Fast Iterative Methods For Least Squares Estimations

Michael K. Ng and Raymond H. Chan *
Department of Mathematics
University of Hong Kong

September 1992

Abstract

Least squares estimations have been used extensively in many applications, e.g. system identification and signal prediction. When the stochastic process is stationary, the least squares estimators can be found by solving a Toeplitz or near-Toeplitz matrix system depending on the knowledge of the data statistics. In this paper, we employ the preconditioned conjugate gradient method with circulant preconditioners to solve such systems. Our proposed circulant preconditioners are derived from the spectral property of the given stationary process. In the case where the spectral density function $s(\theta)$ of the process is known, we prove that if $s(\theta)$ is a positive continuous function, then the spectrum of the preconditioned system will be clustered around 1 and the method converges superlinearly. However, if the statistics of the process is unknown, then we prove that with probability 1, the spectrum of the preconditioned system is still clustered around 1 provided that large data samples are taken. For finite impulse response (FIR) system identification problems, our numerical results show that an n-th order least squares estimators can usually be obtained in $O(n \log n)$ operations when $O(n)$ data samples are used.

Key Words. Least squares estimations, Toeplitz matrix, circulant matrix, preconditioned conjugate gradient method, signal prediction, linear prediction, covariance matrix, windowing methods, finite impulse response (FIR) system identification.

AMS(MOS) Subject Classifications. 65F10, 65F15, 43E10

*Research supported in part by HKRGC grant no. 221600070, ONR contract no. N00014-90-J-1695 and DOE grant no. DE-FG03-87ER25037.

1 Introduction

1.1 Background

Least squares estimations have been used extensively in a wide variety of scientific applications, for instance equalizations [13, p.139], system identifications [22], adaptive signal processing [1, p.87] and speech processing [13, p.343]. In these applications, we usually need to estimate the transmitted signal from a sequence of received signal samples or to model an unknown system by using a linear system model.

To present the problem properly, let us introduce some terminologies used in signal processing. Let x_i be a discrete-time stationary zero-mean complex-valued process (see Fuller [12, pp.10-11] for definition). A linear predictor of order n is of the form

$$\hat{x}_i = \sum_{k=1}^{n} b_k x_{i-k}$$

where $\hat{x}_i$ is the predicted value of x_i based on the data $\{x_k\}_{k=i-1}^{i-n}$ and $\{b_k\}_{k=1}^{n}$ are the predictor coefficients. The difference between the actual value x_i of the process and the predicted value $\hat{x}_i$ is called the prediction error of order n. Since we are interested in predicting the current value of the process based on the previous measurements, the predictor coefficients $\{b_i\}_{i=1}^{n}$ should be chosen to make the prediction error as small as possible.

Usually the predictor coefficients are determined by minimizing the mean square error, i.e. minimizing the prediction error in the least squares sense. The optimal least squares predictor coefficients are then given by the solution of the linear system of equations

$$R_n \mathbf{b} = \mathbf{r}, \tag{1}$$

see Giordano and Hsu [13, pp.41-43]. Here R_n is an n-by-n Hermitian Toeplitz matrix given by

$$R_n = \begin{bmatrix} r_0 & \bar{r}_1 & \cdots & \bar{r}_{n-2} & \bar{r}_{n-1} \\ r_1 & r_0 & \bar{r}_1 & \ddots & \bar{r}_{n-2} \\ \vdots & r_1 & r_0 & \ddots & \vdots \\ r_{n-2} & \ddots & \ddots & \ddots & \bar{r}_1 \\ r_{n-1} & r_{n-2} & \cdots & r_1 & r_0 \end{bmatrix},$$

and $\mathbf{b}$ and $\mathbf{r}$ are vectors of the form $[b_1, b_2, \cdots, b_n]^T$ and $[r_1, r_2, \cdots, r_n]^T$. The entries r_j are the autocovariances of the discrete-time stationary process and are given by

$$r_j = \mathcal{E}[x_i \bar{x}_{i-j}]$$

where $\mathcal{E}$ is the expectation operator.

The matrix R_n is called the covariance matrix of the stationary process and the Toeplitz system (1) is commonly called the Yule-Walker equation, see Yule [36]. We note that if the second-order statistics of the process is known, i.e. the autocovariances r_j of the stationary process are given, then the predictor coefficients $\{b_k\}_{k=1}^n$ can be found by solving (1). Several direct methods [20, 3] have been derived to solve such systems and their complexities vary from $O(n^2)$ to $O(n \log^2 n)$ operations.

We note that in practical cases, no prior knowledge is usually available on the autocovariances of the process. If M data samples have been taken, then all the information we have is contained in the finite number of data points $\{x_k\}_{k=1}^M$. In this case, we can still formulate a well-defined least squares prediction problem by estimating the autocovariances from the data samples $\{x_k\}_{k=1}^M$ with various types of windowing methods; such as the *correlation*, *covariance*, *pre-windowed* and *post-windowed* methods, see for instance Giordano and Hsu [13, pp.65-66]. The least squares estimators can then be found by solving the n-vector $\mathbf{b}$ in

$$\min ||T\mathbf{b} - \mathbf{y}||_2. \tag{2}$$

Here $\mathbf{y}$ is an M-vector, $||\cdot||_2$ denotes the usual Euclidean norm and T is an M-by-n complex Toeplitz matrix with full column rank n, obtained by applying various types of windowing methods on the data samples $\{x_k\}_{k=1}^M$.

The solution $\mathbf{b}$ of (2) can be obtained by solving the normal equation

$$(T^*T)\mathbf{b} = T^*\mathbf{y}. \tag{3}$$

We note that if the correlation method is employed, the normal matrix T^*T is Toeplitz. The other three windowing methods will lead to non-Toeplitz normal matrix T^*T. However, by exploiting the structure of T^*T, some recursive algorithms of complexity $O(M^2)$ have been developed, see Marple [22]. In addition to normal equation approach, orthogonalization schemes of complexity $O(Mn)$ have also been proposed, see for instance Itakura and Saito [17], Lee *et. al.* [21], Cybenko [10] and Qiao [27].

1.2 Iterative Methods For Toeplitz Systems

More recently, the use of preconditioned conjugate gradient method as an iterative method for solving Toeplitz systems $A_n\mathbf{u} = \mathbf{z}$ has been gaining attentions. The idea is to use circulant matrices S_n to precondition Toeplitz systems so as to speed up the convergence rate of the method, see Strang [31]. That means, instead of solving the original Toeplitz system, we solve the preconditioned system

$$S_n^{-1}A_n\mathbf{u} = S_n^{-1}\mathbf{z}$$

by the conjugate gradient method.

Since circulant matrices can always be diagonalized by the Fourier matrix, see Davis [11, p.74], the matrix-vector multiplication $S_n\mathbf{v}$ can be computed easily by the Fast Fourier Transforms (FFTs) in $O(n\log n)$ operations. For $A_n\mathbf{v}$, it can also be computed by FFTs in $O(n\log n)$ operations by first embedding A_n into a $2n$-by-$2n$ circulant matrix, see Strang [31]. It follows that the operations per iteration is of order $O(n\log n)$. The convergence rate of the method has been analyzed by Chan and Strang [5]. They proved that if the diagonals of the Toeplitz matrix A_n are Fourier coefficients of a positive function in the Wiener class, then the spectrum of the preconditioned system $S_n^{-1}A_n$ will be clustered around 1 for large n and the method will converge superlinearly. More precisely, for all $\epsilon > 0$, there exists a constant $c(\epsilon) > 0$ such that the error vector $\mathbf{e_j}$ of the preconditioned conjugate gradient method at the jth iteration satisfies

$$||\mathbf{e_j}||_{S_n^{-1/2}A_nS_n^{-1/2}} \leq c(\epsilon)\epsilon^j||\mathbf{e_0}||_{S_n^{-1/2}A_nS_n^{-1/2}}$$

when n is sufficiently large. Here $||\mathbf{v}||^2_{S_n^{-1/2}A_nS_n^{-1/2}} = \mathbf{v}^*{S_n}^{-1/2}A_n{S_n}^{-1/2}\mathbf{v}$. Hence the complexity of solving a large class of Toeplitz systems can be reduced to $O(n\log n)$ operations.

We remark that circulant approximations to Toeplitz matrices have been considered and used for some time in image processing (e.g. [2]), signal processing (e.g. [24] and [30, pp.75-86]) and time series analysis (e.g. [4, pp.130-131] and [29]). Besides Strang's circulant preconditioner S_n, several other successful circulant preconditioners have been proposed and analyzed, see [9, 16, 19, 32, 33]. Recently, the use of circulant preconditioners for Toeplitz least squares problems was considered by Plemmons and Nagy [23] and Chan, Nagy and Plemmons [8]. They established formal convergence results for the least squares problems and derived some applications in image processing. We remark however that our circulant preconditioner is different from that presented in [23, 8].

1.3 Outline

In this paper, we use the preconditioned conjugate gradient method with circulant preconditioners to solve the systems (1) and (3). For the case of known statistics, our proposed circulant preconditioners are constructed from the spectral density functions of the given discrete-time stationary processes. Using results in [6, 7] straightforwardly, we show that the spectrum of the preconditioned matrix is clustered around 1. Hence if our method is used to solve the Yule-Walker equation (1), then the convergence rate will be superlinear.

For the case of unknown statistics, only a finite number of data measurements from the random process are provided and the convergence analysis must therefore be considered probabilistically. The first thing we do then is to estimate the autocovariances of the given process. Four different windowing methods for estimating these autocovariances are introduced. Our circulant preconditioner C_n is constructed from these estimates and can be generated in $O(M \log n)$ operations where M is the number of data measurements taken. We prove that if the underlying spectral density function of the stationary process is positive and in the Wiener class, then our circulant preconditioner will be positive definite and its smallest eigenvalue will be uniformly bounded away from zero with probability 1, provided that sufficiently large number of data samples are taken. Under the same assumptions, we also prove that the spectrum of the preconditioned matrix $C_n^{-1}(T^*T)$ is clustered around 1 with probability 1. Thus, when we applied conjugate gradient method to the preconditioned system, the method converges superlinearly with probability 1.

As for the cost of our method, since the data matrices T is an M-by-n rectangular Toeplitz matrix, the normal equation and the circulant preconditioner can be formed in $O(M \log n)$ operations. Once they are formed, the cost per iteration of the preconditioned conjugate gradient method will be $O(n \log n)$ operations. Therefore the total work of obtaining the predictor coefficients to a given accuracy is of order $O((M+n)\log n)$.

The outline of the paper is as follows. In §2, we recall some useful results in iterative method for solving Toeplitz systems and apply them to the case where the second-order statistics are known. In §3, we consider processes with unknown statistics. We first formulate the problem for finding the predictor coefficients as a least squares problem. Then we introduce our circulant preconditioner and analyze the convergence rate of our method probabilistically. In §4, numerical experiments are performed for processes with known and unknown statistics. Specifically, we test the performance of our method for the finite impulse response (FIR) system identification. Finally, concluding remarks are given in §5.

2 Results For Known Statistics

In this section, we consider discrete-time stationary process with known second-order statistics, i.e. known autocovariances. In this deterministic case, we can solve the Toeplitz system (1) to obtain the predictor coefficients $\{b_k\}_{k=1}^n$. The convergence rate of the method can be analyzed straightforwardly as in [6, 7] as we will now show.

To begin with, let the n-by-n Toeplitz matrices R_n in (1) be generated by a 2π-periodic continuous function f defined on $[-\pi, \pi]$, i.e. the (j, ℓ)th entry of R_n is given by the $(j-\ell)$th Fourier coefficient of f:

$$r_{j-\ell} = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(\theta) e^{i(j-\ell)\theta} d\theta.$$

The function f is called the generating function of R_n. For simplicity, we will denote an n-by-n Toeplitz matrix generated by a function f by the symbol $A_n[f]$. We note that if

f is real-valued, then $A_n[f]$ is Hermitian and the spectrum $\sigma(A_n[f])$ of $A_n[f]$ satisfies

$$\sigma(A_n[f]) \subseteq [f_{\min}, f_{\max}], \quad \forall n \geq 1, \tag{4}$$

where $f_{\min}$ and $f_{\max}$ are the minimum and maximum values of f respectively, see Grenander and Szegö [15, pp.63-65]. In particular, if f is positive, then $A_n[f]$ is positive definite.

For Toeplitz matrices generated from a function, there are many different choices of circulant preconditioners that can be constructed from its generating function, see Chan and Yeung [7]. In this paper, we only focus on T. Chan's circulant preconditioner $C_n[f]$ which is defined to be the minimizer of $||Q_n - A_n[f]||_F$ over all circulant matrices Q_n, see [9]. Here $||\cdot||_F$ denotes the Frobenius norm. The (j,ℓ)th entry of $C_n[f]$ is given by $c_{j-\ell}$ where

$$c_j = \begin{cases} \dfrac{(n-j)r_j + jr_{j-n}}{n} & 0 \leq j < n, \\ c_{n+j} & 0 < -j < n. \end{cases}$$

Chan and Yeung [7] showed that $C_n[f]$ is closely related to the Fejér kernel $\mathcal{F}_n$ (see Walker [34, p.76] for definition of Fejér kernel). Indeed, the eigenvalues $\lambda_j(C_n[f])$ of $C_n[f]$ are given by

$$\lambda_j(C_n[f]) = (\mathcal{F}_n * f)(\frac{2\pi j}{n}), \quad 0 \leq j < n, \tag{5}$$

where

$$(\mathcal{F}_n * f)(\theta) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \mathcal{F}_n(\theta - \phi) f(\phi) d\phi.$$

One of the interesting spectral property of $C_n[f]$ is that if $A_n[f]$ is positive definite, then $C_n[f]$ is also positive definite. In fact, we have

$$\lambda_{\min}(A_n[f]) \leq \lambda_{\min}(C_n[f]) \leq \lambda_{\max}(C_n[f]) \leq \lambda_{\max}(A_n[f]) \tag{6}$$

where $\lambda_{\min}(\cdot)$ and $\lambda_{\max}(\cdot)$ denote the minimum and maximum eigenvalues respectively, see Tyrtyshnikov [33]. Combining (6) with (4), we see that if the function f is positive, then $||C_n^{-1}[f]||_2$ is uniformly bounded. We remark that most of the other circulant preconditioners do not satisfy (6), see Chan and Yeung [7]. As for the performance of $C_n[f]$ as a preconditioner to $A_n[f]$, we have the following theorem.

Theorem 1 (Chan and Yeung [6, Theorem 1,Corollary 3]) *Let f be a positive 2π-periodic continuous function. Then for all $\epsilon > 0$, there exist positive integers K and N such that for all $n > K$, at most N eigenvalues of $C_n[f] - A_n[f]$ and of $C_n^{-1}[f]A_n[f] - I_n$ have absolute value greater than ϵ.*

Thus the spectrum of $C_n^{-1}[f]A_n[f]$ is clustered around 1 and therefore the conjugate gradient method, when applied to the preconditioned system $C_n^{-1}[f]A_n[f]\mathbf{u} = \mathbf{z}$, will converge superlinearly, see Chan and Yeung [6].

We now apply the result above to the solution of (1). We begin by noting that for a discrete-time stationary process, if the autocovariances of the process are absolutely summable, i.e. $\sum_{k=-\infty}^{\infty} |r_k| < \infty$, then r_k can be expressed in the form

$$r_k = \int_{-\pi}^{\pi} s(\theta) e^{ik\theta} d\theta$$

where $s(\theta)$, called the spectral density function of the stationary process, is given by

$$s(\theta) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} r_k e^{-ik\theta},$$

see [4, p.118]. We note that the covariance matrix R_n is a Toeplitz matrix generated by $s(\theta)$. As examples, we consider the following stationary processes:

1. *Purely random process (White Noise)* [26, p.233]: The process simply consists of a sequence of uncorrelated random variables $\{v_t\}$ and the autocovariances are given by

   $$r_k = \begin{cases} \eta^2, & k = 0, \\ 0, & \text{otherwise}, \end{cases}$$

   where η^2 is the variance of the random variable v_t. The corresponding spectral density function $s(\theta)$ is given by

   $$s(\theta) = \frac{\eta^2}{2\pi}, \quad \forall \theta \in [-\pi, \pi]. \tag{7}$$

   Thus, the covariance matrix is a constant multiple of the identity matrix.

2. *First order auto-regressive process AR(1)* [26, p.238]: The process is given by

   $$x_t = \rho x_{t-1} + v_t,$$

   where $\{v_t\}$ is a white noise process with variance η^2. The autocovariances of the process are given by

   $$r_k = \frac{\eta^2 \rho^{|k|}}{1-\rho^2}, \quad k = 0, \pm 1, \pm 2, \cdots, \tag{8}$$

   where $|\rho| < 1$. The corresponding spectral density function $s(\theta)$ is given by

   $$s(\theta) = \frac{\eta^2}{2\pi(1 - 2\rho\cos\theta + \rho^2)}, \quad \forall \theta \in [-\pi, \pi],$$

and the covariance matrix is a scalar multiple of the Kar-Murdock Szegö matrix, see Kac and Murdock [18].

3. *Second order auto-regressive process AR(2)* [26, p.241]: The process is given by

$$x_t + \tau_1 x_{t-1} + \tau_2 x_{t-2} = v_t,$$

where $\{v_t\}$ is a white noise process with variance η^2. The autocovariances of the process are given by

$$r_k = \frac{[(1-\delta_2^2)\delta_1^{|k|+1} - (1-\delta_1^2)\delta_2^{|k|+1}]\eta^2}{(\delta_1-\delta_2)(1-\delta_1\delta_2)[(1-\delta_1\delta_2)^2-(\delta_1+\delta_2)^2]}, \quad k = 0, \pm 1, \pm 2, \cdots, \tag{9}$$

where $\tau_1 = -(\delta_1+\delta_2)$ and $\tau_2 = \delta_1\delta_2$ such that $|\delta_1| < 1$ and $|\delta_2| < 1$. The spectral density function is given by

$$s(\theta) = \frac{\eta^2}{2\pi[(1+\tau_2)^2 + \tau_1^2 - 2\tau_1(1-\tau_2)\cos\theta - 4\tau_2\cos^2\theta]}, \quad \forall\theta \in [-\pi, \pi].$$

4. *First order moving-average process MA(1)* [4, p.121]: The process is given by

$$x_t = v_t + \chi v_{t-1},$$

where $|\chi| < 1$ and $\{v_t\}$ is a white noise process with variance η^2. The autocovariances of the process are given by

$$r_k = \begin{cases} \eta^2(1+\chi^2), & k = 0, \\ \eta^2\chi, & k = 1, \\ 0, & \text{otherwise.} \end{cases}$$

We see that the covariance matrix is a tridiagonal Toeplitz matrix $A_n[s]$ with

$$s(\theta) = \frac{\eta^2}{2\pi}(1 + 2\chi\cos\theta + \chi^2), \quad \forall\theta \in [-\pi, \pi].$$

If we assume that the spectral density function of the stationary process exists and satisfies the hypothesis of Theorem 1, then the Yule-Walker equation (1) can be solved in $O(n\log n)$ operations by using the preconditioned conjugate gradient method with circulant preconditioner $C_n[f]$. More precisely, we have the following direct consequence of Theorem 1.

Corollary 1 *Let the spectral density function $s(\theta)$ of a discrete-time stationary process be a positive 2π-periodic continuous function. Then for all $\epsilon > 0$, there exist positive integers K and N such that for $n > K$, at most N eigenvalues of $I_n - C_n^{-1}[s]A_n[s]$ have absolute value greater than ϵ.*

We remark that all the above results are derived deterministically. In the least squares estimation algorithms discussed below, we deal with data samples from random processes and the convergence rate will be considered in a probabilistic way.

3 Least Square Solutions With Unknown Statistics

In this section, we consider the more practical case where no prior knowledge on the autocovariances of the discrete-time stationary process is available. In this case, the autocovariances are estimated from the finite number of data samples $\{x_1, x_2, ..., x_M\}$. The usual approach is to formulate the prediction problem as a least square problem by using various types of windowing methods. In §3.1, we will consider four of these windowing methods. The construction of our circulant preconditioner will be given in §3.2 and the convergence rate will be analyzed in §3.3.

3.1 Windowing Methods

Let $\{x_1, \cdots, x_M\}$ be the set of data samples taken. By minimizing the mean square error over the available data, the least squares estimation of the predictor coefficients $\{b_k\}$ can be found by solving the least squares problem

$$\min ||T_w \mathbf{b} - \mathbf{y}||_2. \tag{10}$$

Here $\mathbf{y}$ is a known M-vector and T_w is a data matrix, see Giordano and Hsu [13, pp. 65-66]. The exact form of T_w depends on the assumptions we make to the data outside our observation.

(W1) *Correlation method* assumes that data prior to $k = 0$ and after $k = M$ are zero. The corresponding data matrix is an $(M + n - 1)$-by-n rectangular Toeplitz matrix of the form

$$T_1 = \begin{bmatrix} x_1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & 0 \\ x_n & \cdots & \cdots & x_1 \\ \vdots & & & \vdots \\ \vdots & & & \vdots \\ x_M & \cdots & \cdots & x_{M-n-1} \\ 0 & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & x_M \end{bmatrix}.$$

(W2) *Covariance method* makes no assumptions about the data when $k = 0$ or $k = M$. The corresponding data matrix is an $(M - n + 1)$-by-n rectangular Toeplitz matrix given by

$$T_2 = \begin{bmatrix} x_n & \cdots & \cdots & x_1 \\ \vdots & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ \vdots & & & x_n \\ \vdots & & & \vdots \\ x_M & \cdots & \cdots & x_{M-n+1} \end{bmatrix}.$$

(W3) *Pre-windowed method* assumes that data prior to $k = 0$ are zero but makes no assumptions about data after $k = M$. The $(M + 1)$-by-n data matrix T_3 is given by

$$T_3 = \begin{bmatrix} x_n & \cdots & \cdots & x_1 \\ \vdots & & & \vdots \\ \vdots & & & \vdots \\ x_M & \cdots & \cdots & x_{M-n+1} \\ 0 & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & x_M \\ 0 & \cdots & \cdots & 0 \end{bmatrix}.$$

(W4) *Post-windowed method* assumes that data after $k = M$ are zero but makes no assumptions about data prior to $k = 0$. Thus the $(M+1)$-by-n data matrix T_4 is given by

$$T_4 = \begin{bmatrix} 0 & \cdots & \cdots & 0 \\ x_1 & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & 0 \\ x_n & \cdots & \cdots & x_1 \\ \vdots & & & \vdots \\ \vdots & & & \vdots \\ x_M & \cdots & \cdots & x_{M-n+1} \end{bmatrix}.$$

In all four cases, the least squares solutions to (10) can be obtained by solving the scaled normal equations

$$\frac{1}{M}(T_w^* T_w)\mathbf{b} = \frac{1}{M} T_w^* \mathbf{y}, \quad w = 1, 2, 3, 4. \tag{11}$$

We note that when employing method (W1), the normal matrix $\frac{1}{M}(T_1^* T_1)$ is a Hermitian Toeplitz matrix and can be written in the form

$$\frac{1}{M}(T_1^* T_1) = A_n[g] \tag{12}$$

where

$$g(\theta) = \sum_{k=-(n-1)}^{n-1} \hat{r}_k e^{-ik\theta}, \quad \forall \theta \in [-\pi, \pi] \tag{13}$$

and

$$\hat{r}_k = \frac{1}{M} \sum_{j=1}^{M-|k|} x_j \bar{x}_{j+|k|}. \tag{14}$$

In statistics literature, $\hat{r}_k$ is called an estimator of the autocovariance r_k of the process.

When using the other windowing methods (W2), (W3) and (W4), the normal matrices can be written in the following forms:

For *covariance method* (W2):

$$\frac{1}{M}(T_2^* T_2) = A_n[g] - A_n^*[p] A_n[p] - A_n^*[q] A_n[q]; \tag{15}$$

For *pre-windowed method* (W3):

$$\frac{1}{M}(T_3^* T_3) = A_n[g] - A_n^*[p]A_n[p]; \tag{16}$$

For *post-windowed method* (W4):

$$\frac{1}{M}(T_4^* T_4) = A_n[g] - A_n^*[q]A_n[q]. \tag{17}$$

Here

$$p(\theta) = \sum_{j=1}^{n-1} \frac{x_j}{\sqrt{M}} e^{-ij\theta}, \quad \forall \theta \in [-\pi, \pi],$$

and

$$q(\theta) = \sum_{j=1}^{n-1} \frac{x_{M-j+1}}{\sqrt{M}} e^{ij\theta}, \quad \forall \theta \in [-\pi, \pi].$$

We note that $A_n[p]$ and $A_n[q]$ are lower and upper triangular Toeplitz matrices respectively. As the product of a lower triangular Toeplitz matrix and an upper triangular Toeplitz matrix is not Toeplitz in general, the normal matrices in these cases are non-Toeplitz.

3.2 Construction of Circulant Preconditioner

Let us now generate our circulant preconditioner from the normal matrices $\frac{1}{M}(T_w^* T_w)$ where $w = 1, 2, 3$ and 4. As we can always pad zeros to the bottom rows of the data matrices T_w, we assume here without loss of generality that we can partition T_w as

$$T_w = \begin{bmatrix} T_{(w,1)} \\ T_{(w,2)} \\ \vdots \\ T_{(w,m)} \end{bmatrix}, \tag{18}$$

where each $T_{(w,j)}$ is an n-by-n Toeplitz matrix and m is the number of blocks of n-by-n Toeplitz matrices. Our preconditioner $C_n[g]$ is taken to be the circulant approximation of the Toeplitz part $A_n[g]$ of the normal matrix $\frac{1}{M}(T_w^* T_w)$, see (12), (15)-(17).

Recall that by using FFTs, the cost of matrix-vector multiplications involving the matrix T_w, $w = 1, 2, 3, 4$, can be done in $O(M \log n)$ operations whereas those involving $A_n[p]$ and $A_n[q]$ can be performed in $O(n \log n)$ operations. Hence the Toeplitz matrix $A_n[g]$ can be found in $O(M \log n)$ operations whereas $C_n[g]$ can be found in $O(n \log n)$

operations. Once the first column of $A_n[g]$ has been computed and stored, the cost for each iteration of the preconditioned conjugate gradient method will be of $O(n \log n)$ operations. As for the storage, we need an M-vector to store the set of data samples $\{x_k\}_{k=1}^M$ and five n-vectors in the conjugate gradient method. The diagonals of $A_n[g]$ and the first column of $C_n[g]$ will require another two n-vectors. If the diagonals of $A_n[p]$ and $A_n[q]$ are needed, extra two n-vectors will be required. Thus the overall storage requirement is about $O(M + n)$.

We remark that our circulant preconditioner is different from that recently proposed by Chan, Nagy and Plemmons for Toeplitz least squares problems [8]. They basically take the circulant approximation of each Toeplitz block $T_{(w,j)}$ in (18) and then combine them together to form a circulant preconditioner. The motivation behind our preconditioner is that the Toeplitz matrix $A_n[g]$ is the sample covariance matrix which intuitively should be a good estimation to the covariance matrix R_n of the discrete-time stationary process, provided that sufficiently large number of data samples are taken. Hence we choose to approximate $A_n[g]$ instead of $T_{(w,j)}$ by circulant preconditioners.

The analysis of the performance of $C_n[g]$ will be given later. We first explain why we choose the T. Chan circulant preconditioners $C_n[g]$ instead of the others. We recall that the eigenvalues of $C_n[g]$ are given by $(\mathcal{F} * g)(2\pi j/n)$, see (5). In the deterministic case, $C_n[f]$ is a good preconditioner for $A_n[f]$ because $\mathcal{F}_n * f$ is a good approximation of f, see Chan and Yeung [7]. In the current stochastic case, the following Lemma can serve as a motivation for choosing $C_n[g]$.

Lemma 1 (Grenander and Rosenblatt [14, pp.262-263]) *Let the spectral density function $s(\theta)$ of the discrete-time stationary process be real-valued with bounded second derivative and $g(\theta)$ be given by (13). Then for any given $\epsilon > 0$, there exists a positive integer N such that for $n > N$,*

$$\Pr\{||\mathcal{F}_n * g - s||_\infty < \epsilon\} > 1 - \epsilon,$$

provided that the data samples size M is sufficiently large enough ($M \gg n$). Here $||\cdot||_\infty$ is the supremum norm.

The Lemma basically states that the convolution product $\mathcal{F}_n * g$ converges to the spectral density function $s(\theta)$ in probability. Therefore, we expect $C_n[g]$ to be a good preconditioner for $A_n[g]$.

3.3 Probabilistic Analysis of the Convergence Rate

As we deal with data samples from random processes, the convergence rate will be considered in a probabilistic way which is different from the deterministic case discussed in

§2. We first make the following assumption (**A**) on the discrete-time stationary process so that results of the convergence rate can be derived.

(**A1**) The underlying spectral density function $s(\theta)$ of the process is positive and in the Wiener class, i.e. the autocovariances of the process are absolutely summable:

$$\sum_{k=-\infty}^{\infty} |r_k| \leq \alpha < \infty. \tag{19}$$

(**A2**) The variances of the estimators $\hat{r}_k$ given in (14) are bounded by

$$\mathrm{Var}(\hat{r}_k) \leq \frac{\beta}{M}, \quad k = 0, \pm 1, \pm 2, \cdots, \tag{20}$$

where β is a constant.

(**A3**) The stationary process has zero-mean, i.e. $\mathcal{E}(x_i) = \mu = 0$ for all i.

Some remarks on the assumptions:

1. In time-series analysis, assumption (**A1**) is often valid. For example, the spectral density functions of autoregressive-moving average (ARMA) processes are rational functions [4, p.121]. The positiveness of the spectral density function can be guaranteed by the causality of the process [4, p.85] whereas the absolutely summability of the autocovariances can be assured by the invertibility of the process [4, p.86].

2. Assumption (**A2**) is satisfied when the stationary process is Gaussian (see Priestley [26, p.113] for definition). In fact, in this case, the variances of the estimators $\hat{r}_k$ are given by

$$\mathrm{Var}(\hat{r}_k) = \frac{1}{M} \sum_{j=-(M-k)+k}^{M-k-1} \left(1 - \frac{|j|+k}{M}\right)(r_j^2 + r_{j+k} r_{j-k}), \quad k = 0, \pm 1, \pm 2, \cdots.$$

   As the autocovariances of the process are absolutely summable, inequality (20) is satisfied.

3. If the mean of the stationary process is not equal to zero, then we can consider the stationary process $\{x_i - \mu\}$ instead. Even if μ is unknown, we can estimate it by the sample mean.

4. Under assumption (**A3**), we have

$$\mathcal{E}(\hat{r}_k) = (1 - \frac{|k|}{M})r_k, \quad \forall k \geq 0. \tag{21}$$

see Priestley [26, p.323]. Although the formula of $\mathcal{E}(\hat{r}_k)$ is slightly different when μ is unknown, they are almost the same when a large number of data samples are taken, see [26, p.323].

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

Lemma 2 *Let the discrete-time stationary process satisfy assumption (20). Then for any* $\epsilon > 0$,

$$\Pr\{|\hat{r}_k - \mathcal{E}(\hat{r}_k)| > \epsilon\} \leq \frac{\mathrm{Var}(\hat{r}_k)}{\epsilon^2} \leq \frac{\beta}{M\epsilon^2}.$$

Proof: The first inequality comes from Chebyshev's inequality, see Fuller [12, p.185] and the second inequality is obtained by applying (20). □

Before going into the convergence analysis, we define the function $g_E(\theta)$ which is an approximation to the function $g(\theta)$ in (13):

$$g_E(\theta) = \sum_{k=-(n-1)}^{n-1} \mathcal{E}(\hat{r}_k)e^{-ik\theta}, \quad \forall \theta \in [-\pi, \pi]. \tag{22}$$

The following Lemma gives an estimate of the difference between g_E and g in the supremum norm.

Lemma 3 *Let g and g_E be given by (13) and (22) respectively. Then*

$$\Pr\{||g - g_E||_\infty > \epsilon\} \leq \frac{8\beta n^3}{M\epsilon^2}.$$

Proof: By using a Lemma in Fuller [12, p.182], we have

$$\begin{aligned}\Pr\{||g - g_E||_\infty \geq \epsilon\} &= \Pr\{||\sum_{k=-(n-1)}^{n-1} [\hat{r}_k - \mathcal{E}(\hat{r}_k)]e^{-ik\theta}||_\infty \geq \epsilon\} \\ &\leq \sum_{k=-(n-1)}^{n-1} \Pr\{|\hat{r}_k - \mathcal{E}(\hat{r}_k)| > \frac{\epsilon}{2n-1}\}.\end{aligned}$$

The result now follows by using Lemma 2. □

3.3.1 Correlation Windowing Method

In this subsection, we analyze the spectrum of the preconditioned matrix $C_n^{-1}[g]A_n[g]$ when the *correlation* windowing method (W1) is used. We first prove that the smallest eigenvalue of $A_n[g]$ is uniformly bounded away from zero with probability 1.

Theorem 2 *Let the discrete-time stationary process satisfy assumption* (**A**). *Then for any given $\epsilon > 0$, there exists a positive integer N such that for $n > N$,*

$$\Pr\{\lambda_{\min}(A_n[g]) \text{ is uniformly bounded away from zero}\} > 1 - \epsilon,$$

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

Proof: We first write

$$A_n[g] = \{A_n[g] - A_n[g_E]\} + \{A_n[g_E] - A_n[s]\} + A_n[s].$$

By (**A1**) and (4), we have

$$\lambda_{\min}(A_n[s]) \geq s_{\min} > 0, \tag{23}$$

where $s_{\min}$ is the minimum value of s. Therefore, it suffices to estimate $||A_n[g] - A_n[g_E]||_2$ and $||A_n[g_E] - A_n[s]||_2$ respectively.

For the probabilistic part, i.e. the matrix $A_n[g] - A_n[g_E]$, we note by (4) that

$$\Pr\{||g - g_E||_\infty < \frac{\epsilon}{2}\} \leq \Pr\{||A_n[g] - A_n[g_E]||_2 < \frac{\epsilon}{2}\} \leq 1. \tag{24}$$

On the other hand, by Lemma 2,

$$\Pr\{||g - g_E||_\infty < \frac{\epsilon}{2}\} > 1 - \frac{32\beta n^3}{M\epsilon^2},$$

for sufficiently small $\epsilon > 0$. Thus, it follows from (24) that

$$\Pr\{||A_n[g] - A_n[g_E]||_2 < \frac{\epsilon}{2}\} > 1 - \epsilon, \tag{25}$$

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

For the deterministic part, i.e. the matrix $A_n[g_E] - A_n[s]$, we note by (4), (21) and (22) that

$$||A_n[g_E] - A_n[s]||_2 \leq ||g_E - s||_\infty \leq \sum_{k=-(n-1)}^{n-1} \frac{|k|}{M}|r_k| + \sum_{|k|>n} |r_k|. \tag{26}$$

Using (19), it follows that for any given $\epsilon > 0$, there exist positive integers N_1 and $M_1 > N_1$ such that

$$\sum_{|k| \geq N_1} |r_k| < \frac{\epsilon}{8} \tag{27}$$

and

$$\frac{1}{M_1} \sum_{k=-(N_1-1)}^{N_1-1} |k||r_k| < \frac{\epsilon}{4}. \tag{28}$$

Hence for all $M > n > M_1$, by (27) and (28), we have

$$\sum_{k=-(n-1)}^{n-1} \frac{|k|}{M}|r_k| < \frac{1}{M_1} \sum_{k=-(N_1-1)}^{N_1-1} |k||r_k| + \sum_{|k|>N_1}^{n-1} \frac{|k|}{M}|r_k| < \frac{3\epsilon}{8}.$$

Putting this bound and (27) back into (26), we get

$$||A_n[g_E] - A_n[s]||_2 < \frac{\epsilon}{2}. \tag{29}$$

The lemma now follows by combining (23), (25), (29) and using simple probability arguments. $\square$

Combining Theorem 2 with (6), we immediately have the following corollary on the smallest eigenvalue of $C_n[g]$.

Corollary 2 *Let the discrete-time stationary process satisfy assumption* (**A**). *Then for any given $\epsilon > 0$, there exists a positive integer N such that for $n > N$,*

$$Pr\{\lambda_{\min}(C_n[g]) \textit{ is uniformly bounded away from zero}\} > 1 - \epsilon,$$

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

Next we prove the clustering property of the preconditioned matrices $C_n^{-1}[g]A_n[g]$.

Theorem 3 *Let the discrete-time stationary process satisfy assumption* (**A**). *Then for all $\epsilon > 0$, there exist positive integers K and N such that for $n > N$,*

$$\mathbf{Pr}\ \{\textit{at most } K \textit{ eigenvalues of } C_n[g] - A_n[g] \textit{ have absolute value greater than } \epsilon\} > 1 - \epsilon,$$

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

Proof: We write

$$C_n[g] - A_n[g] = \{C_n[g] - C_n[s]\} + \{C_n[s] - A_n[s]\} + \{A_n[s] - A_n[g]\}.$$

In view of Theorem 1, the eigenvalues of $C_n[s] - T_n[s]$ will be clustered around zero. Hence by applying Cauchy's interlace theorem (see Wilkinson [35, p.103]), it suffices to prove that $||C_n[g] - C_n[s]||_2$ and $||A_n[s] - A_n[g]||_2$ are very small with probability 1. For the difference $A_n[s] - A_n[g]$, we first write it as

$$A_n[s] - A_n[g] = \{A_n[s] - A_n[g_E]\} + \{A_n[g_E] - A_n[g]\}.$$

Then by using arguments similar to those used in Theorem 2, we can prove that

$$\Pr\{||A_n[s] - A_n[g]||_2 < \epsilon\} > 1 - \epsilon,$$

provided that $M = O(n^{3+\nu})$ with $\nu > 0$. By (6), if $||A_n[s] - A_n[g]||_2 < \epsilon$, then we have $||C_n[s] - C_n[g]||_2 < \epsilon$. Now the theorem follows by using simple probability arguments. □

Combining Corollary 2 and Theorem 3, we have the following main theorem about the spectra of the preconditioned system.

Theorem 4 *Let the discrete-time stationary process satisfy assumption* (**A**). *Then for any given $\epsilon > 0$, there exist positive integers K and N such that for $n > N$,*

$$\Pr\ \{\textit{at most } K \textit{ eigenvalues of } I_n - C_n^{-1}[g]A_n[g] \textit{ have absolute value greater than } \epsilon\} > 1 - \epsilon,$$

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

Proof: Let us define the following events:
$E_1 = \{$at most K eigenvalues of $C_n[g] - A_n[g]$ have absolute value greater than $\epsilon\}$,
$E_2 = \{\lambda_{\min}(C_n[g])$ is uniformly bounded away from zero$\}$, and
$E_3 = \{$at most K eigenvalues of $I_n - C_n^{-1}[g]A_n[g]$ have absolute value greater than $\epsilon\}$.
By Theorem 3 and Corollary 2, we see that

$$\Pr\{E_1 \text{ and } E_2\} = \Pr\{E_1\} + \Pr\{E_2\} - \Pr\{E_1 or E_2\} \geq 1 - 2\epsilon.$$

Since events E_1 and E_2 together imply E_3, the theorem follows. □

Using Theorem 4, we can easily show that the conjugate gradient method, when applied to the preconditioned system $C_n^{-1}[g]A_n[g]$, converges superlinearly with probability 1 provided that $M = O(n^{3+\nu})$ with $\nu > 0$. For details of the proof of the superlinearly convergence rate, see Chan and Strang [5].

3.3.2 Other Windowing Methods

To derive the convergence rate for other windowing methods, we first note the following result:

Lemma 4 *Let the variance of the discrete-time stationary process be equal to ξ^2, i.e. $r_0 = \xi^2$. Then for any given $\epsilon > 0$, there exists a positive integer N such that for $n > N$,*

$$\Pr\{||A_n[p]||_2 \leq \epsilon\} > 1 - \epsilon,$$

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

Proof: By (4), $||A_n[p]||_2 \leq 2||p||_\infty$. Thus

$$\Pr\{||p||_\infty \leq \epsilon\} \leq \Pr\{||A_n[p]||_2 \leq \epsilon\} \leq 1.$$

As the variance of x_j is equal to ξ^2, it follows by the Lemma in Fuller [12, p.182] and Chebyshev's inequality [12, p.185] that

$$\Pr\{||p||_\infty \geq \epsilon\} = \Pr\{||\sum_{j=1}^{n-1} \frac{x_j}{\sqrt{M}} e^{-ij\theta}||_\infty \geq \epsilon\} \leq \sum_{j=1}^{n-1} \Pr\{|\frac{x_j}{\sqrt{M}}| \geq \frac{\epsilon}{n}\} \leq \frac{n^3\xi^2}{M\epsilon^2}.$$

Hence, the result follows. $\square$

Following the arguments in Lemma 4, we can establish similar results for the upper triangular Toeplitz matrix $A_n[q]$. Thus, combining with Theorem 3 and using Cauchy's interlace theorem, we can prove that the spectra of the matrices

(i) $C_n[g] - A_n[g] - A_n^*[p]A_n[p] - A_n^*[q]A_n[q]$ (*covariance method* (W2)),

(ii) $C_n[g] - A_n[g] - A_n^*[p]A_n[p]$ (*pre-windowed method* (W3)) and

(iii) $C_n[g] - A_n[g] - A_n^*[q]A_n[q]$ (*post-windowed method* (W4))

are clustered around zero with probability 1 provided that $M = O(n^{3+\nu})$ with $\nu > 0$. To sum up, we have the following main result.

Theorem 5 *Let the discrete-time stationary process satisfy assumption* **(A)**. *Then for all $\epsilon > 0$ and for each $w = 1, 2, 3, 4$, there exist positive integers K and N such that for $n > N$, the probability that at most K eigenvalues of*

$$I_n - C_n^{-1}[g](\frac{1}{M}T_w^*T_w)$$

have absolute value greater than ϵ is greater than $1 - \epsilon$, provided that $M = O(n^{3+\nu})$ with $\nu > 0$.

According to Theorem 5, the preconditioned conjugate gradient method with circulant preconditioner $C_n[g]$ is an efficient algorithm for solving Toeplitz least-square equations derived from different kinds of windowing methods.

4 Numerical Experiments

In this section, numerical experiments are performed to test the convergence performance of the algorithm. Stationary processes with known or unknown second-order statistics (i.e. autocovariances) are considered. All the computations are done by Matlab on a Sparc II workstation at UCLA. In the numerical tests, we use the zero vector as our initial guess and the stopping criterion is $||\mathbf{e_j}||_2/||\mathbf{e_0}||_2 < 10^{-7}$, where $\mathbf{e_j}$ is the residual vector after j iterations. In the tables below, I_n denotes no preconditioner was used whereas C_n signifies T. Chan circulant preconditioner was used.

4.1 Known Statistics

We test our method for first and second order autoregressive processes, i.e. AR(1) and AR(2) processes. Their autocovariances are given by (8) and (9) respectively. We solve the corresponding Yule-Walker equation (1). For each process, two sets of parameters were tried. We note that the spectral density functions $s(\theta)$ of the processes are positive and in the Wiener class. Table 1 gives the number of iterations required to solve (1). From the table, we see that the number of iterations increases for the original matrices as n increases. However, it stays almost the same for the preconditioned systems.

	AR(1) process				AR(2) process			
	$\rho = 0.3$		$\rho = 0.9$		$\delta_1 = 0.1, \delta_2 = 0.5$		$\delta_1 = 0.9, \delta_2 = 0.5$	
n	I_n	C_n	I_n	C_n	I_n	C_n	I_n	C_n
8	8	6	9	5	8	6	9	8
16	12	6	16	6	15	7	21	11
32	14	5	25	6	21	6	38	12
64	14	4	40	7	27	5	68	11
128	14	4	61	6	29	5	121	9
256	14	4	85	6	29	4	198	9

Table 1. Number of iterations for AR(1) and AR(2) process with known statistics.

4.2 Unknown Statistics

We illustrate the convergence rate of our method by using finite impulse response (FIR) system identification as an example. FIR system identification has wide applications in engineering [22, 28]. Figure 1 is a block diagram of an FIR system identification model. The input signal x_k drives the unknown system to produce the output sequence y_k. We model the unknown system as an FIR filter. If the unknown system is actually an FIR system, then the model is exact.

In the tests, we formulate a well-defined least squares prediction problem by estimating the autocovariances from the data samples with *correlation* and *covariance* windowing methods. By solving the normal equations as discussed in §3.1, the FIR system coefficients can be found. We remark that if the autocovariances and the cross-covariances of the input process $\{x_k\}$ and the output process $\{y_k\}$ are known beforehand, then we are just simply solving a system of equations similar to (1), see Marple [22].

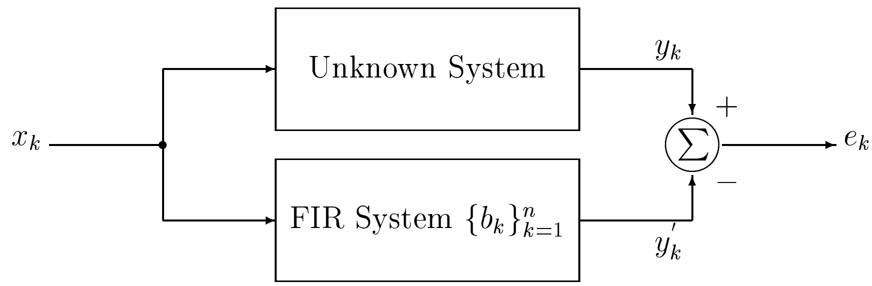

Figure 1. FIR System Identification Model

In the numerical tests, we used Gaussian white noise (purely random process) and colored noise (AR(1) and AR(2) processes) with variance η^2 equal to 1 as input processes. The reference (unknown) system is an n-th order linear phase FIR filter with uncorrelated Gaussian white noise added. The finite impulse response $\{h_k\}_{k=1}^{n}$ we used for the reference system is

$$h_k = 1.1 - \frac{|2k - n - 1|}{n - 1}.$$

We note that the shape of the FIR filter is triangular. Different variances of noise level are used to test the performance of the preconditioned conjugate gradient algorithm. In signal processing, the effect of the background noise to the signal is measured by the signal-to-noise ratio (SNR) which is defined as

$$\text{SNR} = 10 \log_{10} \left(\frac{\text{variance of the reference system output } \{y_k\}}{\text{variance of the additive noise}} \right).$$

In the tables below, $m = M/n$ is the number of blocks of n-by-n Toeplitz matrices in the matrix T_w (c.f. (18)). In Table 2, we first use white noise as input process and employ correlation windowing method to formulate the least square prediction problems. Table 2 shows the average number of iterations (rounded to the nearest integer) of the normal systems and of the preconditioned systems over 100 runs of the algorithm. From the numerical results, we see that the preconditioned systems converge very fast and the number of iterations required for convergence are less than that of the normal systems. However, the reduction of the number of iterations is not significant when m is large. This is because the spectral density function of the white noise process is a constant function, see (7). Hence the number of iterations are almost the same for white noise input process when m is large.

	SNR=50								SNR=30							
n	16		32		64		128		16		32		64		128	
m	I_n	C_n	I_n	C_n	I_n	C_n	I_n	C_n	I_n	C_n	I_n	C_n	I_n	C_n	I_n	C_n
2	12	11	17	14	28	14	36	16	12	9	20	15	27	15	34	16
4	11	9	12	9	16	12	23	12	11	9	16	12	22	14	20	14
8	11	7	11	9	12	9	17	11	10	8	13	9	14	11	17	11
16	9	6	10	8	10	8	13	10	9	7	10	9	12	10	13	10
32	9	6	9	7	9	8	10	9	7	7	8	8	9	8	12	9
64	6	5	8	7	7	7	9	7	7	6	8	6	7	7	10	8

Table 2. Average number of iterations for white noise input process when correlation windowing method is employed.

Tables 3-4 and 5-6 show the average number of iterations over 100 runs of the algorithms when AR(1) (with $\rho = 0.9$) and AR(2) (with $\delta_1 = 0.9$ and $\delta_2 = 0.5$) are used as the input processes respectively. We see that the preconditioned systems also converge very fast and the reduction in number of iterations is much greater than in the case of white noise input process.

In the proof of Lemma 4, we need M to be sufficiently large in order to make the ℓ_2 norm of the matrices $A_n[p]$ or $A_n[q]$ as small as possible. The fact can be seen from the numerical results in Tables 4 and 6 where the number of iterations of the preconditioned systems are greater than the non-preconditioned one when $m = 2$. However, when $m > 2$, the number of iterations of the preconditioned systems reduces significantly. Finally, we note that although the superlinear convergence rate is proved under the assumption that $M = O(n^{3+\nu})$, the numerical results show that the method indeed converges very fast even when the number of data samples M is just of the order $O(n)$.

	Variance of noise=0.0025								Variance of noise=0.025							
n	16		32		64		128		16		32		64		128	
m	I_n	C_n	I_n	C_n	I_n	C_n	I_n	C_n	I_n	C_n	I_n	C_n	I_n	C_n	I_n	C_n
2	14	10	26	12	38	15	48	16	17	9	25	13	42	15	65	15
4	18	10	22	11	32	12	48	13	16	10	25	12	41	11	53	13
8	15	9	18	10	32	10	37	11	13	9	24	10	34	11	42	12
16	13	8	19	9	28	9	37	10	14	8	20	9	28	9	39	10
32	12	8	17	8	26	8	36	9	13	8	20	8	27	9	39	9
64	13	7	17	8	24	8	32	8	12	7	19	7	26	8	37	8

Table 3. Average number of iterations for AR(1) input process when correlation windowing method is employed.

	Variance of noise=0.0025								Variance of noise=0.025							
n	16		32		64		128		16		32		64		128	
m	I_n	C_n	I_n	C_n	I_n	C_n	I_n	C_n	I_n	C_n	I_n	C_n	I_n	C_n	I_n	C_n
2	20	21	25	22	38	52	57	74	19	19	32	19	50	21	57	46
4	18	12	23	16	35	14	49	18	16	15	28	17	34	16	53	19
8	18	11	22	10	31	13	38	13	15	11	23	10	32	11	40	14
16	14	8	20	9	26	10	37	10	14	10	21	9	29	9	40	11
32	14	8	20	9	26	9	36	9	13	8	20	9	29	9	37	10
64	12	7	19	8	23	8	33	9	13	8	19	8	29	9	37	9

Table 4. Average number of iterations for AR(1) process when covariance windowing method is employed.

	Variance of noise=0.0025								Variance of noise=0.025							
n	16		32		64		128		16		32		64		128	
m	I_n	C_n	I_n	C_n	I_n	C_n	I_n	C_n	I_n	C_n	I_n	C_n	I_n	C_n	I_n	C_n
2	19	11	28	12	54	17	95	16	17	9	24	11	51	22	110	21
4	18	9	26	14	41	12	76	15	17	12	31	15	46	14	91	16
8	15	8	26	13	46	14	72	14	15	10	26	12	46	13	90	15
16	18	9	24	13	41	13	60	12	15	8	24	10	38	10	81	12
32	17	9	24	12	41	12	55	10	16	10	23	12	38	11	63	12
64	15	10	23	10	37	10	47	10	16	9	23	12	40	12	60	10

Table 5. Average number of iterations for AR(2) process when correlation windowing method is employed.

	Variance of noise=0.0025								Variance of noise=0.025							
n	16		32		64		128		16		32		64		128	
m	I_n	C_n	I_n	C_n	I_n	C_n	I_n	C_n	I_n	C_n	I_n	C_n	I_n	C_n	I_n	C_n
2	20	19	35	21	56	21	69	75	21	15	34	40	51	112	110	192
4	18	18	31	17	41	18	78	19	19	13	33	16	45	16	111	22
8	16	13	30	15	37	14	72	14	18	12	26	13	45	14	91	15
16	14	9	24	13	39	12	61	13	16	9	25	12	43	13	71	13
32	14	10	25	13	37	12	52	12	17	10	26	12	44	12	68	11
64	14	11	22	12	36	12	53	11	16	10	26	12	37	12	59	11

Table 6. Average number of iterations for AR(2) process when covariance windowing method is employed.

5 Concluding Remarks

Recently, Plemmons [25] proposed to use circulant preconditioner for the recursive (adaptive) least squares problems. We note that our algorithm is also suitable for such problems. For a real-time application of identification and recursive least squares computations, our algorithm can be executed on a parallel machine with multiprocessors. We assign each step of the algorithm to different group of processors. The first group of processors is responsible for the initialization of the data samples (i.e. to generate the right hand side vector of the normal equations and the first column of the Toeplitz matrices $A_n[g]$, $A_n[p]$, $A_n[q]$ and $C_n[g]$). The conjugate gradient iterations can be implemented on the second group of processors.

References

[1] S. Alexander, *Adaptive Signal Processing, Theory and Applications*, Springer-Verlag, New York, 1986.

[2] J. Biedmond, R. Lagendijk and R. Mesereau, *Iterative Methods for Image Deblurring*, Proc. of the IEEE, V78 (1990), pp. 856–883.

[3] R. Brent, F. Gustavson and D. Yun, *Fast Solution of Toeplitz Systems of Equations and Computations of Pade Approximations*, J. Algorithms, V1 (1980), pp. 259–295.

[4] P. J. Brockwell and R. A. Davis, *Time Series: Theory and Methods*, Springer-Verlag, New York, 1987.

[5] R. Chan and G. Strang, *Toeplitz Equations by Conjugate Gradients with Circulant Preconditioner*, SIAM J. Sci. Statist. Comput., V10 (1989), pp. 104–119.

[6] R. Chan and M. Yeung, *Circulant Preconditioners for Toeplitz Matrices with Positive Continuous Generating Functions*, Math. Comp., V58, (1992), pp. 233-240.

[7] R. Chan and M. Yeung, *Circulant Preconditioners Constructed From Kernels*, SIAM J. Numer. Anal., V29 (1992) pp. 1093–1103.

[8] R. Chan, J. Nagy and R. Plemmons, *Circulant Preconditioned Toeplitz Least Squares Iterations*, to appear SIAM J. on Matrix Anal. and Applic., (1992).

[9] T. Chan, *An Optimal Circulant Preconditioner for Toeplitz Systems*, SIAM J. Sci. Statist. Comput., V9 (1988), pp. 766–771.

[10] G. Cybenko, *Fast Toeplitz Orthogonalization Using Inner Products*, SIAM J. Sci. Statist. Comput., V8 (1987), pp. 734–740.

[11] P. Davis, *Circulant Matrices*, John Wiley & Sons, Inc., New York, 1979.

[12] W. Fuller, *Introduction to Statistical Time Series*, John Wiley & Sons, Inc., New York, 1976.

[13] A. Giordano and F. Hsu, *Least Square Estimation with Applications to Digital Signal Processing*, John Wiley & Sons, Inc., New York, 1985.

[14] U. Grenander and M. Rosenblatt, *Statisitcal Analysis of Stationary Time Series*, John Wiley & Sons, Inc., New York, 1957.

[15] U. Grenander and G. Szegö, *Toeplitz Forms and their Applications*, 2nd Ed., Chelsea Pub. Co., New York, 1984.

[16] T. Huckle, *Circulant and Skew-circulant Matrices for Solving Toeplitz Matrix Problems*, SIAM J. Matrix Anal. Appl., V13 (1992), pp. 767–777.

[17] F. Itakura and S. Saito, *Digital Filtering Techniques for Speech Analysis and Sythesis*, Proc. 7th International Congress on Acoustics, budapest, 1971, pp. 261–264.

[18] M. Kac, W. Murdock, and G. Szegö, *On the Eigenvalues of Certain Hermitian Forms*, Arch. Rational Mech. Anal., V2 (1953), pp.767–800.

[19] T. Ku and C. Kuo, *Design and Analysis of Toeplitz Preconditioners*, IEEE Trans. Acoust. Speech Signal Process., V40 (1991), pp. 129–141.

[20] N. Levinson, *The Wiener rms (root-mean-square) Error Criterion in Filter Design and Prediction*, J. Math. Phys., V25 (1947), pp. 261–278.

[21] D. Lee, M. Morf and B. Friedlander, *Recursive Least Squares Ladder Estimation Algorithms*, IEEE Trans. Acoust. Speech Signal Process., V29 (1981), pp. 627–641.

[22] S. Marple, *Efficient Least Squares FIR System Identification*, IEEE TRans. Acoustics, Speech Signal Process., V29 (1981), pp. 62–73.

[23] J. Nagy and R. Plemmons, *Some Fast Toeplitz Least Squares Algorithms*, Proc. SPIE Conference on Advanced Signal Processing Algorithms, Architectures, and Implementations II, V1566, San Diego, CA, July 1991.

[24] J. Pearl, *On Coding and Filtering Stationary Signals by Discrete Fourier Transform*, IEEE Trans. Inform. Theory, V19 (1973), pp. 229–232.

[25] R. Plemmons, *A Proposal for FFT-Based Fast Recursive Least Squares*, Proceedings to the IMA Workshop on Linear Algebra for Signal Processing, April 6-10, 1992.

[26] M. Priestley, *Spectral Analysis and Time Series*, Academic Press, New York, 1981.

[27] S. Qiao, *Recursive Least Squares Algorithm For Linear Prediction Problems*, SIAM J. Matrix Anal. Appl., V9 (1988), pp. 323–328.

[28] L. Rabiner, R. Chrochiere and J. Allen, *FIR System Modelling and Identification in the Presence of Noise and with Band-Limited Inputs*, IEEE Trans. Acoust. Speech Signal Process., V26 (1978), pp. 319–333.

[29] P. Sherman, *Circulant Approximations of the Inverses of Toeplitz Matrices and Related Quantities with Applications to Stationary Random Processes*, IEEE Trans. Acoustics, Speech Signal Process., V33 (1985), pp. 1630–1632.

[30] U. Steimel, *Fast Computation of Toeplitz Forms Under Narrow Band Conditions with Applications to Spectral Estimation*, Digital Signal Processing, Academic Press, New York, 1980.

[31] G. Strang, *A Proposal for Toeplitz Matrix Calculations*, Stud. Appl. Math., V74 (1986), pp. 171–176.

[32] M. Tismenetsky, *A Decomposition of Toeplitz Matrices and Optimal Circulant Preconditioning*, Lin. Alg. Appl., V154 (1991), pp. 105–121.

[33] E. Trytyshnikov, *Optimal and Super-optimal Circulant Preconditioners* , SIAM J. Matrix Anal. Appl., V13 (1992), pp. 459–473.

[34] J. Walker, *Fourier Analysis* , Oxford University Press, New York, 1988.

[35] J. Wilkinson, *The Algebraic Eigenvalue Problem*, Clarendon Press, Oxford, 1965.

[36] U. Yule, *On a Method for Investigating Periodicities in Disturbed Series with Special Reference to Wolfer's Sunspot Numbers*, Philos, Trans. Roy. Soc. London, Ser. A, V226 (1927), pp. 267–298.