Electronic Transactions on Numerical Analysis.
Volume 2, pp. 154-170, December 1994.
Copyright © 1994, Kent State University.
ISSN 1068-9613.

ETNA
Kent State University
etna@mcs.kent.edu

FAST ITERATIVE METHODS FOR SOLVING TOEPLITZ-PLUS-HANKEL LEAST SQUARES PROBLEMS*

MICHAEL K. NG †

Abstract. In this paper, we consider the impulse responses of the linear-phase filter whose characteristics are determined on the basis of an observed time series, not on a prior specification. The impulse responses can be found by solving a least squares problem $\min \|\mathbf{d} - (X_1 + X_2)\mathbf{w}\|_2$ by the fast Fourier transform (FFT) based preconditioned conjugate gradient method, for $(M+2n-1)$-by-n real Toeplitz-plus-Hankel data matrices $X_1 + X_2$ with full column rank. The FFT–based preconditioners are derived from the spectral properties of the given input stochastic process, and their eigenvalues are constructed by the Blackman-Tukey spectral estimator with Bartlett window which is commonly used in signal processing. When the stochastic process is stationary and when its spectral density function is positive and differentiable, we prove that with probability 1, the spectra of the preconditioned normal equations matrices are clustered around 1, provided that large data samples are taken. Hence if the smallest singular value of $X_1 + X_2$ is of order $O(n^\alpha)$, $\alpha > 0$, then the method converges in at most $O((2\alpha+1)\log n+1)$ steps. Since the cost of forming the normal equations and the FFT–based preconditioner is $O(M \log n)$ operations and each iteration requires $O(n \log n)$ operations, the total complexity of our algorithm is of order $O(M \log n + (2\alpha + 1)n \log^2 n + n \log n)$ operations. Finally, numerical results are reported to illustrate the effectiveness of our FFT–based preconditioned iterations.

Key words. least squares estimations, linear-phase filter, Toeplitz-plus-Hankel matrix, circulant matrix, preconditioned conjugate gradient method, fast Fourier transform.

AMS subject classifications. 65F10, 65F15, 43E10.

1. Introduction.

The conjugate gradient (CG) method is an iterative method for solving symmetric positive definite systems $A\mathbf{w} = \mathbf{d}$; see for instance Golub and van Loan [14, pp. 362-374]. When A is a rectangular matrix with full column rank, one can still use the method to find the solution to the least squares problem

$$\min \|\mathbf{d} - A\mathbf{w}\|_2 \tag{1.1}$$

where $\|\cdot\|_2$ denotes the usual Euclidean norm. This can be done by applying the method to the normal equations

$$A^T A\mathbf{w} = A^T\mathbf{d}. \tag{1.2}$$

The convergence rate of the method depends on the eigenvalues of the normal equations matrix $A^T A$; see Axelsson and Barker [1, pp. 24-28]. If the eigenvalues of $A^T A$ cluster around a fixed point, convergence will be rapid. Thus, to make the algorithm a useful iterative method, one usually preconditions the system. That means, instead of solving the original system (1.2), we solve the preconditioned system

$$P^{-1}A^T A\mathbf{w} = P^{-1}A^T\mathbf{d}$$

with preconditioner P. In this paper, we apply the preconditioned conjugate gradient (PCG) method to solve structured least squares problems arising from signal processing applications, where the data matrix A is a rectangular Toeplitz-plus-Hankel matrix with full column rank. A matrix $T = (t_{jk})$ is said to be Toeplitz if $t_{jk} = t_{j-k}$, i.e., T is constant along its diagonals. A matrix $H = (h_{jk})$ is said to be Hankel if $h_{jk} = h_{j+k}$.

* Received March 25, 1994. Accepted for publication November 8, 1994. Communicated by R. J. Plemmons

† Department of Mathematics, The Chinese University of Hong Kong, Shatin, Hong Kong.

ETNA
Kent State University
etna@mcs.kent.edu

1.1. Linear-phase Filtering. Least squares estimations have been used extensively in a wide variety of applications in signal processing, as for instance spectrum analysis [11, 19], system identifications [20], equalizations [13] and speech processing [15, p. 49]. In these applications, one usually uses filters to estimate the transmitted signal from a sequence of received signal samples or to model an unknown system. One important class of filters commonly used in signal processing is the class of *finite impulse response* (FIR) *linear-phase* filters. Such filters are especially important for applications where frequency dispersion, due to nonlinear phase, is harmful, as for example in speech processing.

In this paper, we develop an impulse response vector $\mathbf{w}$ of the linear-phase filter whose characteristics are determined on the basis of an observed time series, and not on a priori specification. It was shown in [16, 21] that given M real data samples $\{x(1), x(2), \ldots, x(M)\}$ and a desired response vector $\mathbf{d}$, the impulse responses can be found by solving the Toeplitz-plus-Hankel least squares problem

$$\min \|\mathbf{d} - (X_1 + X_2)\mathbf{w}\|_2. \tag{1.3}$$

Here X_1 is an $(M + 2n - 1)$-by-n rectangular Toeplitz matrix with its first row and column given by

$$[x(1), 0, \ldots, 0] \quad \text{and} \quad [x(1), x(2), \ldots, x(M), 0, \ldots, 0]^T.$$

respectively. Moreover, X_2 is an $(M + 2n - 1)$-by-n rectangular Hankel matrix with its last column given by

$$[0, \ldots, 0, x(1), x(2), \ldots, x(M), 0, \ldots, 0]^T$$

and a zero vector as its first row.

1.2. Outline. The use of conjugate gradient methods with circulant preconditioners for solving n-by-n Toeplitz systems $T_n\mathbf{z} = \mathbf{v}$ has been studied extensively in recent years; see [6], [8], [9] and [10]. Since circulant matrix can always be diagonalized by the discrete Fourier matrix, an n-by-n linear system with circulant coefficient matrix can be solved in $O(n \log n)$ operations, using fast Fourier transform (FFT). Also, matrix-vector multiplications $T_n\mathbf{u}$ can be computed by using FFT in $O(n \log n)$ operations, by first decomposing T_n into a sum of circulant and skew-circulant matrices; see Chan and Ng [6]. It follows that the number of operations per iteration of the preconditioned conjugate method is of order $O(n \log n)$ operations.

In the practical applications, one always assumes that the data matrix $X_1 + X_2$ is of full column rank. Therefore, the normal equations matrix $(X_1 + X_2)^T(X_1 + X_2)$ is non-singular and positive definite, and the solution $\mathbf{w}$ of (1.3) is obtained by solving the normal equations

$$(X_1 + X_2)^T(X_1 + X_2)\mathbf{w} = (X_1 + X_2)^T\mathbf{d}. \tag{1.4}$$

Note that the normal equations matrix $(X_1 + X_2)^T(X_1 + X_2)$ is an n-by-n Toeplitz-plus-Hankel matrix $T_n + H_n$. By transforming the Hankel matrix H_n to a Toeplitz matrix using the reversal matrix J_n, the Hankel matrix-vector products $H_n\mathbf{u}$ can be computed by using FFT in $O(n \log n)$ operations.

In this paper, we apply the preconditioned conjugate gradient algorithm with circulant (FFT–based) preconditioners to solve the normal equations $(T_n + H_n)\mathbf{w} = (X_1 + X_2)^T\mathbf{d}$. The main result of the paper is that, under some practical signal

ETNA
Kent State University
etna@mcs.kent.edu

processing assumptions, the spectrum of the Hankel matrix H_n is clustered around zero with probability 1. The contribution of the term H_n is not significant as far as the conjugate gradient method is concerned, and we therefore do not approximate it by a circulant matrix. Thus, the preconditioner $c(T_n)$ is just defined to be the minimizer of $\|Q_n - T_n\|_F$ over all n-by-n circulant matrices Q_n. Here $\|\cdot\|_F$ denotes the Frobenius norm. As T_n is a Toeplitz matrix, the circulant preconditioner $c(T_n)$ can be found in $O(n \log n)$ operations. We also show that the eigenvalues of $c(T_n)$ can be derived from the *Blackman-Tukey* spectral estimator with the *Bartlett window* that is a commonly used non-parametric spectral estimation method in signal processing.

As for the convergence rate of the method, we prove that if the stochastic process $\{x(i)\}$ is stationary and its underlying spectral density function is $(\ell + 1)$-times differentiable function for $\ell > 0$, then the spectra of the preconditioned matrices $c(T_n)^{-1}(T_n + H_n)$ are clustered 1 with probability 1. If the smallest singular value of $X_1 + X_2$ is of order $O(n^\alpha)$ with $\alpha > 0$, the method converges in at most $O((2\alpha + 1)\log n + 1)$ steps with probability 1. Since the data matrices X_1 and X_2 are Toeplitz and Hankel matrices respectively, the normal equations and the circulant preconditioner can be formed in $O(M \log n)$ operations; see Ng and Chan [23]. Once they are formed, the cost per iteration of the preconditioned conjugate gradient method is of order $O(n \log n)$ operations, as only Toeplitz, Hankel and circulant matrix-vectors multiplications are required in each iteration. Therefore, the total work of obtaining the impulse responses to a given accuracy is of order $O(M \log n + (2\alpha + 1)n \log^2 n + n \log n)$.

The outline of the paper is as follows. In Section 2, we study some properties of the normal equations matrices and introduce our FFT–based preconditioners. In Section 3, we analyze the convergence rate of the method probabilistically. In Section 4, numerical experiments are discribed which illustrate the effectiveness of the method. Some concluding remarks are given in Section 5.

2. FFT–based Preconditioners. The least squares solutions to (1.3) can be obtained by solving the scaled (normalized version of) normal equations

$$\frac{1}{2M}(X_1^T X_1 + X_2^T X_2 + X_2^T X_1 + X_1^T X_2)\mathbf{w} = \frac{1}{2M}(X_1 + X_2)^T \mathbf{x}. \tag{2.1}$$

We note that X_1 and X_2 have special structures. Each row of X_1 is a right-shifted version of the previous row and each row of X_2 is a left-shifted version of the previous row. By utilizing these special rectangular Toeplitz and Hankel structures, the matrices $\frac{1}{2M}(X_1^T X_1 + X_2^T X_2)$ and $\frac{1}{2M}(X_2^T X_1 + X_1^T X_2)$ can be written as

$$\frac{1}{2M}(X_1^T X_1 + X_2^T X_2) = T_n \quad \text{and} \quad \frac{1}{2M}(X_2^T X_1 + X_1^T X_2) = H_n, \tag{2.2}$$

respectively. Here T_n is an n-by-n symmetric Toeplitz matrix and H_n is an n-by-n symmetric Hankel matrix. The first column of T_n is given by

$$[\gamma_0, \gamma_1, \ldots, \gamma_{n-1}]^T,$$

and the first row and the last column of H_n are given by

$$[\gamma_{2n-1}, \gamma_{2n-2}, \ldots, \gamma_n] \quad \text{and} \quad [\gamma_n, \gamma_{n-1}, \ldots, \gamma_1]^T,$$

respectively, where

$$\gamma_k = \frac{1}{M}\sum_{j=1}^{M-k} x(j)x(j+k), \quad k = 0, 1, \ldots, 2n-1.$$

ETNA
Kent State University
etna@mcs.kent.edu

In the statistics literature, if the input stochastic process is stationary, the parameters γ_k are called estimators of the autocorrelation of the stationary process. The parameters γ_k have a smaller mean square error than other estimators; see for instance Priestley [24, p. 322].

Our preconditioner is taken to be the circulant approximation of the Toeplitz part T_n of the normal equations matrix. We remark that our preconditioner is different from that recently proposed by Ku and Kuo [18] for Toepltiz-plus-Hankel systems. They basically take the circulant approximations of Toepltiz matrix and Hankel matrix and then combine them together to form a preconditioner. We note that under the assumptions in [18], the spectrum of the Hankel matrix is not clustered around zero. The motivation behind our preconditioner is that the Toeplitz matrix T_n is the sample autocorrelation matrix which intuitively should be a good estimation to the autocorrelation matrix of the discrete-time stationary process, provided that a sufficiently large number of data samples are taken. Moreover, we prove in §3 that under practical signal processing assumptions, the spectrum of the Hankel matrix H_n is clustered around zero. Hence it suffices to approximate T_n by circulant preconditioner.

In this paper, we only focus on an "optimal" circulant preconditioner $c(T_n)$ for T_n which is defined to be the minimizer of $\|Q_n - T_n\|_F$ over all n-by-n circulant matrices Q_n; see T. Chan [10]. The (j,k) entry of $c(T_n)$ is given by the diagonals c_{j-k} where

$$c_k = \begin{cases} \dfrac{(n-k)\gamma_k + k\gamma_{n-k}}{n}, & 0 \le k \le n, \\ c_{n+k}, & 0 < -k < n. \end{cases} \tag{2.3}$$

As T_n is a Toeplitz matrix, the circulant preconditioner $c(T_n)$ is found in $O(n)$ operations. An interesting spectral property of $c(T_n)$ is that if T_n is symmetric positive definite, the corresponding "optimal" circulant matrix $c(T_n)$ is also symmetric positive definite. In fact, we have that

$$\lambda_{\min}(T_n) \le \lambda_{\min}(c(T_n)) \le \lambda_{\max}(c(T_n)) \le \lambda_{\max}(T_n), \tag{2.4}$$

where $\lambda_{\min}$ and $\lambda_{\max}$ denote the minimum and maximum eigenvalues respectively; see Tyrtyshnikov [26].

In addition, the preconditioner is closely related to the Blackman-Tukey spectral estimator with the Bartlett window that is one of the popular method for nonparametric spectral analysis in signal processing; see [2]. The Bartlett spectral estimator can be expressed as

$$s(\omega) = \sum_{k=-M}^{M} W(k)\gamma_k e^{i\omega k}, \quad \forall \omega \in [0, 2\pi],$$

where

$$W(k) = \begin{cases} 1 - \dfrac{|k|}{n}, & |k| \le n, \\ 0, & |k| > n; \end{cases}$$

see [17, p. 80]. On the other hand, as $c(T_n)$ is a circulant matrix, it can be diagonalized by the discrete Fourier matrix F_n with entries $[F_n]_{j,k} = \frac{1}{\sqrt{n}} e^{2\pi ijk/n}$, we have that

$$c(T_n) = F_n^* \Lambda_n F_n,$$

ETNA
Kent State University
etna@mcs.kent.edu

where Λ_n is a diagonal matrix whose diagonal entries are the eigenvalues of $c(T_n)$. Using the relationship between the first column of $c(T_n)$ and its eigenvalues, the eigenvalues $\lambda_j(c(T_n))$ of $c(T_n)$ can be expressed as

$$\lambda_j(c(T_n)) = \gamma_0 + \sum_{k=1}^{n-1} \left[\frac{n-k}{n}\gamma_k + \frac{k}{n}\gamma_{n-k} \right] \xi_j^k, \quad 0 \le j < n, \tag{2.5}$$

where $\xi_j = e^{2\pi ij/n}$; see also Chan and Yeung [9]. After some rearrangement of the terms in (2.5), we note that the eigenvalues of $c(T_n)$ are equal to the values of $s(\omega)$ sampled at the points $\{2\pi j/n\}_{j=0}^{n-1}$ on $[0, 2\pi]$.

3. Spectra of Preconditioned Normal Equations Matrices. As we deal with data samples from stochastic processes, the convergence rate will be considered in a probabilistic way which is different from the deterministic case discussed in [1, pp. 24-28]. We first make the following practical signal processing assumptions (**A**) on the input discrete-time real-valued process $\{x(i)\}$:

- (**A1**) The process is stationary with non-zero constant mean μ;
- (**A2**) The underlying spectral density function of the process is a $(\ell+1)$-times differentiable function for $\ell > 0$;
- (**A3**) The spectral density function of the process is positive;
- (**A4**) The variances of $\frac{1}{M}\sum_{j=1}^{M-k} x(j)$ and $\frac{1}{M}\sum_{j=1}^{M-k}[x(j)-\mu][x(j+k)-\mu]$ are bounded by

$$\mathrm{Var}\left(\frac{1}{M} \sum_{j=1}^{M-k} x(j) \right) \le \frac{\beta_1}{M}, \quad k = 0, 1, 2, \ldots, M-1, \tag{3.1}$$

and

$$\mathrm{Var}\left(\frac{1}{M} \sum_{j=1}^{M-k} [x(j)-\mu][x(j+k)-\mu] \right) \le \frac{\beta_2}{M}, \quad k = 0, 1, 2, \ldots, M-1, \tag{3.2}$$

where β_1 and β_2 are positive constants depending on the input stochastic process.

The remarks on the assumptions can be found in [23]. The following lemma, which gives the spectrum of the covariance matrix, R_n appeared in Haykin [15, p.139].

LEMMA 3.1. *Let the stochastic process $\{x(i)\}$ be stationary with zero mean and let its spectral density function be $f(\theta)$ with minimum and maximum values $f_{\min}$ and $f_{\max}$, respectively. Then the spectrum $\sigma(R_n)$ of R_n satisfies*

$$\sigma(R_n) \subseteq [f_{\min}, f_{\max}], \quad \forall n \ge 1. \tag{3.3}$$

In the following, we express $x(j)x(j+k)$ in terms of μ:

$$x(j)x(j+k) = [x(j)-\mu][x(j+k)-\mu] + \mu[x(j)+x(j+k)] - \mu^2.$$

Thus, the matrices T_n and H_n can be written as

$$T_n = T_n^{(1)} + \mu T_n^{(2)} - \mu^2 T_n^{(3)} \quad \text{and} \quad H_n = H_n^{(1)} + \mu H_n^{(2)} - \mu^2 H_n^{(3)},$$

respectively. The (j,k)th entries of Toeplitz matrices $T_n^{(1)}$, $T_n^{(2)}$ and $T_n^{(3)}$ are given by

$$[T_n^{(1)}]_{j,k} = \frac{1}{M} \sum_{p=1}^{M-|j-k|} [x(p)-\mu][x(p+|j-k|)-\mu], \quad 0 \le j,k < n, \tag{3.4}$$

ETNA
Kent State University
etna@mcs.kent.edu

$$[T_n^{(2)}]_{j,k} = \frac{1}{M}\sum_{p=1}^{M-|j-k|}[x(p)+x(p+|j-k|)], \quad 0 \le j,k < n, \tag{3.5}$$

and

$$[T_n^{(3)}]_{j,k} = \frac{(M-|j-k|)}{M}, \quad 0 \le j,k < n, \tag{3.6}$$

respectively. The (j,k)th entries of Hankel matrices $H_n^{(1)}$, $H_n^{(2)}$ and $H_n^{(3)}$ are given by

$$[H_n^{(1)}]_{j,k} = \frac{1}{M}\sum_{p=1}^{M-2n+1+j+k}[x(p)-\mu][x(p+2n-1-j-k)-\mu], \quad 0 \le j,k < n, \tag{3.7}$$

$$[H_n^{(2)}]_{j,k} = \frac{1}{M}\sum_{p=1}^{M-2n+1+j+k}[x(p)+x(p+2n-1-j-k)], \quad 0 \le j,k < n, \tag{3.8}$$

and

$$[H_n^{(3)}]_{j,k} = \frac{(M-2n+1+j+k)}{M}, \quad 0 \le j,k < n, \tag{3.9}$$

respectively. By the linearity of the "optimal" circulant approximation, $c(T_n)$ is decomposed into three parts:

$$c(T_n) = c(T_n^{(1)}) + \mu c(T_n^{(2)}) - \mu^2 c(T_n^{(3)}). \tag{3.10}$$

In the following discussions, we let $\mathcal{E}(Z)$ be the expected value of a random matrix Z, so that the entries of $\mathcal{E}(Z)$ are the expected value of the elements of Z, i.e.,

$$[\mathcal{E}(Z)]_{j,k} = \mathcal{E}([Z]_{j,k}), \quad 0 \le j,k < n.$$

The following two lemmas will be useful later in the analysis of the convergence rate of the method.

LEMMA 3.2. (Ng and Chan [23, Theorem 1]) *Let the stochastic process* $\{x(i)\}$ *satisfy assumptions* (**A1**), (**A2**) *and* (**A4**). *Then for any given* $\epsilon > 0$ *and* $0 < \eta < 1$, *there exist positive integers* K *and* N *such that for* $n > N$,

$$\|\mathcal{E}(T_n^{(1)}) - R_n\|_2 \le \epsilon,$$

and Pr $\{$ *at most* K *eigenvalues of* $T_n^{(1)} - c(T_n^{(1)})$ *have absolute value greater than* ϵ $\}$ $> 1-\eta$, *provided that* $M = \Omega(n^{3+\nu})$ *with* $\nu > 0$, *i.e., number of data samples taken is at least as large as* $n^{3+\nu}$.

We remark that in Lemma 3.2, the parameter ν is theoretically used to let the probability of the event tend to 1.

LEMMA 3.3. *Let the stochastic process* $\{x(i)\}$ *satisfy assumption* (**A4**). *Then, for any given* $\epsilon > 0$, *we have that*

$$\Pr\left\{\sum_{k=\tau_1}^{\tau_2}\left|\frac{1}{M}\sum_{j=1}^{M-k}[x(j)+x(j+k)] - \mathcal{E}\left(\frac{1}{M}\sum_{j=1}^{M-k}[x(j)+x(j+k)]\right)\right| \le \epsilon\right\}$$

$$\ge 1 - \frac{8|\tau_2-\tau_1+1|^3\beta_1}{\epsilon^2 M},$$

ETNA
Kent State University
etna@mcs.kent.edu

and that

$$\Pr\left\{\sum_{k=\tau_1}^{\tau_2}\left|\frac{1}{M}\sum_{j=1}^{M-k}[x(j)-\mu][x(j+k)-\mu]-\mathcal{E}\left(\frac{1}{M}\sum_{j=1}^{M-k}[x(j)-\mu][x(j+k)-\mu]\right)\right|\le\epsilon\right\}$$

$$\ge 1-\frac{|\tau_2-\tau_1+1|^3\beta_2}{\epsilon^2 M},$$

for any integers τ_1 *and* τ_2.

Proof. By using a lemma in Fuller [12, p.182] and Chebyshev's inequality [12, p.185], we have

$$\begin{aligned}
&\Pr\left\{\sum_{k=\tau_1}^{\tau_2}\left|\frac{1}{M}\sum_{j=1}^{M-k}[x(j)+x(j+k)]-\mathcal{E}\left(\frac{1}{M}\sum_{j=1}^{M-k}[x(j)+x(j+k)]\right)\right|\ge\epsilon\right\}\\
\le\ &\sum_{k=\tau_1}^{\tau_2}\Pr\left\{\left|\frac{1}{M}\sum_{j=1}^{M-k}[x(j)+x(j+k)]-\mathcal{E}\left(\frac{1}{M}\sum_{j=1}^{M-k}[x(j)+x(j+k)]\right)\right|\ge\frac{\epsilon}{|\tau_2-\tau_1+1|}\right\}\\
\le\ &\sum_{k=\tau_1}^{\tau_2}\frac{4|\tau_2-\tau_1+1|^2\left[\mathrm{Var}\left(\frac{1}{M}\sum_{j=1}^{M-k}x(j)\right)+\mathrm{Var}\left(\frac{1}{M}\sum_{j=1}^{M-k}x(j+k)\right)\right]}{\epsilon^2}\\
\le\ &\frac{8|\tau_2-\tau_1+1|^3\beta_1}{\epsilon^2 M}.
\end{aligned}$$

The other part can be derived similarly; it is therefore omited. □

Using Lemma 3.3, we prove that the ℓ_2 norm of the difference between the random matrices and their expected values is sufficiently small with probability 1.

COROLLARY 3.4. *Let the stochastic process* $\{x(i)\}$ *satisfy assumptions* (**A1**), (**A2**) *and* (**A4**). *Then, for any given* $\epsilon>0$ *and* $0<\eta<1$, *we have that*

$$\Pr\left\{\|T_n^{(1)}-\mathcal{E}(T_n^{(1)})+\mu[T_n^{(2)}-\mathcal{E}(T_n^{(2)})]\|_2\le\epsilon\right\}>1-\eta \tag{3.11}$$

and that

$$\Pr\left\{\|H_n^{(1)}-\mathcal{E}(H_n^{(1)})+\mu[H_n^{(2)}-\mathcal{E}(H_n^{(2)})]\|_2\le\epsilon\right\}>1-\eta, \tag{3.12}$$

provided that $M=\Omega(n^{3+\nu})$ *with* $\nu>0$.

Proof. We note that

$$\begin{aligned}
\|T_n^{(1)}-\mathcal{E}(T_n^{(1)})+\mu[T_n^{(2)}-\mathcal{E}(T_n^{(2)})]\|_2 &\le \|T_n^{(1)}-\mathcal{E}(T_n^{(1)})\|_2+\mu\|T_n^{(2)}-\mathcal{E}(T_n^{(2)})\|_2\\
&\le \|T_n^{(1)}-\mathcal{E}(T_n^{(1)})\|_1+\mu\|T_n^{(2)}-\mathcal{E}(T_n^{(2)})\|_1.
\end{aligned}$$

It can be shown that both $\|T_n^{(1)}-\mathcal{E}(T_n^{(1)})\|_1$ and $\|T_n^{(2)}-\mathcal{E}(T_n^{(2)})\|_1$ are bounded by 2 times the ℓ_1 norm of their corresponding first column vectors; see (3.4) and (3.5). Then the result follows by setting $\tau_1=0$ and $\tau_2=n-1$ in Lemma 3.3. Using similar arguments, we establish the same bound for $\|H_n^{(1)}-\mathcal{E}(H_n^{(1)})+H_n^{(1)}-\mathcal{E}(H_n^{(1)})\|_2$. In fact, $\|H_n^{(1)}-\mathcal{E}(H_n^{(1)})\|_1$ and $\|H_n^{(2)}-\mathcal{E}(H_n^{(2)})\|_1$ are bounded by 2 times the ℓ_1 norm of their corresponding last column vectors. Hence, (3.12) follows. □

ETNA
Kent State University
etna@mcs.kent.edu

Next we prove the main result of the paper about the clustering property of the matrices H_n.

THEOREM 3.5. *Let the stochastic process $\{x(i)\}$ satisfy assumptions* (**A1**), (**A2**) *and* (**A4**). *Then for any given $\epsilon > 0$ and $0 < \eta < 1$, there exist positive integers K and N such that for $n > N$, Pr $\{$ at most K eigenvalues of H_n have absolute value greater than ϵ $\} > 1 - \eta$, provided that $M = \Omega(n^{3+\nu})$ with $\nu > 0$.*

Proof. We write H_n as follows:

$$\begin{aligned} H_n &= H_n^{(1)} - \mathcal{E}(H_n^{(1)}) + \mu[H_n^{(2)} - \mathcal{E}(H_n^{(2)})] - \mu^2 H_n^{(3)} + \\ &\quad \mu\mathcal{E}(H_n^{(2)}) - \mu^2 L_n + \mu^2 L_n + \mathcal{E}(H_n^{(1)}), \end{aligned}$$

where L_n is an n-by-n matrix with all entries being 1. By (3.8) and (**A1**), we obtain

$$\mu\mathcal{E}(H_n^{(2)}) - \mu^2 H_n^{(3)} = \mu^2 H_n^{(3)}.$$

By using (3.12) in Corollary 3.4 and $\|H_n^{(3)} - L_n\|_2 \leq \frac{n(n-1)}{M}$, we have that

$$\Pr\left\{\|H_n^{(1)} - \mathcal{E}(H_n^{(1)}) + \mu[H_n^{(2)} - \mathcal{E}(H_n^{(2)})] - \mu^2 H_n^{(3)} + \mu\mathcal{E}(H_n^{(2)}) - \mu^2 L_n\|_2 \leq \epsilon\right\} > 1-\eta,$$

provided that $M = \Omega(n^{3+\nu})$ with $\nu > 0$. We remark that the rank $(L_n) = 1$. Therefore, it suffices to prove that the spectrum of $\mathcal{E}(H_n^{(1)})$ is clustered around zero deterministically. By (3.7), the entries of $\mathcal{E}(H_n^{(1)})$ are given by

$$[\mathcal{E}(H_n^{(1)})]_{j,k} = \frac{(M - 2n + 1 + j + k)r_{2n-1-j-k}}{M}, \quad 0 \leq j, k < n,$$

where r_k is the k-lag autocovariance of the stationary process. By (**A2**), the autocovariances of the stationary process are absolutely summable. Hence, for any given $\epsilon > 0$, there exists an $N > 0$ such that

$$\sum_{j=N+1}^{\infty} |r_j| < \epsilon. \tag{3.13}$$

Let U_n be the n-by-n matrix obtained from $\mathcal{E}(H_n^{(1)})$ by replacing the $(n-N)$-by-$(n-N)$ leading principal submatrix of $\mathcal{E}(H_n^{(1)})$ by the zero matrix. Then, rank $(U_n) \leq 2N$. Let $V_n \equiv \mathcal{E}(H_n^{(1)}) - U_n$. The leading $(n-N)$-by-$(n-N)$ block of V_n is the leading $(n-N)$-by-$(n-N)$ principal submatrix of $\mathcal{E}(H_n^{(1)})$. Hence, this block is a Hankel matrix, and using (3.13) and

$$|r_k| \leq \frac{\beta_3}{|k|^{\ell+1}},$$

where β_3 is a positive constant, the ℓ_1 norm of V_n is attained at the $(n-N-1)$th column. As V_n is a symmetric matrix, the result follows by noting that $\|V_n\|_2 \leq \|V_n\|_1 \leq \epsilon$. □

Under the assumptions, the smallest eigenvalues of T_n and $c(T_n)$ are uniformly bounded away from zero with probability 1. Therefore, $c(T_n)$ is uniformly invertible. As we consider the process with non-zero mean in general, the theorem below extends the result of Theorem 2 in Ng and Chan [23].

ETNA
Kent State University
etna@mcs.kent.edu

THEOREM 3.6. *Let the stochastic process* $\{x(i)\}$ *satisfy assumptions* (**A1**) *and* (**A4**). *Then for any given* $\epsilon > 0$ *and* $0 < \eta < 1$, *there exist a positive integer* N *such that for* $n > N$,

$$\Pr\{\lambda_{\min}(T_n) \geq f_{\min} - \epsilon\} > 1 - \frac{\eta}{2} \quad \text{and} \quad \Pr\left\{\lambda_{\max}(T_n) \leq \mu^2 n + f_{\max} + \epsilon\right\} > 1 - \frac{\eta}{2},$$

provided that $M = \Omega(n^{3+\nu})$ *with* $\nu > 0$. *In particular, we have that*

$$\Pr\{\lambda_{\min}(c(T_n)) \geq f_{\min} - \epsilon\} > 1 - \frac{\eta}{2} \quad \text{and} \quad \Pr\left\{\lambda_{\max}(c(T_n)) \leq \mu^2 n + f_{\max} + \epsilon\right\} > 1 - \frac{\eta}{2}.$$

Proof. We write

$$\begin{aligned} T_n &= T_n^{(1)} - \mathcal{E}(T_n^{(1)}) + \mu[T_n^{(2)} - \mathcal{E}(T_n^{(2)})] - \mu^2 T_n^{(3)} + \mu\mathcal{E}(T_n^{(2)}) - \mu^2 L_n + \\ &\quad \mathcal{E}(T_n^{(1)}) - R_n + R_n + \mu^2 L_n. \end{aligned}$$

By (3.5), (3.6) and (**A1**), we obtain

$$\mu\mathcal{E}(T_n^{(2)}) - \mu^2 T_n^{(3)} = \mu^2 T_n^{(3)} \quad \text{and} \quad \|T_n^{(3)} - L_n\|_2 \leq \frac{n(n-1)}{M}.$$

Using Lemma 3.1 and the fact that R_n and L_n are symmetric and the eigenvalues of $\mu^2 L_n$ are 0 and $\mu^2 n$, it follows by Corollary in [14, p.269] that the smallest and largest eigenvalues of $R_n + \mu^2 L_n$ are bounded below by $f_{\min}$ and above by $f_{\max} + \mu^2 n$, respectively. Then, the result follows by using Lemma 3.2, (3.11) in Corollary 3.4 and some simple probability arguments, provided that $M = \Omega(n^{3+\nu})$ with $\nu > 0$. Using (2.4), we immediately have the result for the smallest and largest eigenvalues of $c(T_n)$. □

THEOREM 3.7. *Let the stochastic process* $\{x(i)\}$ *satisfy assumptions* (**A1**), (**A2**) *and* (**A4**). *Then for any given* $\epsilon > 0$ *and* $0 < \eta < 1$, *there exist positive integers* K *and* N *such that for* $n > N$, *Pr* $\{$ *at most* K *eigenvalues of* $T_n + H_n - c(T_n)$ *have absolute value greater than* ϵ $\} > 1 - \frac{\eta}{2}$, *provided that* $M = \Omega(n^{3+\nu})$ *with* $\nu > 0$.

Proof. By (3.10),

$$\begin{aligned} T_n - c(T_n) &= T_n^{(1)} - c(T_n^{(1)}) + \mu[T_n^{(2)} - \mathcal{E}(T_n^{(2)}) + \mathcal{E}(T_n^{(2)}) - \mathcal{E}(c(T_n^{(2)})) + \\ &\quad \mathcal{E}(c(T_n^{(2)})) - c(T_n^{(2)})] - \mu^2[T_n^{(3)} - c(T_n^{(3)})]. \end{aligned}$$

We note from (3.5), (3.6) and (**A1**) that

$$\mu\mathcal{E}(T_2^{(n)}) - \mu^2 T_n^{(3)} = \mu^2 T_n^{(3)} \quad \text{and} \quad -\mu\mathcal{E}(c(T_2^{(n)})) + \mu^2 c(T_n^{(3)}) = -\mu^2 c(T_n^{(3)}).$$

Thus, (3.14) becomes

$$\begin{aligned} T_n - c(T_n) &= T_n^{(1)} - c(T_n^{(1)}) + \mu[T_n^{(2)} - \mathcal{E}(T_n^{(2)}) + c(T_n^{(2)}) - \mathcal{E}(c(T_n^{(2)}))] + \\ &\quad \mu^2 T_n^{(3)} - L_n + L_n - \mu^2 c(T_n^{(3)}). \end{aligned}$$

Using (2.4), the commutative property of circulant approximation and expectation operator and the circulant structure of L_n, we obtain

$$\|c(T_n^{(2)}) - \mathcal{E}(c(T_n^{(2)}))\|_2 \leq \|T_n^{(2)} - \mathcal{E}(T_n^{(2)})\|_2 \tag{3.14}$$

ETNA
Kent State University
etna@mcs.kent.edu

and

$$\|c(T_n^{(3)}) - L_n\|_2 \leq \|T_n^{(3)} - L_n\|_2 \leq \frac{(n-1)n}{M}. \tag{3.15}$$

In view of Lemmas 3.2 and 3.5, (3.11) in Corollary 3.4, (3.14) and (3.15), the theorem follows. □

By combining Theorems (3.6) and (3.7), the main theorem concerning the spectra of the preconditioned matrices is proved.

THEOREM 3.8. *Let the stochastic process* $\{x(i)\}$ *satisfy assumption* (**A**). *Then for any given* $\epsilon > 0$ *and* $0 < \eta < 1$, *there exist positive integers* K *and* N *such that for* $n > N$, $Pr\ \{$ *at most* K *eigenvalues of* $c(T_n)^{-1}(T_n + H_n)$ *have absolute value larger than* $\epsilon\ \} > 1 - \eta$, *provided that* $M = \Omega(n^{3+\nu})$ *with* $\nu > 0$.

As for the convergence rate of the preconditioned conjugate gradient method for our circulant preconditioned Toeplitz-plus-Hankel matrix $c(T_n)^{-1}(T_n + H_n)$, the method converges in at most $O((2\alpha+1)\log n + 1)$ steps when the smallest singular value of the data matrix $X_1 + X_2$ is of order $O(n^\alpha)$. We begin by noting the following error estimate of the conjugate gradient method; see [5].

LEMMA 3.9. *Let* G_n *be an* n*-by-*n *positive definite matrix and* $\mathbf{z}$ *be the solution to* $G_n\mathbf{z} = \mathbf{v}$. *Let* $\mathbf{z}_j$ *be the* j*th iterant of the ordinary conjugate gradient method applied to the equation* $G_n\mathbf{z} = \mathbf{v}$. *If the eigenvalues* $\{\lambda_k\}$ *of* G_n *are such that*

$$0 < \lambda_1 \leq ... \leq \lambda_p \leq b_1 \leq \lambda_{p+1} \leq ... \leq \lambda_{n-q} \leq b_2 \leq \lambda_{n-q+1} \leq ... \leq \lambda_n,$$

then

$$\frac{||\mathbf{z} - \mathbf{z}_j||_{G_n}}{||\mathbf{z} - \mathbf{z}_0||_{G_n}} \leq 2\left(\frac{b-1}{b+1}\right)^{j-p-q} \cdot \max_{\lambda\in[b_1,b_2]} \left\{ \prod_{k=1}^{p} \left(\frac{\lambda - \lambda_k}{\lambda_k}\right) \right\}. \tag{3.16}$$

Here $b \equiv (b_2/b_1)^{\frac{1}{2}} \geq 1$ *and* $||\mathbf{v}||_{G_n} \equiv \mathbf{v}^* G_n \mathbf{v}$.

For the preconditioned system

$$c(T_n)^{-1}(T_n + H_n)\mathbf{w} = (X_1 + X_2)^T\mathbf{d}, \tag{3.17}$$

the iteration matrix G_n is given by $G_n = c(T_n)^{-1/2}(T_n + H_n)c(T_n)^{-1/2}$. Theorem 3.8 implies that we can choose $b_1 = 1 - \epsilon$ and $b_2 = 1 + \epsilon$, with probability 1. Then, p and q are constants that depend only on ϵ but not on n. By choosing $\epsilon < 1$, we have that

$$\frac{b-1}{b+1} = \frac{1 - \sqrt{1-\epsilon^2}}{\epsilon} < \epsilon.$$

In order to use (3.16), we need a lower bound for λ_k, $1 \leq k \leq p$. We note that with probability 1

$$||G_n^{-1}||_2 = ||(T_n + H_n)^{-1}c(T_n)||_2 \leq \|c(T_n)\|_2\|(T_n + H_n)^{-1}\|_2 \leq (n + f_{\max} + \epsilon)n^{2\alpha}.$$

We then see that for all n sufficiently large,

$$||G_n^{-1}||_2 \leq \beta_4 n^{2\alpha+1},$$

for some constant β_4 that does not depend on n. Therefore,

$$\lambda_k \geq \min_\ell \lambda_\ell = \frac{1}{||G_n^{-1}||_2} \geq \beta_4 n^{-(2\alpha+1)}, \quad 1 \leq k \leq n.$$

ETNA
Kent State University
etna@mcs.kent.edu

Thus, for $1 \leq k \leq p$ and $\lambda \in [1-\epsilon, 1+\epsilon]$, we have that

$$0 \leq \frac{\lambda - \lambda_k}{\lambda_k} \leq \beta_4 n^{2\alpha+1}.$$

Hence, (3.16) becomes

$$\frac{||\mathbf{w} - \mathbf{w}_j||_{G_n}}{||\mathbf{w} - \mathbf{w}_0||_{G_n}} < \beta_4^p n^{p(2\alpha+1)} \epsilon^{j-p-q}.$$

Given an arbitrary tolerance $\delta > 0$, an upper bound for the number of iterations required to make

$$\frac{||\mathbf{w} - \mathbf{w}_j||_{G_n}}{||\mathbf{w} - \mathbf{w}_0||_{G_n}} < \delta$$

is therefore given by

$$j_0 \equiv p + q - \frac{p \log \beta_4 + (2\alpha+1) p \log n - \log \delta}{\log \epsilon} = O(2\alpha \log n + 1),$$

with probability 1.

Since by using FFTs, the Toeplitz, Hankel and circulant matrix-vector products in the PCG method can be done in $O(n \log n)$ operations, the cost per iteration of the conjugate gradient method is of order $O(n \log n)$. Thus, we conclude that the work of solving (3.17) to a given accuracy δ is of order $O((2\alpha+1) n \log^2 n + n \log n)$ when $\alpha > 0$. We remark that the order of complexity of our iterative method is less than that of direct methods (see Merchant and Parks [22] and Yagle [27]) which requires $O(n^2)$ operations.

4. Numerical Experiments. In this section, the results of numerical experiments which test the convergence performance of the algorithm are described. All the computations are done by Matlab on a Sparc workstation. We used AR(2) and MA(2) processes given by

$$x(t) - 1.4x(t-1) + 0.5x(t-2) = v(t) \quad \text{and} \quad x(t) = v(t) + 0.75v(t-1) + 0.25v(t-2),$$

respectively, to generate the Toeplitz-plus-hankel matrices $T_n + H_n$. Here $\{v(t)\}$ is a Gaussian process with zero mean and variance 1 as input stationary process. In Figures 1 and 2, we depict the spectra of the normal equations matrix and the preconditioned normal equations matrix in one of the realizations of the AR(2) and MA(2) processes, respectively, with $n = 128$ and $M = 1024$. We note that the spectra of the preconditioned matrices indeed are clustered around 1.

In the numerical tests, we uses the zero vector and a random vector as our initial guess and right hand side vector. The stopping criterion of the preconditioned conjugate gradient method was $\|\mathbf{e}_j\|_2 / \|\mathbf{e}_0\|_2 < 10^{-7}$, where $\mathbf{e}_j$ is the residual vector after j iterations. In the tables below, $M' = M/n$ is the number of blocks of data samples with size n and I denotes no preconditioner is used whereas C signifies the "optimal" circulant preconditioner is used. Tables 1–2 show the average number of iterations (rounded to the nearest integer) over 100 runs of the algorithms when AR(2) and MA(2) processes are used. We see that the preconditioned system converges very fast and the average number of iterations of preconditioned systems is much less than that of non-preconditioned one when n is large. As for the comparison of times in

ETNA
Kent State University
etna@mcs.kent.edu

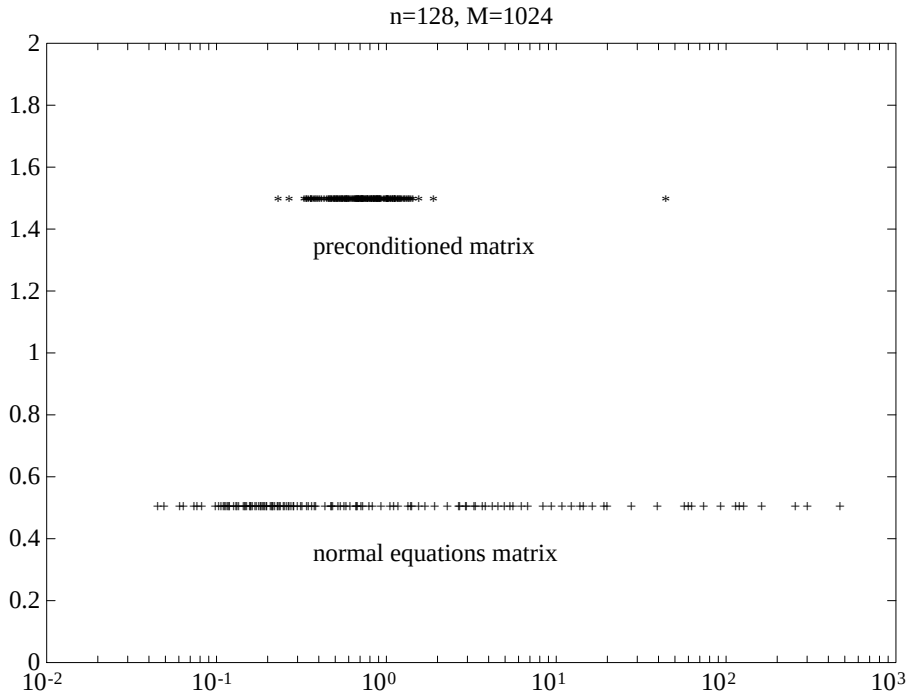

FIG. 4.1. *Eigenvalues for normal equations matrix and preconditioned matrix when AR(2) process is used. (autocorrelation windowing method)*

n	16		32		64		128	
M'	I	C	I	C	I	C	I	C
4	22	16	53	23	113	27	233	30
8	22	15	50	21	102	23	208	24
16	22	15	49	18	96	20	183	19
32	23	14	47	17	94	18	169	16
64	23	14	47	16	90	16	165	14

TABLE 4.1
Average number of iterations when AR(2) process is used.

conjugate gradient iterations, Tables 3 and 4 show the average number of kilo-flops (counted by Matlab) used for the cases of the AR(2) and MA(2) processes tested in Tables 1 and 2. We see from the tables that the number of kilo-flops used for the preconditioned systems is significantly less than that of non-preconditioned systems especially when n is large.

In this paper, we employed the *autocorrelation windowing method* to formulate the Toeplitz-plus-Hankel least squares problem. Other windowing methods can be used, for instance, the *covariance windowing method*, the *pre-windowed method* and the *post-windowed method*; see Haykin [15, p.373]. We remark that the other windowing methods lead to non-Toeplitz-plus-Hankel normal equations matrices. However, by

ETNA
Kent State University
etna@mcs.kent.edu

n	16		32		64		128	
M'	I	C	I	C	I	C	I	C
4	17	14	34	19	55	22	74	25
8	17	13	32	15	48	14	59	19
16	16	11	29	13	42	14	50	15
32	16	10	28	11	39	12	44	12
64	16	9	28	10	36	10	39	9

TABLE 4.2
Average number of iterations when MA(2) process is used.

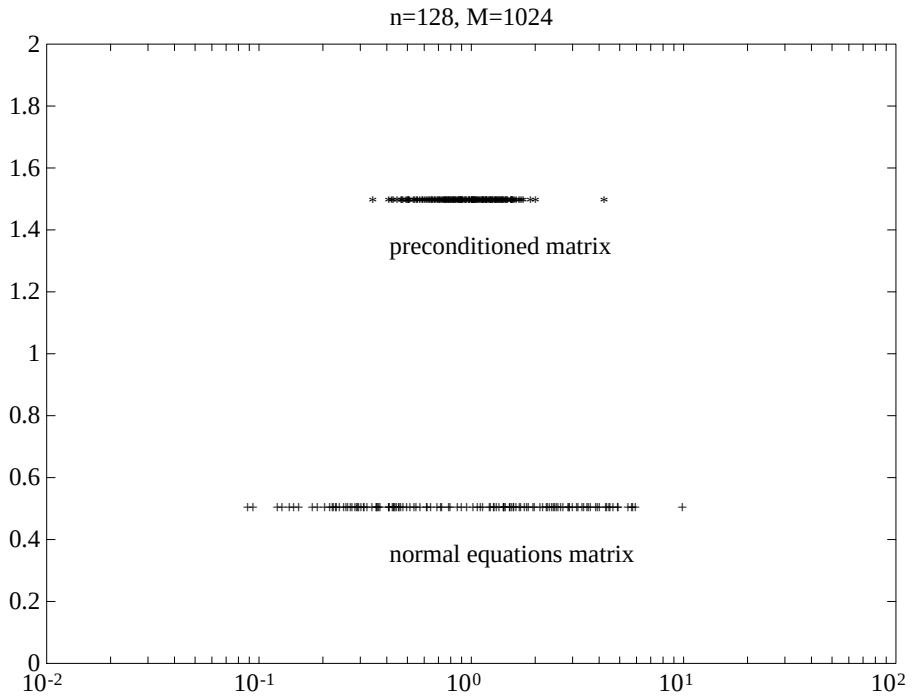

FIG. 4.2. *Eigenvalues for normal equations matrix and preconditioned matrix when MA(2) process is used. (autocorrelation windowing method)*

n	16		32		64		128	
M'	I	C	I	C	I	C	I	C
4	52	54	277	145	1306	451	5925	1106
8	52	51	262	132	1179	384	5290	885
16	52	51	256	114	1109	334	4654	700
32	54	47	246	107	1086	300	4298	590
64	54	47	246	101	1040	267	4196	516

TABLE 4.3
Average number of kilo-flops (rounded to nearest kilo-flops) when AR(2) process is used.

ETNA
Kent State University
etna@mcs.kent.edu

n	16		32		64		128	
M'	I	C	I	C	I	C	I	C
4	40	47	178	120	636	367	1882	922
8	40	44	167	95	555	234	1500	700
16	38	37	152	82	485	234	1272	553
32	38	34	146	69	451	200	1119	442
64	38	30	146	63	416	167	992	332

TABLE 4.4
Average number of kilo-flops (rounded to nearest kilo-flops) when MA(2) process is used.

n	16		32		64		128	
M'	I	C	I	C	I	C	I	C
4	22	16	53	23	113	27	233	30
8	22	15	50	21	102	23	208	24
16	22	15	49	18	96	20	183	19
32	23	14	47	17	94	18	169	16
64	23	14	47	16	90	16	165	14

TABLE 4.5
Average number of iterations when AR(2) processes is used.

exploiting the structure of the normal equations matrices, it can still be written as

$$\frac{1}{2M}(X_1 + X_2)^T(X_1 + X_2) = T_n + H_n - S_n^{(1)} - S_n^{(2)} - S_n^{(3)} - S_n^{(4)},$$

where $S_n^{(i)}$ are non-Toeplitz and non-Hankel matrices. By considering similar arguments as in Ng and Chan [23], it can be shown that the ℓ_2 norm of these matrices S_i $(i = 1, 2, 3, 4)$ are sufficiently small when M is sufficiently large. Therefore, our algorithm can handle the Toeplitz-plus-Hankel least squares problems with the use of different windowing methods. To illustrate the performance of our preconditioner for these problems, we use the AR(2) process to generate the covariance windowing data matrices X_1 and X_2. In Figure 3, we depict the spectra of the normal equations matrix and the preconditioned normal equations matrix in one realization of the AR(2) process where $n = 128$ and $M = 1024$. The figure shows clustering of the eigenvalues of the FFT–based preconditioned matrices. Also Tables 5 and 6 show the average number of iterations (rounded to the nearest integer) and the corresponding average number of kilo-flops required respectively over 100 runs of the algorithms when the AR(2) process is used. We see that both the average number of iterations and the average kilo-flops used by the preconditioned systems are much less than those of the non-preconditioned systems especially when n is large.

5. Concluding Remarks. In this paper, we have proposed a new FFT–based preconditioned Toeplitz-plus-Hankel least squares iteration. Our preliminary numerical results show the effectiveness of our algorithm. As a summary, we list the following remarks concerning our algorithm:

(i) In signal processing applications, the linear-phase filters can also be characterized by antisymmetric impulse responses. We solve the Toeplitz-plus-Hankel

ETNA
Kent State University
etna@mcs.kent.edu

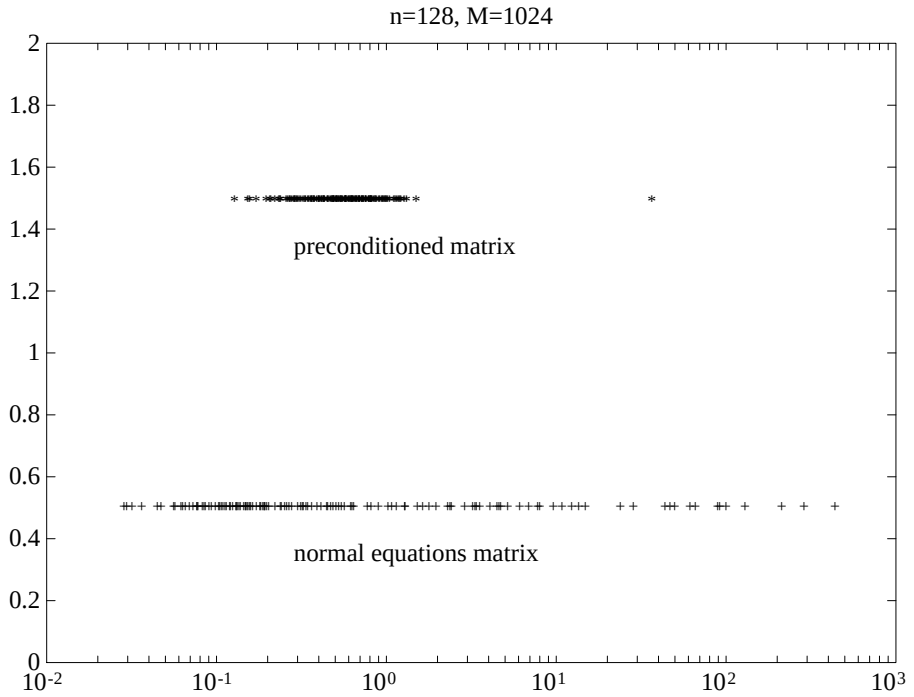

FIG. 4.3. *Eigenvalues for normal equations matrix and preconditioned matrix when AR(2) process is used. (covariance windowing method)*

n	16		32		64		128	
M'	I	C	I	C	I	C	I	C
4	488	370	2640	1194	12574	3133	57481	7721
8	488	346	2491	1091	11350	2669	51313	6177
16	488	346	2441	935	10682	2321	45146	4890
32	510	323	2341	883	10459	2088	41692	4118
64	510	323	2341	831	10014	1856	40705	3603

TABLE 4.6
Average number of kilo-flops (rounded to nearest kilo-flops) when AR(2) process is used.

least squares problem

$$\min \|\mathbf{d} - (X_1 - X_2)\mathbf{w}\|_2,$$

and the normal equations become

$$\frac{1}{2M}(X_1^T X_1 + X_2^T X_2 - X_2^T X_1 - X_1^T X_2)\mathbf{w} = \frac{1}{2M}(X_1 - X_2)^T \mathbf{x}.$$

The preconditioned conjugate gradient algorithm can also be applied to solve normal equations in this case.

ETNA
Kent State University
etna@mcs.kent.edu

(ii) Recently, other discrete transform matrices W_n have been used to construct the "optimal" preconditioners to symmetric Toeplitz matrices. These transform matrices include the sine transform [7] and the Hartley transform [3]. These preconditioners are defined to be the minimizer of $\|Q_n - T_n\|_F$ over all n-by-n matrices Q_n that can be diagonalized by W_n. We note that they are defined similarly to $c(T_n)$. In [7] and [3], it was shown that the these preconditioners perform very well when solving symmetric Toeplitz systems. Thus, we expect these preconditioners to be good alternatives to our FFT–based ones for solving Toeplitz-plus-Hankel least squares problems.

(iii) In [22], it has been shown that a Toeplitz-plus-Hankel system of equations can be reformulated as a block-Toeplitz system of equations with 2×2 blocks, i.e.

$$\begin{pmatrix} T_n & H_n \\ H_n & T_n \end{pmatrix} \begin{pmatrix} \mathbf{w} \\ J_n\mathbf{w} \end{pmatrix} = \begin{pmatrix} \mathbf{d} \\ J_n\mathbf{d} \end{pmatrix}.$$

In this case, a block-circulant preconditioner

$$\begin{pmatrix} c(T_n) & 0 \\ 0 & c(T_n) \end{pmatrix}$$

can be used to precondition the block equations. By Theorem 3.5, we note that the block-circulant matrix is also a good preconditioner. However, the approach doubles the dimension of the problem being solved, and hence it doubles the operations per iteration.

(iv) Our algorithm presented in this paper is of the fixed order n and the block-processing type, i.e. M data samples are collected over a finite time interval; the estimates of the autocorrelations are then computed and an n-by-n Toeplitz-plus-Hankel system as in (2.1) is formed and solved by the preconditioned conjugate gradient method. The complexity of solving Toeplitz-plus-Hankel systems is $O(n \log^2 n)$ operations as compared to $O(n^2)$ operations required by direct solvers. We note that the basic tool of our fast iterative algorithm is the fast Fourier transform (FFT). Since the FFT algorithm is highly parallelizable and has been implemented on multiprocessors efficiently (see for instance Swarztrauber [25]), our algorithm is expected to perform efficiently in a parallel environment.

6. Acknowledgement. The author thanks the referees for numerous valuable comments and helpful suggestions.

REFERENCES

[1] O. Axelsson and V. Barker, *Finite Element Solution of Boundary Value Problems, Theory and Computation*, Academic, New York, 1984.
[2] M. S. Barlett, *Periodogram Analysis and Continuous Spectra*, Biometrika, 37 (1950), pp. 1–16.
[3] D. Bini and P. Favat, *On a Matrix Algebra Related to the Discrete Hartley Transform*, SIAM J. Matrix Anal. Appl., 14 (1993), pp. 500–507.
[4] P. J. Brockwell, *Time Series: Theory and Methods*, Springer-Verlag, New York, 1987.
[5] R. Chan, J. Nagy and R. Plemmons, *Circulant Preconditioned Toeplitz Least Squares Iterations*, to appear SIAM J. Matrix Anal. Appl., 15 (1994), pp. 80–97.
[6] R. Chan and M. Ng, *Toeplitz Preconditioners for Hermitian Toeplitz Systems*, Linear Algebra Appl., 190 (1993), pp.181–208.

ETNA
Kent State University
etna@mcs.kent.edu

[7] R. Chan, M. Ng and C. Wong, *Sine Transform Based Preconditioners For Symmetric Toeplitz Systems*, to appear in Linear Algebra Appl., (1994).
[8] R. Chan and G. Strang, *Toeplitz Equations by Conjugate Gradients with Circulant Preconditioner*, SIAM J. Sci. Statist. Comput., 10 (1989), pp. 104–119.
[9] R. Chan and M. Yeung, *Circulant Preconditioners Constructed From Kernels*, SIAM J. Numer. Anal., 29 (1992), pp. 1093–1103.
[10] T. Chan, *An Optimal Circulant Preconditioner for Toeplitz Systems*, SIAM J. Sci. Statist. Comput., 9 (1988), pp. 766–771.
[11] B. Friedlander and M. Morf, *Least Squares Algorithms for Adaptive Linear-Phase Filtering*, IEEE Trans. Acoust. Speech and Signal Processing, 30 (1982), pp. 381–390.
[12] W. Fuller, *Introduction to Statistical Time Series*, John Wiley & Sons, Inc., New York, 1976.
[13] D. Godard, *Channel Equalizations Using a Kalman Filter for Fast Data Transmission*, IBM J. Res. Develop., 18 (1974), pp. 267–273.
[14] G. Golub and C. Van Loan, *Matrix Computations*, Second ed., The Johns Hopkins University Press, Baltimore, MD, 1989.
[15] S. Haykin, *Adaptive Filter Theory*, Second ed., Prentice-Hall, Englewood Cliffs, NJ, 1991.
[16] J. Hsue and A. Yagle, *Fast Algorithms for Close-to-Toeplitz-plus-Hankel Systems of Equations and Two-Sided Linear Prediction*, IEEE Trans. Signal Processing 41 (1993), pp. 2349–2361.
[17] S. Kay, *Spectral Estimation*, Advanced Topics in Signal Processing, Prentice-Hall, Englewood Cliffs, NJ, 1988, pp. 58–122.
[18] T. Ku and C. Kuo, *Preconditioned Iterative Methods for Solving Toeplitz-plus-Hankel Systems*, SIAM J. Numer. Anal. 30 (1993), pp. 824–845.
[19] S. Marple, *A New Autoregressive Spectrum Analysis Algorithm*, IEEE Trans. Acoust. Speech and Signal Processing, 28 (1980), pp. 441–454.
[20] ———, *Efficient Least Squares FIR System Identification*, IEEE Trans. Acoust. Speech and Signal Processing, 29 (1981), pp. 62–73.
[21] ———, *Fast Algorithms for Linear Prediction and System Identification Filters with Linear Phase*, IEEE Trans. Acoust. Speech and Signal Processing, V30 (1982), pp. 942–953.
[22] G. Merchant and T. Parks, *Efficient Solution of a Toeplitz-plus-Hankel Coefficient Matrix System of Equations*, IEEE Trans. Acoust. Speech and Signal Processing, 30 (1982), pp. 40–44.
[23] M. Ng and R. Chan, *Fast Iterative Methods for Least Squares Estimations*, Numer. Algorithms, 6 (1994), pp. 353–378.
[24] M. Priestley, *Spectral Analysis and Time Series*, Academic Press, New York, 1981.
[25] P. Swarztrauber, *Multiprocessors FFTs*, Parallel Comput., 5 (1987), pp. 197–210.
[26] E. Tyrtyshnikov, *Optimal and Super-optimal Circulant Preconditioners*, SIAM J. Matrix Anal. Appl., 13 (1992), pp. 459–473.
[27] A. Yagle, *New Analogues of Split Algorithms for Arbitrary Toeplitz-plus-Hnakel Matrices*, IEEE Trans. Signal Processing, 39 (1991), pp. 2457–2463.