RECURSIVE SOLUTIONS OF ESTIMATING EQUATIONS
AND ADAPTIVE SPECTRAL FACTORIZATION

BY

TZE LEUNG LAI AND ZHILIANG YING

TECHNICAL REPORT NO. 14
June 1990

PREPARED UNDER THE AUSPICES
OF
NATIONAL SCIENCE FOUNDATION GRANT DMS87-15614

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
Recursive Solutions of Estimating Equations
and Adaptive Spectral Factorization

By

Tze Leung Lai and Zhiliang Ying

Technical Report No. 14

June 1990

Prepared Under the Auspices

of

National Science Foundation Grant DMS87-15614

Department of Statistics

Stanford University

Stanford, California
Recursive Solutions of Estimating Equations
and Adaptive Spectral Factorization

TZE LEUNG LAI AND ZHILIANG YING

Abstract. In [1] a recursive spectral factorization algorithm is developed to be used with recursive instrumental variables for consistent estimation of the parameters of an ARMAX system, but the convergence of the spectral factorization algorithm is questionable. Herein a modified adaptive spectral factorization algorithm is shown to converge and a general method for constructing convergent recursive solutions of nonlinear equations that are used to define parameter estimates is also presented.

Keywords. Recursive estimation, spectral factorization.

T. L. Lai, Department of Statistics, Stanford University, Stanford, CA 94305 (Address for correspondence). Tel. #(415)494-7986.

Z. Ying, Department of Statistics, University of Illinois, Champaign, IL 61820. Tel. #(217)333-4792.

This work was supported by the National Science Foundation, the National Security Agency and the Air Force Office of Scientific Research.
I. INTRODUCTION

Consider a moving average $MA(k)$ process $y_t = C(q^{-1}) \varepsilon_t$, where $C(q^{-1}) = c_0 + c_1 q^{-1} + \ldots + c_k q^{-k}$ (with $c_0 > 0$) is a stable polynomial in the unit delay operator $q^{-1}$, and the $\varepsilon_t$ are uncorrelated random variables with mean 0 and variance 1. Let $\gamma_i = \gamma_{-i}$ be the autocovariances of $\{y_t\}$, $i = 0, 1, \ldots, k$, and let $\Gamma(z) = \sum_{j=-k}^{k} \gamma_j z^j$. Note that

$$\gamma_i = \sum_{j=0}^{k-i} c_j c_{j+i} \ (i = 0, \ldots, k),$$

or equivalently, $\Gamma(z) = C(z)C(z^{-1})$ for $z \neq 0$. Given $\gamma_0, \ldots, \gamma_k$, the spectral factorization problem is to solve (1) for $c_0(>0), \ldots, c_k$, or equivalently to factorize $\Gamma(z)$ as $C(z)C(z^{-1})$ with $C(z)$ stable (i.e., having all zeros outside the unit circle). The nonlinear equation (1) can be solved by standard Newton-Raphson iterations. Solo [1] proposed a modification of the Newton-Raphson iterative scheme, proved by Wilson [2] to be quadratically convergent, for on-line implementation in applications where the autocovariances $\gamma_i$ are not known in advance and have to be estimated recursively from the data. Noting that the sample autocovariances are strongly consistent estimates of $\gamma_i$ under certain assumptions, Solo [1] extended Wilson’s argument in [2] to prove that his adaptive spectral factorization algorithm is convergent, but the convergence demonstrated is questionable as explained below.

Let $\Gamma_t(z) = \sum_{j=-k}^{k} \gamma_{t,j} z^j$, where the $\gamma_{t,j}(= \gamma_{t,-j})$ are estimates of $\gamma_j$ at stage $t$. Letting $c_{t,j}$ be estimates of $c_j$ at stage $t$, define $C_t(z) = \sum_{j=0}^{k} c_{t,j} z^j$. Solo [1, (4b)] modified Wilson’s argument to show that for the $c_{t,j}$ generated by his algorithm, $\text{Re}\{C_{t+1}(z)/C_t(z)\}$ is a harmonic function for $|z| \leq 1$ and satisfies the inequality

$$1 \leq 2 \text{Re}\{C_{t+1}(z)/C_t(z)\} \leq 1 + \Gamma_{t+1}(z)/\Gamma_t(z)$$

(2)

for $|z| = 1$. In the case $\gamma_{t,j} = \gamma_j$ so that $\Gamma_t(z) = \Gamma(z)$, Wilson [2] used the maximum principle (that $\text{Re}\{C_{t+1}(z)/C_t(z)\}$ is either constant for $|z| \leq 1$ or attains its maximum and minimum in this region at points on the boundary $|z| = 1$) to conclude from (2) that

$$1 \leq 2 \text{Re}\{C_{t+1}(z)/C_t(z)\} \leq 2 \text{ for } |z| \leq 1,$$

since $\Gamma_{t+1}(z)/\Gamma_t(z) = 1$ in this case. Solo [1] likewise applied the maximum principle to conclude that (2) also holds for $|z| \leq 1$, and
then applied (2) to prove convergence of his algorithm. However, the upper bound in (2) is typically a non-constant function of $z$ when $\gamma_j$ is replaced by a time-varying estimate $\gamma_{t,j}$, and the denominator $\Gamma_t(z)$ in fact has zeros inside $|z| < 1$. Therefore (2) cannot be extended from $|z| = 1$ to $|z| \leq 1$ by applying the maximum principle. What one can conclude from the inequality (2) for $|z| = 1$ and the maximum principle is that

$$1 \leq 2 \text{Re}\{C_{t+1}(z)/C_t(z)\} \leq 2 + \max_{|z|=1} |\Gamma_{t+1}(z)/\Gamma_t(z) - 1| \quad \text{for} \quad |z| \leq 1. \quad (3)$$

In Section II we shall make use of (3) to show that a modified version of Solo’s algorithm is indeed convergent if $\gamma_{t,j}$ converges to $\gamma_j$ with probability 1 for $j = 0, \ldots, k$. Convergence rates for the algorithm are also provided in terms of those for the $\gamma_{t,j}$. This convergent adaptive spectral factorization algorithm can be regarded as an example of a general method for constructing convergent recursive solutions to nonlinear estimating equations presented in Section III. Finally, applications of the convergent spectral factorization algorithm to adaptive control and recursive identification of ARMAX systems are discussed in Section IV.

II. ADAPTIVE SPECTRAL FACTORIZATION

Since the $\gamma_i$ and $c_j$ in (1) are usually unknown, an obvious modification is to replace at each stage $t$ the unknown $\gamma_i$ in (1) by an estimate $\gamma_{t,i}$ and then use the resultant equation to determine an estimate of $c_j$. Suppose that the estimates $\gamma_{t,i}$ are strongly consistent, so that with probability 1

$$\gamma_{t,i} - \gamma_i = o(t^{-b}), \quad i = 0, \ldots, k, \quad (4)$$

for some $0 < b < 1/2$. In particular, if the innovations $\epsilon_n$ form a martingale difference sequence with respect to an increasing sequence of $\sigma$-fields $\mathcal{F}_n$ such that $E(\epsilon_n^2|\mathcal{F}_{n-1}) = 1$ and $\sup_n E(\epsilon_n^4|\mathcal{F}_{n-1}) < \infty$, then (4) is satisfied for every $b < 1/2$ by the sample autocovariances $\gamma_{t,i} = (t-i)^{-1} \sum_{s=1}^{t-i} y_s y_{s+i}$, in view of martingale strong laws, cf [3].

Let $t(0) = 0$ and choose positive integers $t(1) < t(2) < \cdots$ such that

$$\lim_{n \to \infty} t(n)/n^\alpha = \infty \quad \text{for some} \quad \alpha > 1/b. \quad (5)$$

For example, take $t(n) = 2^n$. We first use Solo’s approach [1] to recursively construct approximate solutions of the equations $\gamma_{t(n),i} = \sum_{j=0}^{k-i} \theta_j^{(n)} \theta_{j+i}^{(n)}, i = 0, \ldots, k, n = 1, 2, \ldots,$
which can be regarded as adaptive versions of (1) at the stages \( t(n) \). The approximate solution \( \theta^{(n)} = (\theta_0^{(n)}, \ldots, \theta_k^{(n)})^T \) at stage \( t(n) \) is essentially a one-step Newton-Raphson iteration initialized at \( \theta^{(n-1)} \). Specifically, let \( \Delta_{t(n-1)} \) be a \((k+1) \times (k+1)\) matrix whose \((i,j)\)th element is \( \hat{\theta}_i^{(n-1)} - \hat{\theta}_j^{(n-1)} \), setting \( \hat{\theta}_\nu^{(n-1)} = 0 \) if \( \nu > k \) or \( \nu < 0 \). Let \( \phi_{n,i} = \sum_{j=0}^{k-i} \theta_j^{(n-1)} \theta_{j+i}^{(n-1)} - \gamma_{t(n),i}, i = 0, \ldots, k \). Define \( \theta_0^{(0)} > 0 = \theta_1^{(0)} = \ldots = \theta_k^{(0)} \) and

\[
\theta^{(n)} = \theta^{(n-1)} - \Delta_{t(n-1)}^{-1} (\phi_{n,0}, \ldots, \phi_{n,k})^T, \quad n \geq 1.
\] (6)

Perform a stability test (cf. [4, p. 153]) for the polynomial \( \sum_{j=0}^{k} \theta_j^{(n)} z^j \), redefining (6) by \( \theta^{(n)} = \theta^{(n-1)} \) if and only if the stability test fails. Take \( c > 0 \). Let \( \theta_0 = \theta^{(0)} \), and for \( t(n-1) < t \leq t(n) \) define \( \theta_t = (\theta_{t,0}, \ldots, \theta_{t,k})^T \) recursively by

\[
\tilde{\theta}_t = \theta_{t-1} - \Delta_{t(n-1)}^{-1} \left( \sum_{j=0}^{k-i} \theta_{t-1,j} \theta_{t-1,j+i} - \gamma_{t,i} \right)_{0 \leq i \leq k},
\] (7a)

\[
\theta_t = \tilde{\theta}_t \quad \text{if} \quad \| \tilde{\theta}_t - \theta^{(n-1)} \| \leq c/n, \quad \text{and} \quad \theta_t = \theta^{(n)} \quad \text{otherwise.} \] (7b)

Let \( \theta = (c_0, \ldots, c_k)^T \). The following theorem establishes the strong consistency of the adaptive spectral factorization scheme \( \theta_t \) defined recursively by (7).

**Theorem 1**: Suppose that \( c_0 > 0, \sum_{j=0}^{k} c_j z^j \neq 0 \) for all \( |z| \leq 1 \), and that the \( \gamma_{t,i} \) satisfy (4) for some \( 0 < b < 1/2 \). Then with probability 1, \( \theta_t - \theta = o(t^{-b}) \).

**Proof**: We first show that \( \theta^{(n)} \to \theta \) with probability 1 (abbreviated henceforth as “w.p.1”). Setting \( \gamma_{t,-j} = \gamma_{t,j} \), let \( \rho_n(z) = \sum_{j=-k}^{k} \gamma_{t(n),j} z^j \), \( \theta_n(z) = \sum_{j=0}^{k} \theta_j^{(n)} z^j \). Since \( \gamma_{t(n),i} \to \gamma_i \) w.p.1, \( \theta_n(z) \) passes the stability test for all large \( n \) w.p.1. Moreover, in view of the definitions of \( \Delta_{t(n-1)} \) and \( \phi_{n,i} \), (6) can be expressed in the form

\[
\theta_{n-1}(z) \theta_n(z)^{-1} + \theta_n(z) \theta_{n-1}(z)^{-1} = \theta_{n-1}(z) \theta_{n-1}(z)^{-1} + \rho_n(z),
\] (8)

cf. [1, p. 1049], [2, p. 3]. Hence the same argument as in [1] or [2] shows that \( \text{Re}\{\theta_n(z)/\theta_{n-1}(z)\} \) is a harmonic function for \( |z| \leq 1 \) and that

\[
1 \leq 2\text{Re}\{\theta_n(z)/\theta_{n-1}(z)\} \leq 1 + \rho_n(z)/\rho_{n-1}(z) \quad \text{on} \quad |z| = 1.
\] (9)

Therefore, analogous to (3), it follows from (9) and the maximum principle that

\[
1/2 \leq \text{Re}\{\theta_n(z)/\theta_{n-1}(z)\} \leq 1 + \beta_n \quad \text{for} \quad |z| \leq 1, \quad \text{where}
\]

\[
\beta_n = \max_{|z|=1} |\rho_n(z)/\rho_{n-1}(z) - 1|/2.
\] (10)
Clearly $\beta_n = O(\max_{j \leq k} |\gamma_{t(n), j} - \gamma_{t(n-1), j}|) = o((t(n - 1))^{-b})$ w.p.1 by (4), and therefore $\sum_1^\infty \beta_n < \infty$ w.p.1 by (5), implying that $\prod_{i=1}^n (1 + \beta_i)$ converges w.p.1. From (10) it follows that

$$2(1 + \beta_n)^{-1} \leq \Theta_n(x)/\Theta_{n-1}(x) \leq 1 \quad \text{for } x \in [-1, 1],$$

where

$$\Theta_n(x) = \theta_n(x)/\prod_{i=1}^n (1 + \beta_i).$$

Hence $\Theta_n(x)$, and therefore $\theta_n(x)$ also, converge uniformly for $x \in [-1, 1]$ w.p.1, implying that $\theta^{(n)} \rightarrow \theta$ w.p.1, cf. [2, p. 4].

We next show that $\|\theta^{(n)} - \theta\| = o(n^{-1})$ w.p.1. Let

$$f_t(\xi) = (f_{t, 0}(\xi), \ldots, f_{t, k}(\xi))^T, \quad f_{t, i}(\xi) = \sum_{j=0}^{k-i} \xi_j \xi_{j+i} - \gamma_{t, i},$$

for $\xi = (\xi_0, \ldots, \xi_k)^T$. By the definition of $\phi_{n, i}$ and (1),

$$\phi_{n, i} = [f_{t(n), i}(\theta^{(n-1)}) - f_{t(n), i}(\theta)] + f_{t(n), i}(\theta), \quad f_{t, i}(\theta) = \gamma_i - \gamma_{t, i}.$$  \hspace{1cm} (13)

Moreover, the Jacobian matrix $Df_t := (\partial f_{t, i}/\partial \xi_j)_{0 \leq i, j \leq k}$ does not depend on $t$, and at $\xi = \theta$ is equal to $\Delta$, which is nonsingular, cf. [2]. Since $\theta^{(n)} \rightarrow \theta$ w.p.1, $\Delta_{t(n-1)} \rightarrow \Delta$ and $\theta^{(n-1)} - \theta - \Delta_{t(n-1)}^{-1}[f_{t(n), i}(\theta^{(n-1)}) - f_{t(n), i}(\theta)] = o(\|\theta^{(n-1)} - \theta\|)$ w.p.1. Combining this with (6), (13), (4) and (5) gives

$$\|\theta^{(n)} - \theta\| = o(\|\theta^{(n-1)} - \theta\|) + o((t(n))^{-b}) = o(\|\theta^{(n-1)} - \theta\|) + o(n^{-1}) \text{ w.p.1.}$$ \hspace{1cm} (14)

By Lemma 2 below, (14) implies that $\|\theta^{(n)} - \theta\| = o(n^{-1})$ w.p.1.

Let $\theta'_t = \theta^{(n-1)}$ and $\delta_t = c/n$ for $t(n - 1) < t \leq t(n)$. In view of (5), for $t(n - 1) < t \leq t(n)$, $t^{-b} < (t(n - 1))^{-b} = o(n^{-1}) = o(\delta_t)$. Defining $f_t$ as in (12), we can write the recursion (7) in the form $\theta_t = \tilde{\theta}_t$ if $\|\tilde{\theta}_t - \theta'_t\| \leq \delta_t$ and $\theta_t = \theta'_t$ otherwise, where $\tilde{\theta}_t = \theta_{t-1} - (Df_t(\theta'_t))^{-1} f_t(\theta_{t-1})$. By (13) and (4), $f_t(\theta) = o(t^{-b}) = o(\delta_t)$ w.p.1. Since we have shown that $\|\theta'_t - \theta\| = o(\delta_t)$ w.p.1, the desired conclusion for $\theta_t$ is an immediate corollary of the more general Theorem 2 in Section III.

Lemma 1: Let $a_n, b_n$ be positive numbers such that $b_n \rightarrow 0, b_{n-1} + b_{n+1} + n^{-\delta} = O(b_n)$ as $n \rightarrow \infty$, for some $\delta > 0$. If $a_n = o(a_{n-1} + b_n)$, then $a_n = o(b_n)$.

Proof: Let $\epsilon > 0$. Choose $K$ and $n_0 \geq 1$ such that

$$b_n \leq K b_{n-1}, b_{n-1} \leq K b_n \quad \text{and} \quad a_n \leq \epsilon(K + \epsilon)^{-2}(a_{n-1} + b_n) \quad \text{for all} \quad n \geq n_0.$$ \hspace{1cm} (15)
There exists \( n_1 > n_0 \) such that \( a_{n_1} \leq \varepsilon b_{n_1} \), because otherwise \( a_n > \varepsilon b_n \) for all large \( n \), implying that \( a_n = o(a_{n-1}) \) and therefore \( a_n = o(n^{-\delta}) \), contradicting that \( n^{-\delta} = O(b_n) \). Making use of (15), an induction argument then shows that \( a_n \leq \varepsilon b_n \) for all \( n \geq n_1 \).

III. ESTIMATING EQUATIONS AND APPROXIMATE RECURSIVE SOLUTIONS

Defining \( f_t \) as in (12), the equation \( f_t(\xi) = 0 \), whose solution corresponds to the estimate of the moving average parameters of the MA\((k)\) process by the method of moments (that involves substituting in (1) the “moments” \( \gamma_i \) by their sample estimates \( \gamma_{t,i} \)), is an example of “estimating equations” commonly used to define statistical estimators, cf. [5]. Another well known example of estimating equations is the system

\[
\frac{\partial}{\partial \xi_j} \log \psi_t(y_1, \ldots, y_t; \xi) = 0, \quad j = 1, \ldots, p, \tag{16}
\]

used to determine the maximum likelihood estimate of a parameter vector \( \theta = (\theta_1, \ldots, \theta_p)^T \), where \( \psi_t \) denotes the joint density function of \( y_1, \ldots, y_t \).

Suppose that an unknown parameter \( \theta = (\theta_1, \ldots, \theta_p)^T \) is estimated at each stage \( t \) by solving an estimating equation of the form \( f_t(\xi) = 0 \), where \( f_t = (f_{t,1}, \ldots, f_{t,p})^T \) is a continuously differentiable vector-valued function defined either from all the data up to stage \( t \) (as in (16)) or from summary statistics thereof (such as the \( \gamma_{t,0}, \ldots, \gamma_{t,k} \) in (12)). Let \( Df_t = (\partial f_{t,i}/\partial \xi_j)_{1 \leq i,j \leq p} \) denote the Jacobian matrix, and let \( \|A\| = \sup_{\|x\|=1} \|Ax\| \) denote the norm of a \( p \times p \) matrix \( A \). Suppose that w.p.1,

\[
f_t(\theta) = o(t^{-b}) \quad \text{as} \quad t \to \infty, \tag{17}
\]

\[
\sup_{\|\xi-\theta\| \leq \epsilon} \|Df_t(\xi) - A\| \to 0 \quad \text{as} \quad t \to \infty \quad \text{and} \quad \epsilon \to 0, \tag{18}
\]

for some \( 0 < b < 1/2 \) and nonsingular matrix \( A \). Then a standard argument involving Taylor’s expansion shows that w.p.1, the equation \( f_t(\xi) = 0 \) has a solution \( \hat{\xi}_t \) for all large \( t \) such that \( \hat{\xi}_t - \xi = o(t^{-b}) \). If furthermore

\[
t^{1/2}f_t(\theta) \quad \text{has a limiting normal distribution with mean} \ 0 \ \text{and covariance} \ V, \tag{19}
\]

then \( t^{1/2}(\hat{\xi}_t - \xi) \) also has a limiting normal distribution with mean 0 and covariance matrix \( A^{-1}V(A^{-1})^T \), cf. [3], [5]. In particular, for the maximum likelihood example (16), let
\( f_t = t^{-1}(\partial \log \psi_t / \partial \xi_j)^T_{1 \leq j \leq p} \) and note that (17) and (18) correspond to standard regularity conditions with \( -A \) being the Fisher information matrix \( F \). Furthermore (19) holds with \( V = F \), and it is well known that under these regularity conditions, the maximum likelihood estimate is asymptotically normal with mean \( \xi \) and covariance \( t^{-1}F^{-1} \).

In order to find a solution \( \hat{\xi}_t \) of the estimating equation that is near \( \theta \), one typically first evaluates some preliminary estimate \( \theta'_t \) of \( \theta \) and then uses \( \theta'_t \) as a starting value in iterative solution of the equation \( f_t(\xi) = 0 \). Suppose that these preliminary estimates \( \theta'_t \) are strongly consistent, so that there exist constants \( \delta_t \to 0 \) such that

\[
\|\theta'_t - \theta\| = o(\delta_t) \quad \text{w.p.1,} \quad \delta_{t-1} + \delta_{t+1} + t^{-b} = O(\delta_t). \tag{20}
\]

Instead of carrying out the iterations until convergence at every stage \( t \), consider the following recursive algorithm analogous to (7). If \( Df_t(\theta'_t) \) is singular, let \( \theta_t = \theta'_t \). If \( Df_t(\theta'_t) \) is nonsingular, define

\[
\tilde{\theta}_t = \theta_{t-1} - (Df_t(\theta'_t))^{-1}f_t(\theta_{t-1}), \tag{21a}
\]

\[
\theta_t = \tilde{\theta}_t \quad \text{if} \quad \|\tilde{\theta}_t - \theta'_t\| \leq \delta_t \quad \text{and} \quad \theta_t = \theta'_t \quad \text{otherwise}. \tag{21b}
\]

The following theorem shows that the recursive estimator \( \theta_t \) has the same asymptotic properties as the solution \( \hat{\xi}_t \) of the estimating equation \( f_t(\xi) = 0 \).

**Theorem 2:** Let \( f_t : \mathbb{R}^p \to \mathbb{R}^p \) be continuously differentiable such that (17) and (18) hold for some \( 0 < b < 1/2 \) and nonsingular matrix \( A \). Let \( \theta'_t \to \theta \) w.p.1 and \( \delta_t \to 0 \) such that (20) holds. Define \( \theta_t \) recursively by (21). Then \( \theta_t - \theta = o(t^{-b}) \) w.p.1. If furthermore (19) holds and

\[
\sup_t t^{1/2} \max_{t \geq s \geq t - \log t} |f_s(\theta)| \geq x \to 0 \quad \text{as} \quad x \to \infty, \tag{22}
\]

then \( t^{1/2}(\theta_t - \theta) \) has a limiting normal distribution with mean 0 and covariance matrix \( A^{-1}V(A^{-1})^T \).

**Proof:** By (20), \( \|\theta'_t - \theta\| = o(\delta_t) \) w.p.1, and therefore (21b) implies that \( \|\theta_t - \theta\| = O(\delta_t) \) w.p.1. Moreover,

\[
f_t(\theta_{t-1}) = f_t(\theta) + (Df_t(\theta^*_{t-1}))(\theta_{t-1} - \theta), \quad \text{with} \quad \|\theta^*_{t-1} - \theta\| \leq \|\theta_{t-1} - \theta\|. \tag{23}
\]
Applying (18) and (23) to (21a) then shows that

$$\tilde{\theta}_t - \theta = -(Df_t(\theta'_t))^{-1} f_t(\theta) + o(||\theta_{t-1} - \theta||) \text{ w.p.1},$$

(24)

noting that $A$ is nonsingular and that $Df_t(\theta'_t)$ is nonsingular for all large $t$ w.p.1 by (18). From (24) and (17), it follows that $||\tilde{\theta}_t - \theta|| = o(t^{-b}) + o(||\theta_{t-1} - \theta||) = o(\delta_{t-1}) = o(\delta_t)$ w.p.1, in view of (20). Since $||\theta'_t - \theta|| = o(\delta_t)$ w.p.1, it then follows that w.p.1, $||\tilde{\theta}_t - \theta'_t|| = o(\delta_t)$ and therefore $\tilde{\theta}_t = \theta_t$ for all large $t$. Hence by (21a) and (23), w.p.1, for all large $t$,

$$\theta_t - \theta = -(Df_t(\theta'_t))^{-1} f_t(\theta) + R_t(\theta_{t-1} - \theta), \quad \text{where } ||R_t|| \rightarrow 0. \quad (25)$$

From Lemma 1 together with (25), (17) and (18), $\theta_t - \theta = o(t^{-b})$ w.p.1.

Let $x_t = \theta_t - \theta$, $z_t = -(Df_t(\theta'_t))^{-1} f_t(\theta)$, so (25) is of the form $x_t = R_t x_{t-1} + z_t$. Hence w.p.1, for all large $m$,

$$x_t = R_t \cdots R_{t-m} x_{t-m-1} + z_t + \sum_{j=1}^{m} R_t \cdots R_{t-j+1} z_{t-j} \quad \text{for } t > m. \quad (26)$$

Let $m$ be the largest integer $\leq \log t$. Since $||R_t|| \rightarrow 0$ and $x_{t-m-1} = o(t^{-b})$ w.p.1, $t^{1/2} R_t \cdots R_{t-m} x_{t-m-1} \rightarrow 0$ w.p.1. Moreover, $\sum_{j=1}^{m} R_t \cdots R_{t-j+1} t^{1/2} z_{t-j}$ converges to 0 in probability by (22), since $||R_t|| \rightarrow 0$ and $Df_t(\theta'_t) \rightarrow A$ w.p.1. Since $t^{1/2} z_t$ has a limiting normal distribution with mean 0 and covariance matrix $A^{-1} V (A^{-1})^T$, it then follows that $t^{1/2} x_t$ has the same limiting distribution.

The assumption (22) says that $t^{1/2} \max_{t \geq s \geq t - \log t} |f_s(\theta)|$ is bounded in probability. For the spectral factorization example in which $f_t$ is defined by (12), $f_t(\theta) = (\gamma_{t-i})_{0 \leq i \leq k}$.

When the $\gamma_{t,i}$ are sample autocovariances (i.e., $\gamma_{t,i} = (t-i)^{-1} \sum_{s=1}^{t-i} y_s y_{s+i}$) and $y_n = \sum_{j=0}^{k} c_j \epsilon_{n-j}$ is a moving average of martingale differences $\epsilon_n$ such that $E(\epsilon_n^2 | \mathcal{F}_{n-1}) = 1$ and $\sup_n E(\epsilon_n^4 | \mathcal{F}_{n-1}) < \infty$, conditions (19) and (22) are satisfied in view of martingale central limit theorems, cf. [3].

IV. CONCLUDING REMARKS

The recursive scheme (21) can be made very flexible by redefining $f_t$ if necessary. For example, it can incorporate $k$ iterative solution steps for an estimating equation $f_s(\xi) = 0$.
by simply setting \( f_t = f_s \) for \( t = t_s, t_s + 1, \ldots, t_s + k - 1 \). This may be particularly useful during initial stages of the recursion when the \( f_s \) are based on relatively few data. In particular, we can incorporate such iterations into the auxiliary estimator \( \theta^{(n)} \) in our spectral factorization algorithm.

While the adaptive spectral factorization algorithm in Section II is only discussed in the context of moving average processes, we can also apply it to general ARMAX systems of the form \( A(q^{-1})y_t = B(q^{-1})u_{t-d} + C(q^{-1})\epsilon_t \) (where \( A(q^{-1}) = 1 + a_1 q^{-1} + \cdots + a_p q^{-p} \) and \( B(q^{-1}) = b_0 + \cdots + b_h q^{-h} \)) by a suitable choice of \( \gamma_{t,i} \). Solo [1] suggested first using recursive instrumental variables to estimate \( A(q^{-1}) \) and \( B(q^{-1}) \) at stage \( t \) by \( \hat{A}_t(q^{-1}) \) and \( \hat{B}_t(q^{-1}) \), and then defining the \( \gamma_{t,i} \) as the sample autocovariances of \( w_t := \hat{A}_t(q^{-1})y_t - \hat{B}_t(q^{-1})u_{t-d} \). We have recently used this idea, denoted by \( IV - W \) in [1], to provide a consistent estimate of the parameters of the ARMAX system in [6], where we give a relatively complete solution to the adaptive control problem of choosing the inputs \( u_t \) sequentially so that the outputs \( y_t \) are as close as possible to certain target values \( y_t^* \), without assuming prior knowledge of the system parameters. The solution involves parallel implementation of the \( IV - W \) algorithm and the stochastic gradient and recursive maximum likelihood algorithms. It also involves occasional use of a dither signal probing the system for information, to ensure consistency of the \( IV - W \) algorithm via an argument similar to the proof of Theorem 1.

REFERENCES


