by the following.

Problem $\mathcal{D}$:

$$\text{Maximize } \sum_{s,x} s \cdot r_{s,x}$$

subject to

$$r_{s,x} \geq 0, \quad \text{and}$$

$$\sum_{s,x} r_{s,x} \phi_j(x) s = 0, \quad j = 0, 1, \ldots, \ell - 1.$$

It is observed in [18] that even without discretizing the domain $[0, \pi]$, the Simplex algorithm can be applied to Problem $\mathcal{D}$. The preconditioners in the next section are obtained by this computation.

Now, suppose that $f(x_0) = 0$. In practice, x_0 is often known. Because $f \geq 0$ (lest A_n has negative eigenvalues for large enough n), we have $f'(x_0) = 0$ also. Suppose $f''(x_0) \neq 0$, then we would determine P by imposing the constraint $g(x_0) = 0$, that is,

$$p_0 + 2 \sum_{j=1}^{\ell-1} p_j \cos(jx_0) = 0.$$

This linear constraint on the coefficients p_j's can be naturally added to Problem $\mathcal{P}$ and translated to its dual form in Problem $\mathcal{D}$. In general, the case when $f^{(k)}(x_0) = 0$ for $k = 0, 1, \ldots, m$ can be handled by the constraints $g^{(k)}(x_0) = 0$ for $k = 0, 1, \ldots, m-1$.

We note that when A_n is complex Hermitian, f will not be even necessarily (but still continuous real-valued and 2π-periodic). The domain of approximation becomes $[-\pi, \pi]$ and the approximant will be trigonometric polynomials with sin and cos. Similar constraints can be imposed when $f(x_0) = 0$ for some x_0.

Let us end the section by discussing the computational cost of our method. As pointed out in [18], the number of Simplex iterations needed to determine g is proportional to ℓ. In practice, all of our experiments took less than 2ℓ iterations. Moreover, after an initial LU decomposition of an $\ell \times \ell$ matrix, each Simplex iteration requires only a modification to the decomposition after a rank-1 change. The total effort for location g is $O(\ell^3)$. We stress the fact that g is independent of n. Thus, as long as f is fixed and a sufficient band width ℓ is reached, the entries for $B_{n,\ell}$ are determined for all n.

In each iteration of the preconditioned conjugate gradient method, we have to compute matrix vector multiplications of the form $A_n x$ and $B_{n,\ell}^{-1} y$. We note that $A_n x$ can be computed in $O(n \log n)$ operations by first embedding A_n into a $2n$-by-$2n$ circulant matrix and then perform the multiplication by the Fast Fourier Transform (see Strang [17]). The vector $z = B_{n,\ell}^{-1} y$ can be obtained by solving the banded system $B_{n,\ell} z = y$ with any band solvers, see for instance Golub and Van Loan [13], or Wright [19] for a parallel one. Typically, we will decompose $B_{n,\ell}$ into some triangular factors and then solve the system by a backward and forward solve. The cost of obtaining the triangular factors is $O(\ell^2 n)$, and each subsequent solve will cost $O(\ell n)$, as the triangular factors will also be banded.

Recall that the number of iterations is independent of the size of the matrix n, we therefore conclude that the total complexity of our method is $O(n \log n + \ell^2 n)$.

4. Numerical Results. In this section, we compare the convergence rate of the band-Toeplitz preconditioner with circulant preconditioner on five different generating functions. They are $\cosh x$, $x^4 + 1$, $1 - e^{-x^2}$, $(x-1)^2(x+1)^2$ and x^4. The first two functions are positive while the others have either one or two distinct zeros. The matrices A_n are formed by evaluating the Fourier coefficients of the generating functions.

We note that when $f(x) = 1 - e^{-x^2}$, its Fourier coefficients cannot be evaluated exactly. In this case, we approximate them by

$$\begin{aligned} a_j &= \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) e^{-ijx} dx \\ &\approx \frac{1}{2n} \sum_{k=0}^{2n-1} f(\frac{k\pi}{n} - \pi) e^{-ij(k\pi/n - \pi)}, \quad j = 0, \pm 1, \pm 2, \cdots \end{aligned}$$

where the last expression is evaluated by using the Fast Fourier Transform.

In our tests, the vector of all ones is the right hand side vector, the zero vector is the initial guess and the stopping criterion is $||r_q||_2 / ||r_0||_2 \leq 10^{-7}$, where r_q is the residual vector after q iterations. All computations are done by Matlab on a Sun workstation. Tables 1-5 show the numbers of iterations required for convergence with different choices of preconditioners. In the tables, I denotes no preconditioner is used, C is the T. Chan circulant preconditioner [8], and $B_{n,\ell}$ is the band-Toeplitz preconditioner with half-bandwidth ℓ.

We note that for the cases when the f's are positive, our preconditioners, with half-bandwidths 4 to 5, work as well as the circulant preconditioners. In the cases when the f's have zeros, our preconditioned systems still converge at a rate that is independent of the sizes of the matrices. For the circulant preconditioned systems, however, the numbers of iterations required grow as the sizes of the matrices increase.

TABLE 1
Numbers of Iterations for $f(x) = \cosh x$.

n	I	C	$B_{n,2}$	$B_{n,3}$	$B_{n,4}$	$B_{n,5}$
16	9	6	9	7	6	5
32	16	6	10	7	6	6
64	21	5	11	8	6	6
128	23	5	10	8	6	6
256	24	5	10	7	6	6

TABLE 2
Numbers of Iterations for $f(x) = x^4 + 1$.

n	I	C	$B_{n,2}$	$B_{n,3}$	$B_{n,4}$	$B_{n,5}$
16	10	9	9	8	8	7
32	22	7	16	11	8	7
64	37	7	22	12	8	7
128	56	6	25	12	8	7
256	67	6	26	12	8	7

TABLE 3
Numbers of Iterations for $f(x) = 1 - e^{-x^2}$.

n	I	C	$B_{n,2}$	$B_{n,3}$	$B_{n,4}$	$B_{n,5}$
16	9	6	9	7	4	3
32	14	7	15	7	5	3
64	24	8	17	8	5	3
128	42	10	17	8	5	3
256	77	13	17	8	5	3

5. Regularization. We note that when f has zeros, the system of equations $A_n x = b$ will become very ill-conditioned when n is large. Thus the usefulness of the solution x can be in doubt even though we can solve for it quickly using our preconditioners. In this case, one can employ the technique of regularization to alleviate the problem. One approach is to solve the appended least squares problem:

$$\min_x || \begin{bmatrix} A_n \\ \mu P_n \end{bmatrix} x - \begin{bmatrix} b \\ 0 \end{bmatrix} ||_2.$$

Here μ is the regularization parameter and the n-by-n matrix P_n is the regularization operator that tries to smooth the solution x to a certain degree. Choosing P_n as the $2k$-th difference operator will force the solution to have a small $2k$-th derivative. We note that the corresponding P_n will be a banded Hermitian matrix with half-bandwidth $k+1$. Typical choices of P_n are the n-by-n identity matrix and the 1-dimensional discrete Laplacian matrix. Choosing the regularization parameter μ on the other hand is usually not a trivial problem. One may need to solve the least squares problem for several values of μ to determine the best one, see Eldén [11].

TABLE 4
Numbers of Iterations for $f(x) = (x-1)^2(x+1)^2$.

n	I	C	$B_{n,3}$	$B_{n,4}$	$B_{n,5}$	$B_{n,6}$
16	11	9	9	9	8	7
32	27	14	13	11	9	7
64	74	17	16	11	8	7
128	193	22	18	11	8	7
256	465	28	19	11	8	7

TABLE 5
Numbers of Iterations for $f(x) = x^4$.

n	I	C	$B_{n,3}$	$B_{n,4}$	$B_{n,5}$	$B_{n,6}$
16	12	10	9	9	9	7
32	34	16	15	10	11	9
64	119	26	21	13	11	9
128	587	77	24	15	12	10
256	> 1000	179	27	16	12	10

The solution to the least squares problem can be obtained by solving the normal equation:

$$(A_n^2 + \mu^2 P_n^2)x = A_n b.$$

An obvious choice of preconditioners for the normal equation is the band matrix $B_{n,\ell}^2 + \mu^2 P_n^2$. Its half-bandwidth is $\max(2\ell - 1, 2k + 1)$. By using (1), we can easily show that

$$\kappa\left\{(B_{n,\ell}^2 + \mu^2 P_n^2)^{-1}(A_n^2 + \mu^2 P_n^2)\right\} \leq \left(\frac{1+h}{1-h}\right)^2.$$

In contrast, even if C_n is a good circulant preconditioner for A_n, the matrix $C_n^2 + \mu^2 P_n^2$ will no longer be circulant. However, we remark that regularization techniques using other circulant preconditioners have been considered in Chan, Nagy and Plemmons [7].

6. Concluding Remarks. By understanding Toeplitz preconditioner from the point of view of minimax approximation of the corresponding generating functions, we can construct band-Toeplitz preconditioners that offer fast convergence rates even when the matrix to be preconditioned has a generating function with a zero. Moreover, our preconditioner with modest bandwidth is also an excellent choice for f without a zero. We emphasize that for a given f, the entries of the preconditioners are unchanged as n increases. Thus, we need to invoke the Remez algorithm once for each f. We note moreover that the Cholesky factors of $B_{n,\ell}$ can be used to build the Cholesky factors of $B_{n+1,\ell}$. That can reduce the cost of factorization of the band-Toeplitz preconditioner. We finally remark that our preconditioner can also be adapted easily to give a good preconditioner for Toeplitz-plus-band systems of the form $(A_n + D_n)x = b$, where D_n is an arbitrary band matrix. Toeplitz-plus-band systems appear in solving Fredholm integral-differential equations, see Delves and Mohamed [9, p.343] and also in signal processing literature, see Carayannis et. al. [2].

For such systems, direct Toeplitz solvers and the preconditioned conjugate gradient method with circulant preconditioners will not work. However, one can use $B_{n,\ell} + D_n$ as a preconditioner and use the proof mentioned in Chan and Ng [6] to

derive basically the same result that we have in Theorem 1, namely that

$$\kappa\left\{(B_{n,\ell}+D_n)^{-1}(A_n+D_n)\right\} \le \frac{1+h}{1-h}.$$

Hence the number of iterations required for convergence is still fixed independent of the size of the matrices. Since $B_{n,\ell}+D_n$ is a band matrix, the system $(B_{n,\ell}+D_n)x = y$ can still be solved efficiently by band solvers for any vector y.

REFERENCES

[1] O. Axelsson and V. Barker, *Finite Element Solution of Boundary Value Problems, Theory and Computation*, Academic Press Inc., New York, 1984.

[2] G. Carayannis, N. Kalouptsidis and D. Manolakis, *Fast Recursive Algorithms for a Class of Linear Equations*, IEEE Trans. Acoust. Speech Signal Process., 30 (1982), pp. 227-239.

[3] R. Chan and G. Strang, *Toeplitz Equations by Conjugate Gradients with Circulant Preconditioner*, SIAM J. Sci. Statst. Comput., 10 (1989), pp. 104-119.

[4] R. Chan, *Toeplitz Preconditioners for Toeplitz Systems with Nonnegative Generating Functions*, IMA J. Numer. Anal., 11 (1991), pp. 333-345.

[5] R. Chan and M. Yeung, *Circulant Preconditioners for Toeplitz Matrices with Positive Continuous Generating Functions*, Math. Comp., 58 (1992), pp. 233-240.

[6] R. Chan and M. Ng, *Fast Iterative Solvers for Toeplitz-plus-Band Systems*, SIAM J. Sci. Statist. Comput., to appear.

[7] R. Chan, J. Nagy and R. Plemmons, *Circulant Preconditioned Toeplitz Least Squares Iterations*, SIAM J. Matrix Anal. Appl., to appear.

[8] T. Chan, *An Optimal Circulant Preconditioner for Toeplitz Systems*, SIAM J. Sci. Statist. Comput., 9 (1988), pp. 766-771.

[9] L. Delves and J. Mohamed, *Computational Methods for Integral Equations,* Cambridge University Press, Cambridge, 1985.

[10] E. Cheney, *Introduction to Approximation Theory*. Chelsea, New York, 1986.

[11] L. Eldén, *An Algorithm for the Regularization of Ill-conditioned, Banded Least Squares Problems*, SIAM J. Sci. Statist. Comput., 5 (1984), pp. 237-254.

[12] I. Gohberg and I. Fel'dman, *Convolution Equations and Projection Methods for Their Solutions*, Volume 41, Transaction of Mathematical Monographs, American Mathematical Society, Rhode Island, 1974.

[13] G. Golub and C. Van Loan, *Matrix Computations*, 2nd Ed., The Johns Hopkins University Press, Maryland, 1989.

[14] U. Grenander and G. Szegö, *Toeplitz Form and Its Applications*, 2nd Ed., Chelsea Pub. Co., New York, 1984.

[15] K. Ku and C. Kuo, *Minimum-Phase LU Factorization Preconditioners for Toeplitz Matrices*, SIAM J. Sci. Stat. Comput., to appear.

[16] A. Oppenheim, *Applications of Digital Signal Processing*, Prentice-Hall, Englewood Cliffs, New Jersey, 1978.

[17] G. Strang, *A Proposal for Toeplitz Matrix Calculations*, Studies in App. Math., 74 (1986), pp. 171–176.

[18] P. Tang, *A Fast Algorithm for Linear Complex Chebyshev Approximation*, Math. Comp., 51 (1988), pp. 721–739.

[19] S. J. Wright, *Parallel Algorithms for Banded Linear Systems*, SIAM J. Sci. Statist. Comput., 12 (1991), pp. 824–842.