Stochastics, 1986, Vol. 16, pp. 137–155
0090-9491/86/1602-0137 $18.50/0

Detecting Changes in the AR Parameters of a Nonstationary ARMA Process

GEORGE V. MOUSTAKIDES and ALBERT BENVENISTE

Institut de Recherche en Informatique et Systemes Aleatoires, Avenue du General Leclerc, Rennes 35042, France

(*Accepted for publication July 15, 1985*)

We present a method for detecting changes in the AR parameters of an ARMA process with arbitrarily time varying MA parameters. Assuming that a collection of observations and a set of nominal time invariant AR parameters are given, we test if the observations are generated by the nominal AR parameters or by a different set of time invariant AR parameters. The detection method is derived by using a local asymptotic approach and it is based on an estimation procedure which was shown to be consistent under nonstationarities.

KEY WORDS: Detection of changes, nonstationary ARMA processes.

I. INTRODUCTION

The problem of detecting changes in the spectral parameters of processes is frequently encountered in practice. Most methods (likelihood or innovation based techniques) solve this problem by using a complete description of the spectral parameters, (see for example [2, 8, 11]). Sometimes this description is not possible for all parameters, this is for instance the case when nonstationarities are present. In such situations one is mainly interested in those parameters that can be described, while regarding the others as nuisance. A problem of this type is the problem of vibration monitoring. Here one is interested in detecting changes in the vibrating modes of a system, without taking into account the cause that generates these vibrations. If we use an ARMA model to describe the system then the problem translates into detecting changes in the AR parameters (assumed time invariant) while the MA part is assumed nonstationary and unknown. Clearly the classical techniques do not apply directly here because of the high variability of the MA part. In this paper we give a solution to this problem. The method that we present is based on an estimation method for the AR parameters that is known to be consistent even under nonstationarities. To derive the detection scheme we use asymptotic techniques assuming that the number of observations goes to infinity and the magnitude of the change goes to zero. Similar asymptotic techniques are used in [5, 9], but they are mainly applied to the likelihood function, something which is not possible here.

II. PROBLEM STATEMENT

Let us consider the following system of difference equations

$$\begin{aligned} X_{k+1} &= FX_k + W_{k+1} \quad X_0 = 0 \\ y_k &= c^T X_k \end{aligned} \tag{1}$$

where F is a real square matrix of dimension m, c is a real vector of dimension m and $\{W_k\}$ is a sequence of zero mean independent nonstationary vectors. With the superscript "T" we denote the transpose. The model in (1) is often used to model real systems. The vector X_k is called the state of the system, W_k the input and the scalar y_k the output (observations). The process y_k is an ARMA process and if we write it in this form then we have

$$y_k - \alpha_1 y_{k-1} - \cdots - \alpha_m y_{k-m} = \beta_k^0 e_k + \cdots + \beta_k^{m-1} e_{k-m+1} \tag{2}$$

where $\{e_k\}$ is a standard i.i.d. sequence. The vector $\alpha^T = [\alpha_1 \ldots \alpha_m]$ is the vector of the AR parameters and the vector $\beta_k^T = [\beta_k^0 \ldots \beta_k^{m-1}]$ is the vector of the MA parameters. The vector α is related only to the matrix F, its components are the coefficients of the characteristic polynomial of this matrix. Since the roots of this polynomial are the spectral modes of the system, all the spectral information is contained in the vector α. The vector β_k is related to the input W_k thus it is time varying. Notice that with the model in (1) we can have a MA part in (2) of order at most $m-1$.

Since any change in the spectral modes reflects into a change in the vector α we concentrate on this vector. Thus we are interested in detecting changes in the vector α in a *nonsequential way*, assuming that we have available the observation sequence $\{y_k\}$. We will assume that a nominal vector $\alpha=\alpha_0$ is known and that we do not know the vector α after the change. More specifically we will assume that a collection $\{y_1,\ldots,y_n\}$ is given and we would like to decide between the two hypotheses, $H_0:\alpha=\alpha_0$ and $H_1:\alpha\neq\alpha_0$. As we said in the introduction we will follow an asymptotic approach, thus we suppose that under change, α is of the form $\alpha=\alpha_0+\theta/\sqrt{n}$, where θ is an unknown direction of change. Although for our test we will never use the matrix F we will assume that α_0 corresponds to some nominal F_0 and $\alpha_0+\theta/\sqrt{n}$ to $F_0+\Theta/\sqrt{n}$. Let us now consider the following vector:

$$U_n=\frac{1}{\sqrt{n}}\sum_{k=1}^{n}(y_{k+1}-Y_k^T\alpha_0)Z_k \tag{3}$$

where $Y_k^T=[y_k\ldots y_{k-m+1}]$ and $Z_k^T=[y_{k-m+1}\cdots y_{k-m-M+2}]$. Equation (3) can be used to estimate the AR parameters of the observation process. Since U_n (disregarding $1/\sqrt{n}$) is the sample correlation of the MA part with delayed enough observation and since this correlation is zero, we can estimate α by solving the system $U_n=0$ (in the mean square sense if $M>m$). This estimate is known to be consistent even under nonstationarities [4]. We use a similar idea to derive a test. We expect that under no change the vector U_n will have a zero mean and under change a nonzero mean. Indeed we have

$$\begin{aligned} &\Sigma_n^{-1/2}U_n \xrightarrow{d} N(0,I) && \text{under } H_0 \\ &\Sigma_n^{-1/2}(U_n-H_n\theta) \xrightarrow{d} N(0,I) && \text{under } H_1 \end{aligned} \tag{4}$$

where $\xrightarrow{d}$ means convergence in distribution, $N(0,I)$ is the standard vector Gaussian density and Σ_n is the covariance matrix of U_n. The matrix H_n is defined as

$$H_n = \frac{1}{n} \sum_{k=1}^{n} Z_k Y_k^T. \tag{5}$$

We can see that under both hypotheses U_n has asymptotically the same variance Σ_n but not the same mean. Next step is to use U_n as if we had equality in (4) and define a log-likelihood

$$T_n' = U_n^T \Sigma_n^{-1} H_n \theta - \tfrac{1}{2} \theta^T H_n^T \Sigma_n^{-1} H_n \theta. \tag{6}$$

Notice that Z_k is of length M thus H_n is of dimension $M \times m$, assuming $M \geqq m$ let us substitute θ by its maximum likelihood estimate, this yields

$$T_n = U_n^T \Sigma_n^{-1} H_n (H_n^T \Sigma_n^{-1} H_n)^{-1} H_n^T \Sigma_n^{-1} U_n. \tag{7}$$

The quantity T_n will be our test statistic. To define the threshold we have that under H_0, T_n is asymptotically χ^2 with m degrees of freedom. Under H_1, T_n is a noncentered χ^2 with noncentricity equal to $\theta H_n^T \Sigma_n^{-1} H_n \theta$. Our test will be able to detect the change if the noncentricity is nonzero. As we will show in Section IV, with our assumptions this is always the case. The covariance matrix Σ_n is not known and thus we must estimate it. We can for example use the sample covariance matrix defined as

$$\hat{\Sigma}_n = \frac{1}{n} \sum_{k=1}^{n} v_k^2 Z_k Z_k^T + \frac{1}{n} \sum_{j=1}^{m-1} \sum_{k=1}^{n} v_k v_{k-j} (Z_k Z_{k-j}^T + Z_{k-j} Z_k^T), \tag{8}$$

where $v_k = y_k - Y_{k-1}^T \alpha_0$. Thus everything is defined in terms of known things. In the next sections we will prove the validity of (4), i.e. that the Central Limit Theorem (CLT) holds for U_n under both hypotheses. Notice also that in (7) we take the inverse of the matrices Σ_n and $H_n^T \Sigma_n^{-1} H_n$. We will show that these two matrices are full rank. For Σ_n this will be necessary for the proof of the CLT, for the other matrix it is important in showing that the noncentricity factor is nonzero. Notice that the second matrix will be full rank if H_n is full rank and Σ_n is nonsingular and bounded from above. The proofs will be presented in several steps.

III. ASSUMPTIONS

Let us now introduce our assumptions. Let $\{y_k\}$ be an ARMA process generated by a system of the form of (1) or (2). Let F_0, α_0 be the nominal values of F and α. We assume:

A1. The matrix F_0 is full rank and has all its eigenvalues distinct and strictly inside the unit circle.

A2. The matrix $0=[cF_0^Tc\ldots(F_0^T)^{m-1}c]$ is of full rank.

A3. There exists a real $p>0$ such that for every vector λ and every integer k we have $E\{[\lambda^T W_k]^4\}\leqq p(\lambda^T\lambda)^2$, where by $E\{\ \}$ we denote expectation.

A4. If $Q_k=E\{W_kW_k^T\}$ is the input covariance matrix at time k and if $\bar{Q}_n=(1/n)\sum_{k=1}^n Q_k$ is the average input covariance up to time n, we assume that there exists a real $\delta>0$ such that for every eigenvalue r_i, $i=1,\ldots,m$ of F_0 we have

$$\left|\lim_{z\to r_i}\{(z-r_i)c^T(zI-F_0)^{-1}\bar{Q}_n(z^{-1}-F_0)^{-1}c\}\right|\geqq\delta.$$

A5. There exists a nonzero vector s such that for every k we have $Q_k\geqq ss^T$.

Assumption A1 is to ensure the stability of our system, i.e. bounded inputs will produce bounded outputs. A2 is an observability condition, it ensures that any mode excited by the input will be observed in the output. Assumption A3 is rather technical: it requires that the input has uniformly bounded fourth order moments. A4 means that in the average the input excites all the modes of the system and that these modes are present in the second order statistics. Finally Assumption A5 is to ensure that the input excites constantly at least at the direction of the vector s.

Discussion We will comment now on our assumptions. In A1 the assumption that the eigenvalues are distinct was made only to facilitate certain proofs; it can be relaxed to include multiple eigenvalues. Notice also that we assume knowledge of the exact system order m. A2 seems necessary. Assumption A3 is quite strong, probably it can be relaxed to conditions involving only second order moments. Assumption A4, as we said before, ensures that all the modes are present in the second order statistics. This assumption is very important since our test is based on second order statistics. If for example A4 was not true for some mode, in other words, if a mode even though excited by the input was not present in the second order statistics (this is the case when the MA part of a stationary ARMA process has zeros that are mirror images, with respect to the unit circle, of poles of the AR part), then any change on this mode cannot be detected by simply using second order statistics. Notice also that A4 involves only the average input covariance, that is, instantaneously we can have cancellation of modes. As one can see this assumption will be the base for proving that H_n is of full rank. In the stationary case usually the assumptions up to A4 are sufficient to show the same things we like to show here. When nonstationarities are present this is no longer true, one can find examples where only with the first four assumptions we have a nonsingular covariance matrix Σ_n. Thus, it seems that an assumption of the form of A5 is also necessary.

IV. RESULTS

Before going to the proof of the CLT we will first present some lemmas that will be useful for this proof. Let us denote by A and B two real square matrices of dimension m, with distinct eigenvalues. Let ξ_X denote the eigenvalue of a square matrix X with the maximum magnitude and D_X a constant that depends only on X.

LEMMA 1 *There exist constants D_A, $D_{A,B}$ such that for any two vectors a, b and any integer $k \geqq 0$ we have*

i) $|a^T A^k b| \leqq D_A |\xi_A|^k (a^T a)^{1/2} (b^T b)^{1/2}$

ii) $|a^T (A^k - B^k)(A^k - B^k)^T a|^{1/2} \leqq D_{A,B} |\xi_{A,B}|^k \|A - B\| (a^T a)^{1/2}$

where by $\|X\|$ we denote the maximum singular value of the matrix X.

Proof The proof of (i) is easy. Notice that it is obvious when A is diagonal. When it is not diagonal we make a diagonalization and proceed in a similar way. Inequality (ii) can be proved by using (i) and that $A^k - B^k = A^{k-1}(A-B) + A^{k-2}(A-B)B + \cdots + (A-B)B^{k-1}$. we omit the details.

LEMMA 2 *Let F, W_k satisfy A1 and A3, then we also have that X_k has uniformly bounded fourth order moments.*

Proof From (1) we have that

$$X_k = \sum_{i=0}^{k-1} F^i W_{k-i}$$

thus using the fact that $\{W_k\}$ is a sequence of independent vectors, we have for every vector λ that

$$E\{[\lambda^T X_k]^4\}$$

$$= E\left\{\sum_{i=0}^{k-1}\sum_{j=0}^{k-1}\sum_{l=0}^{k-1}\sum_{s=0}^{k-1} [\lambda^T F^i W_{k-i}][\lambda^T F^j W_{k-j}][\lambda^T F^l W_{k-l}][\lambda^T F^s W_{k-s}]\right\}$$

$$= \sum_{i=0}^{k-1} E\{[\lambda^T F^i W_{k-i}]^4\} + 3\sum_{\substack{j=0 \\ j\neq l}}^{k-1}\sum_{l=0}^{k-1} E\{[\lambda^T F^j W_{k-j}]^2\} E\{[\lambda^T F^l W_{k-l}]^2\}$$

$$\leqq \sum_{i=0}^{k-1} E\{[\lambda^T F^i W_{k-i}]^4\} + 3\left[\sum_{i=0}^{k-1} E\{[\lambda^T F^i W_{k-i}]^2\}\right]^2.$$

Using Lemma 1(i) and A3 and the fact that $|\xi_F| < 1$ (from A1), we have that there exists constant D such that for every k we have

$$E\{[\lambda^T X_k]^4\} \leqq D(\lambda^T \lambda)^2.$$

LEMMA 3 *Let F, W_k satisfy A1 and A3. Consider the following sequence of covariance matrices P_k*

$$P_{k+1} = F P_k F^T + Q_{k+1},\ P_0 = 0. \tag{9}$$

The matrix P_k is the covariance matrix of the state vector X_k. Let now $\bar{Q}_n = 1/n \sum_{k=1}^n Q_k$ and $\bar{P}_n = 1/n \sum_{k=1}^n P_k$. Let P be the solution of the equation $P = FPF^T + \bar{Q}_n$, then $\|\bar{P}_n - P\| \to 0$. The matrix P is nothing *but the state covariance matrix of a stationary system with input covariance $\bar{Q}_n$ (assumed constant).*

Proof Since W_k has uniformly bounded fourth-order moments, it will also have uniformly bounded second order moments, thus from (9), assuming $Q_0 = 0$ we have

$$\bar{P}_n = \frac{1}{n}\sum_{k=1}^{n} P_k = \frac{1}{n}\sum_{k=1}^{n}\sum_{j=0}^{k-1} F^j Q_{k-j}(F^T)^j = \sum_{j=0}^{n-1} F^j\left(\frac{1}{n}\sum_{k=j+1}^{n} Q_{k-j}\right)(F^T)^j$$

$$= \sum_{j=0}^{\infty} F^j \bar{Q}_n (F^T)^j - \sum_{j=n}^{\infty} F^j \bar{Q}_n (F^T)^j - \frac{1}{n}\sum_{j=0}^{n-1} F^j\left(\sum_{k=n-j+1}^{n} Q_k\right)(F^T)^j. \quad (10)$$

The first term in (10) is equal to P, the second is uniformly of order $|\xi_F|^{2n}$ and the third term of order $1/n$. Thus both last terms tend to zero. And this concludes the proof. Using this lemma we can show that we can approximate (order $1/n$) any average of matrices of the form $\bar{V}_n = 1/n \sum_{k=1}^{n} E\{[y_k \cdots y_{k-j}]^T [y_{k-j} \cdots y_{k-j-i}]\}$, by the corresponding matrix V of a stationary system with input covariance $\bar{Q}_n$ (assumed constant). This is true because $E\{y_k y_{k+j}\} = c^T F^j P_k c$ and thus we can approximate every term in $\bar{V}_n$ by the corresponding stationary term.

LEMMA 4 *Let Q, Q' be two nonnegative definite matrices with $Q \geqq Q'$. Let F satisfy A1. Consider the following two systems in stationary situation.*

$$\begin{aligned} X_{k+1} &= FX_k + W_{k+1} \quad & y_k &= c^T X_k \\ X'_{k+1} &= FX'_k + W'_{k+1} \quad & y'_k &= c^T X'_k, \end{aligned} \quad (11)$$

with W_k having covariance Q and W'_k covariance Q'. For any integer N we then have

$$E\{[y_k \cdots y_{k-N}]^T [y_k \cdots y_{k-N}]\} \geqq E\{[y'_k \cdots y'_{k-N}]^T [y'_k \cdots y'_{k-N}]\}. \quad (12)$$

Proof The proof is easy. Since we consider only second order statistics we can decompose W_k into two independent processes V_k, R_k such that $E\{V_k V_k^T\} = Q'$, $E\{R_k R_k^T\} = Q - Q'$ and $E\{V_k R_k^T\} = 0$. Because of linearity, the process y_k can be also decomposed into two independent processes one due to V_k and the other to R_k. The process due to V_k will have exactly the same second order properties with y'_k. Thus we conclude that the left hand side of (12) is equal to the right hand side plus another nonnegative term due to R_k.

LEMMA 5 *Let F, F' be two matrices satisfying A1 and W_k a process satisfying A3. Let X_k be the state vector of a system of the form of* (1) *and X'_k another state vector of a similar system but F replaced by F'. For any vector λ with $\lambda^T\lambda = 1$ we have*

$$E\{[\lambda^T(X_k - X'_k)]^2\} = 0(\|F - F'\|^2)$$

uniformly in k. Where by $z = 0(x)$ we mean z is of the order of x.

Proof By writing $X_k = \sum_{i=0}^{k-1} F^i W_{k-i}$, using uniform boundedness of $\{Q_k\}$ and Lemma 1(ii) we have

$$E\{[\lambda^T(X_k - X'_k)]^2\} = \sum_{i=0}^{k-1} \lambda^T(F^i - (F')^i)Q_{k-i}(F^i - (F')^i)^T\lambda$$

$$\leq p \sum_{i=0}^{k-1} \lambda^T(F^i - (F')^i)(F^i - (F')^i)^T\lambda$$

$$\leq Dp\|F - F'\|^2 \sum_{i=0}^{\infty} |\xi_{F,F'}|^{2i} = D'\|F - F'\|^2.$$

And this concludes the proof of the lemma.

LEMMA 6 *Let s be a nonzero vector, let F satisfy A1 and A2 and let $E\{W_k W_k^T\} = ss^T$. Consider the system in* (1) *in the stationary situation, then for any fixed integer N the covariance matrix*

$$V = E\{[y_k \ldots y_{k-N+1}]^T[y_k \ldots y_{k-N+1}]\}$$

is of full rank.

Proof Let $\lambda = [\lambda_1 \ldots \lambda_N]^T$ be any vector, then

$$\lambda^T V \lambda = \frac{1}{2\pi} \int_0^{2\pi} |\lambda_1 + \cdots + \lambda_N e^{j\omega(N-1)}|^2 |c^T(e^{-j\omega}I - F)^{-1}s|^2 \, d\omega.$$

If $\lambda^T V \lambda = 0$ then we will also have that the product under the integral is identically zero. Since the first term in this product is a trigonometric polynomial, if it is not identically zero, it can have only a finite number of zeros in the interval $[0, 2\pi]$. Thus if we assume that $\lambda \neq 0$ we conclude that $c^T(e^{-j\omega}I - F)^{-1}s$ must be identically zero. But we have that

$$c^T(e^{-j\omega}I - F)^{-1}s = \sum_{i=0}^{\infty} e^{j\omega(i+1)}c^T F^i s.$$

The above can be identically zero only if $c^T F^i s = 0$ for every i. This is not possible since because of A2 we have that for at least one i between zero and $m-1$ we have $c^T F^i s \neq 0$, contradiction. Thus V is of full rank.

With the next theorem we prove the Law of Large Numbers for expressions that appear in the proof of the CLT.

THEOREM 1 *Let F, W_k satisfy A1 and A3, let X_k be the state vector of a system of the form of* (1). *If $\{a_k\}$, $\{b_k\}$ are two uniformly bounded real vector sequences, then for any fixed integer j*

$$A_n = \frac{1}{n}\sum_{k=1}^{n} b_k^T[X_{k+j}X_k^T - E\{X_{k+j}X_k^T\}]a_k \to 0 \quad \text{a.s.}$$

Proof Since $X_{k+j} = F^j X_k + \sum_{i=1}^{j} F^{j-i}W_{k+i}$ and W_{k+i} is independent of X_k for $i \geqq 1$ we have

$$A_n = \sum_{i=1}^{j} \frac{1}{n}\sum_{k=1}^{n} b_k^T F^{j-i} W_{k+i} X_k^T a_k + \frac{1}{n}\sum_{k=1}^{n} b_k^T F^j [X_k X_k^T - E\{X_k X_k^T\}]a_k. \quad (13)$$

The sums of the form $\sum_{k=1}^{n} b_k^T F^{j-i} W_{k+i} X_k^T a_k$ are martingales. Following Feller ([6], page 243) in order to show that these sums normalized by n go a.s. to zero, it is enough to show that

$$\sum_{n=1}^{\infty} \frac{E\{[b_n^T F^{j-i} W_{n+i} X_n^T a_n]^2\}}{n^2} < \infty. \quad (14)$$

But (14) is true because

$$E\{[b_n^T F^{j-i} W_{n+i} X_n^T a_n]^2\} = E\{[b_n^T F^{j-i} W_{n+i}]^2\} E\{[X_n^T a_n]^2\}$$

and from Lemma 2 we have that both terms in the product are uniformly bounded. Thus the first j terms in (13) go to zero a.s. To show that this is also true for the last term we will use the theory of mixingales (see Hall [7], page 41). If $\{\mathscr{F}_n\}_{n\geq 1}$ is a sequence of σ-algebras with $\mathscr{F}_n$ the σ-algebra generated by $\{W_1 \dots W_n\}$ and if $V_n = b_n^T F^j [X_n X_n^T - E\{X_n X_n^T\}] a_n$ then if we can find $\psi_k \to 0$ and $f_n \geq 0$ such that

1) $\|V_n - E\{V_n/\mathscr{F}_{n+k}\}\|_2 \leq \psi_{k+1} f_n$
2) $\|E\{V_n/\mathscr{F}_{n-k}\}\|_2 \leq \psi_k f_n$
3) $\sum_{n=1}^{\infty} \frac{f_n}{n^2} < \infty$ and $\psi_k = o(k^{-1/2}(\log k)^{-2})$

then $1/n \sum_{k=1}^{n} V_k \to 0$ a.s. For a proof of this statement see Hall ([7], page 41). We will show that we can define f_n and ψ_k to satisfy conditions 1 through 3. First notice that 1 is trivially satisfied for any nonnegative ψ_k, f_n because $E\{V_n/\mathscr{F}_{n+k}\} = V_n$. Let us now for simplicity denote $d_n^T = b_n^T F^j$, then d_n is also uniformly bounded. We can see that

$$E\{V_n/\mathscr{F}_{n-k}\} = d_n [F^k (X_{n-k} X_{n-k}^T - E\{X_{n-k} X_{n-k}^T\})(F^T)^k] a_n. \quad (15)$$

Using the fact that the second moment of a random variable is larger than the variance we have from (15)

$$\|E\{V_n/\mathscr{F}_{n-k}\}\|_2^2 \leq E\{[d_n^T F^k X_{n-k}]^2 [a_n^T F^k X_{n-k}]^2\}$$

$$\leq E\{[d_n^T F^k X_{n-k}]^4\} + E\{[a_n^T F^k X_{n-k}]^4\}. \quad (16)$$

Using now Lemma 1(i) and Lemma 2 we have from (16) that there exists a constant D such that $\|E\{V_n/\mathscr{F}_{n-k}\|_2^2 \leq D^2 |\xi_F|^{4k}$. We can thus define $f_n = D$ and $\psi_k = |\xi_F|^{2k}$. Clearly with this definition we have validity of conditions 2 and 3 and thus the Strong Law of Large Numbers holds also for the last term in (13).

We are now ready to prove the CLT defined by (4). This is done in the following theorem.

THEOREM 2 *Let $\{y_k\}$ be the observation process defined by a system of the form of* (1). *If F_0 and W_k satisfy conditions A1 through A5, then* (4) *holds.*

Proof We first show the CLT under hypothesis H_0. We will use a version of the martingale CLT, thus we will put U_n defined in (3) under a martingale form. Let λ be a vector with $\lambda^T\lambda = 1$. What we would like to show is that $\lambda^T \Sigma_n^{-1/2} U_n \xrightarrow{d} N(0, 1)$. Notice that

$$E\{[\lambda^T \Sigma_n^{-1/2} U_n]^2\} = 1, \tag{17}$$

this is true because Σ_n was defined as the covariance of U_n. From (17) we see that we have the right variance required by the CLT. Dropping for simplicity the subscript "0" we have from (1)

$$y_{k-i} = c^T F^{m-i-1} X_{k-m+1} + c^T \sum_{j=0}^{m-i-2} F^j W_{k-j}. \tag{18}$$

Substituting (18) in (3) and using the Cayley–Hamilton theorem we have

$$\lambda^T \Sigma_n^{-1/2} U_n = \frac{\lambda_n^T}{\sqrt{n}} \sum_{k=1}^{n} [c_1^T W_{k+1} + c_2^T W_k + \cdots + c_m^T W_{k-m+2}] Z_k \tag{19}$$

where $\lambda_n^T = \lambda^T \Sigma_n^{-1/2}$ and $c_i^T = c^T [F^{i-1} - \alpha_1 F^{i-2} - \cdots - \alpha_{i-1} I]$. Let us for the moment assume that the covariance Σ_n is uniformly bounded away from zero for large enough n. We will prove this statement in Theorem 3. With this assumption λ_n is uniformly bounded. Now rearranging the sum in (19) we have

$$\lambda^T \Sigma_n^{-1/2} U_n = \frac{\lambda_n^T}{\sqrt{n}} \sum_{k=1}^{n} [Z_k c_1^T + \cdots + Z_{k+m-1} c_m^T] W_{k+1} + 0\left(\frac{1}{\sqrt{n}}\right). \tag{20}$$

Since in (20) W_{k+1} is independent with whatever is in the brackets, the sum in (20) is a martingale. Defining

$$v_k = \lambda_n^T [Z_k c_1^T + \cdots + Z_{k+m-1} c_m^T] W_{k+1} \tag{21}$$

in order to show the CLT, it is enough to show Lindeberg's condition for v_k and that

$$\frac{1}{n} \sum_{k=1}^{n} v_k^2 \to 1 \tag{22}$$

in probability (see Hall [7], page 52). Lindeberg's condition is the following, for any $\varepsilon > 0$ we want

$$\sum_{k=1}^{n} Pr\left\{\frac{v_k^2}{n} I(|v_k| > \varepsilon\sqrt{n})\right\} \to 0$$

where with $I(A)$ we denote the indicator function of the set A. To show Lindeberg's condition we have

$$Pr\left\{\frac{v_k^2}{n} I(|v_k| > \varepsilon\sqrt{n})\right\} \leqq \frac{E\{v_k^4\}}{\varepsilon^4 n^2}.$$

Applying now the Schwarz inequality at (21) yields $v_k^2 \leqq [(\lambda_n^T Z_k)^2 + \cdots + (\lambda_n^T Z_{k-m+1})^2][(c_1^T W_{k+1})^2 + \cdots + (c_m^T W_{k+1})^2]$. And thus using Lemma 2 and the independence of W_{k+1} with the Z_{k+i} vectors, we can easily see that $E\{v_k^4\}$ is uniformly bounded by a constant D. This yields

$$\sum_{k=1}^{n} Pr\left\{\frac{v_k^2}{n} I(|v_k| > \varepsilon\sqrt{n})\right\} \leqq \frac{D}{\varepsilon^4 n} \to 0,$$

and thus Lindeberg's condition holds. To show (22) since using (17) and (20) we have $1/n\sum_{k=1}^{n} E\{v_k^2\} = E\{[\lambda^T \Sigma_n^{-1/2} U_n]^2\} + 0(1/n) = 1 + 0(1/n)$, it is enough to show that $1/n\sum_{k=1}^{n}[v_k^2 - E\{v_k^2\}] \to 0$ in probability. The random variable v_k^2 is a finite sum of terms of the form $b_n y_{k-i} y_{k-j}(c_p^T W_{k+1})(c_q^T W_{k+1})$ where b_n depends on the vector λ_n. Thus we can write

$$\frac{b_n}{n}\sum_{k=1}^{n} [y_{k-i}y_{k-j}(c_p^T W_{k+1})(c_q^T W_{k+1}) - E\{y_{k-i}y_{k-j}\}E\{(c_p^T W_{k+1})(c_q^T W_{k+1})\}]$$

$$= \frac{b_n}{n}\sum_{k=1}^{n} y_{k-i}y_{k-j}[(c_p^T W_{k+1})(c_q^T W_{k+1}) - E\{(c_p^T W_{k+1})(c_q^T W_{k+1})\}]$$

$$+ \frac{b_n}{n}\sum_{k=1}^{n} E\{(c_p^T W_{k+1})(c_q^T W_{k+1})\}[y_{k-i}y_{k-j} - E\{y_{k-i}y_{k-j}\}]. \tag{23}$$

The first sum in (23) is a martingale. Using similar reasoning as in Theorem 1 we can show that it converges to zero a.s. For the second term in (23) we apply Theorem 1 and thus this term also goes to zero a.s. With these arguments we have shown that (22) is true not only in probability but a.s.

Up to this point we have shown the CLT under H_0. To prove it also under H_1, notice that the observation sequence $\{y_k\}$ is now generated not by the nominal matrix F_0 but by the matrix F that satisfies $\|F-F_0\|=0(1/\sqrt{n})$. Let us call $\{y_k^0\}$ the observation sequence generated by the nominal matrix F_0 when as input we have exactly the same like the system with F. If we show for every vector λ with $\lambda^T\lambda=1$ that we have

$$E\left\{\left[\frac{1}{\sqrt{n}}\sum_{k=1}^{n}(y_{k+1}-\alpha^T Y_k)\lambda^T Z_k-\frac{1}{\sqrt{n}}\sum_{k=1}^{n}(y_{k+1}^0-\alpha_0^T Y_k^0)\lambda^T Z_k^0\right]^2\right\}\to 0, \tag{24}$$

where Y_k^0, Z_k^0 are the Y_k, Z_k vectors corresponding to the nominal system, then this means that the two terms in (24) have the same asymptotic distribution. Since the second term, as it was shown in the first part of this theorem, is asymptotically Gaussian, the same will be true for the first term as well. To show (24) define

$$\omega_k=(y_{k+1}-\alpha^T Y_k)\lambda^T Z_k-(y_{k+1}^0-\alpha_0^T Y_k^0)\lambda^T Z_k^0.$$

The terms $y_{k+1}-\alpha^T Y_k$ and $y_{k+1}^0-\alpha_0^T Y_k^0$ are the MA parts of the two observation processes and since they are at most $(m-1)$-dependent we have that $E\{\omega_k\omega_{k-j}\}=0$ for $j\geqq m$. Thus (24) is equivalent to

$$\frac{1}{n}E\left\{\left[\sum_{k=1}^{n}\omega_k\right]^2\right\}=\frac{1}{n}\sum_{k=1}^{n}E\{\omega_k^2\}+\frac{2}{n}\sum_{j=1}^{m-1}\sum_{k=1}^{n}E\{\omega_k\omega_{k-j}\}\to 0. \tag{25}$$

Since $2|\omega_k\omega_{k-j}|\leqq\omega_k^2+\omega_{k-j}^2$, in order to prove (25) it is enough to prove that the first term in (25) goes to zero. Notice that we have

$$\omega_k=(y_{k+1}-\alpha^T Y_k)[\lambda^T(Z_k-Z_k^0)]+[(y_{k+1}-\alpha^T Y_k)-(y_{k+1}^0-\alpha_0^T Y_k^0)]\lambda^T Z_k^0.$$

The two MA parts are independent of Z_k and Z_k^0, thus

$$E\{\omega_k^2\}\leqq 2E\{[y_{k+1}-\alpha^T Y_k]^2\}E\{[\lambda^T(Z_k-Z_k^0)]^2\}$$
$$+2E\{[\lambda^T Z_k]^2\}E\{[(y_{k+1}-\alpha^T Y_k)-(y_{k+1}^0-\alpha_0^T Y_k^0)]^2\}. \tag{26}$$

Using Lemma 5 the quantities y_{k+1}, Y_k and Z_k differ from the corresponding nominal quantities (in the mean square sense) by an amount $0(\|F-F_0\|^2)=0(1/n)$. Thus for every $k \leqq n$ we have $E\{\omega_k^2\}=0(1/n)$ which yields $1/n\sum_{k=1}^{n} E\{\omega_k^2\}=0(1/n)$. And thus we have shown that (24) is true. What is left now to prove in order for Theorem 2 to be complete is that the covariance matrix Σ_n is uniformly bounded away from zero. This is shown in Theorem 3.

THEOREM 3 *Let F_0, W_k satisfy A1 through A5. Let Σ_n be the covariance matrix of U_n defined in* (3) *and H_n the matrix defined in* (5), *then for large enough n the matrix Σ_n is uniformly of full rank and uniformly bounded from above also H_n is a.s. uniformly of full rank.*

Proof The proof is based on certain properties that hold under stationarity. From Eq. (20) we have that

$$U_n=\frac{1}{\sqrt{n}}\sum_{k=1}^{n}[Z_k c_1^T+\cdots+Z_{k+m-1}c_m^T]W_{k+1}+0\left(\frac{1}{\sqrt{n}}\right).$$

We first prove that Σ_n is full rank. Let λ be a vector with $\lambda^T\lambda=1$. Let us call $t_k=\lambda^T Z_k$ and $T_k^T=[t_k\, t_{k+1}\ldots t_{k+m-1}]$, then

$$\lambda^T U_n=\frac{1}{\sqrt{n}}\sum_{k=1}^{n} T_k^T\bar{0}W_{k+1}+0\left(\frac{1}{\sqrt{n}}\right),$$

where $\bar{0}^T=[c_1\ldots c_m]$. Using A5 and that W_{k+1} is independent of T_k, we have

$$E\{[\lambda^T U_n]^2\}\geqq\frac{1}{n}\sum_{k=1}^{n}E\{[T_k^T\bar{0}s]^2\}+0\left(\frac{1}{n}\right)=\frac{1}{n}\sum_{k=1}^{n}E\{[T_k^T l]^2\}+0\left(\frac{1}{n}\right), \quad (27)$$

where $l=\bar{0}s$. Notice that $\bar{0}$ comes from the matrix 0 defined in A2 by using linear operations. Since 0 is assumed full rank so is $\bar{0}$, thus the vector l is nonzero. Let $l^T=[l_1\ldots l_m]$ and $\lambda^T=[\lambda_1\ldots\lambda_M]$, then we can see that

$$T_k^T l=[\lambda_1\ldots\lambda_M]\begin{bmatrix} l_m & l_{m-1} & \ldots & l_1 & 0 & \ldots & 0\\ 0 & l_m & \ldots & l_2 & l_1 & \ldots & 0\\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots\\ 0 & 0 & \ldots & 0 & l_m & \ldots & l_1\end{bmatrix}\begin{bmatrix} y_k\\ y_{k-1}\\ \vdots\\ y_{k-m-M+2}\end{bmatrix} \quad (28)$$

Denote with L the matrix in (28). This matrix is of dimensions $M\times(M+m-1)$. It is easy to see that $\lambda^T L$ gives the convolution of the two sequences $\{\lambda_1,\ldots,\lambda_M\}$ and $\{l_m,\ldots,l_1\}$. It is known that the convolution of two sequences cannot be identically zero unless one of the two sequences is. Since $l\neq 0$ this means that there is no $\lambda\neq 0$ such that $\lambda^T L=0$ or, that L is of full rank. Let us call $\sigma_L>0$ the smallest singular value of LL^T. Now define

$$\bar{V}_n=\frac{1}{n}\sum_{k=1}^{n} E\{[y_k\cdots y_{k-m-M+2}]^T[y_k\cdots y_{k-m-M+2}]\}, \tag{29}$$

the average covariance matrix of the random vector in (28). From Lemma 3 $\bar{V}_n$ can be approximated (order $1/n$) by the corresponding covariance of a stationary system with input $\bar{Q}_n$. Since from A5 we also have $\bar{Q}_n\geqq ss^T$, using Lemma 4 this last covariance can be lower bounded by the corresponding covariance V of a stationary system with input ss^T. The covariance V from Lemma 6 is of full rank. Call $\sigma_V>0$ the smallest singular value of V. Going back to (27) and using (28) and (29) we have

$$E\{[\lambda^T U_n]^2\}\geqq\lambda^T L\bar{V}_n L^T\lambda\geqq\sigma_V\sigma_L+0\left(\frac{1}{n}\right), \tag{30}$$

and thus for large enough n, Σ_n is uniformly bounded away from zero. Notice that for the proof of the nonsingularity of Σ_n we did not use A4. To prove that Σ_n uniformly bounded from above we proceed in a similar way but we use the uniform upper bounds of the covariance matrices involved. We omit the details.

To prove that H_n is a.s. of full rank notice that because of Theorem 1 we have that $H_n-E\{H_n\}\to 0$ a.s., thus it is enough to show that $E\{H_n\}$ is of full rank. Using Lemma 4, $E\{H_n\}$ can be approximated by the corresponding matrix H of a stationary system with input $\bar{Q}_n$. We will show that this matrix H is of full rank. We have that $H=E\{Y_k Z_k^T\}$, where Y_k and Z_k are now generated by a stationary system with input $\bar{Q}_n$ (assumed constant). We would like to show that there exists real $\sigma>0$ such that for any vector λ with $\lambda^T\lambda=1$ we have $\lambda^T HH^T\lambda\geqq\sigma$. Since Z_k has length $M\geqq m$ we consider only the first m components of this vector, that is, we will show that $E\{Y_k Y_{k-m+1}^T\}$ is of full rank. Consider the vector $E\{\lambda^T Y_k Y_{k-m+1}^T\}$;

using similar ideas as in [1, 10], the ith component of this vector (say d_i) is given by

$$d_i = \frac{1}{2\pi j} \oint (\lambda_1 + \cdots + \lambda_m z^{-m+1}) c^T (z^{-1}I - F_0)^{-1} \bar{Q}_n (zI - F_0)^{-1} c z^{m-2+i} \frac{dz}{z}$$

$$= \frac{1}{2\pi j} \oint \lambda(z) c^T (z^{-1}I - F_0)^{-1} \bar{Q}_n (zI - F_0)^{-1} c z^{i-1} \frac{dz}{z} \quad i = 1, \ldots, m \qquad (31)$$

where $\lambda(z) = \lambda_1 z^{m-1} + \cdots + \lambda_m$ and the integration path is the unit circle. The only poles inside the unit circle in (31) are the eigenvalues r_i of F_0. Thus from (31)

$$d_i = \sum_{k=1}^{m} \frac{\lambda(r_k)}{r_k} \mu_k r_k^{i-1} \quad i = 1, \ldots, m \qquad (32)$$

where $\mu_k = \lim_{z \to r_k} \{(z - r_k) c^T (z^{-1} - F_0)^{-1} \bar{Q}_n (zI - F_0)^{-1} c\}$. Since the Vandermonde matrix

$$V_r = \begin{bmatrix} 1 & r_1 & \ldots & r_1^{m-1} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & r_m & \ldots & r_m^{m-1} \end{bmatrix},$$

is of full rank, call σ_r its smallest singular value. Notice that the vector $[\lambda(r_1) \ldots \lambda(r_m)]$ can be written as $\bar{\lambda}^T V_r$, where $\bar{\lambda}^T = [\lambda_m \ldots \lambda_1]$. Let us call

$$\rho^T = \left[\frac{\lambda(r_1)}{r_1} \mu_1 \ldots \frac{\lambda(r_m)}{r_m} \mu_m \right],$$

then using A4, i.e. that $|\mu_k| \geqq \delta$ and also from A1 that $|r_k| < 1$ we have

$$\sum_{i=1}^{m} d_i^2 = \rho^T V_r (V_r^T \rho)^* \geqq \sigma_r^2 \rho^T \rho^* \geqq \sigma_r^2 \delta^2 \sum_{k=1}^{m} |\lambda(r_k)|^2$$

$$= \sigma_r^2 \delta^2 \bar{\lambda}^T V_r^T V_r^* \bar{\lambda} \geqq \sigma_r^4 \delta^2 \lambda^T \lambda,$$

where by the superscript "*" we mean complex conjugate. And this concludes the proof. As we have proved, Σ_n is uniformly bounded from above and H_n is uniformly of full rank, thus we have that $H_n^T \Sigma_n^{-1} H_n$ is also uniformly bounded away from zero.

V. CONCLUSION

We have presented a method for detecting changes in the AR parameters of a nonstationary ARMA process. The detection scheme was derived by using the same idea that is used for the estimation of these parameters, i.e. that the MA part of an ARMA process is independent from delayed enough observations. Following a local asymptotic approach the detection of a change in the AR parameters was reduced to the detection of a nonzero mean of a Gaussian random vector. With the assumptions we have introduced here we can actually show a stronger result, namely that the CLT in (4) remains valid if Σ_n is replaced by $\hat{\Sigma}_n$. This is true basically because it is possible to show that the difference between the two matrices goes to zero a.s. The proof of this last statement is unfortunately long, thus we have decided to present only the weaker version of the CLT defined in (4). To say a few things about estimating the covariance matrix Σ_n. Even though $\hat{\Sigma}_n$ has expectation equal to Σ_n it has the drawback that it is not always positive definite. This can lead to a negative test statistics T_n (Eq. 7). For practical purposes we can for example use only the first term of $\hat{\Sigma}_n$ which is positive definite. This method was successfully used in detecting changes in vibrating modes of linear systems [3]. Finally another practical problem is the knowledge of the system order m. In simulations our method performed very well even when the true order was overestimated.

References

[1] K. J. Åström and T. Söderström, Uniqueness of the maximum likelihood estimates of the parameters of an ARMA model, *IEEE Trans. on Automatic Control* **AC-19**(6) (1974), 769–772.

[2] M. Basseville and A. Benveniste, Sequential detection of abrupt changes in spectral characteristics of digital signals, *IEEE Trans. on Information Theory* **IT-29** (1983), 709–723.

[3] M. Basseville, A. Benveniste and G. V. Moustakides, Detection and diagnosis of abrupt changes in modal characteristics of nonstationary digital signals, *Rapports de Recherche INRIA*, No. 348, (1984).

[4] A. Benveniste and J. J. Fuchs, Single sample modal identification of a non-stationary stochastic process, *IEEE Trans. on Automatic Control* **AC-30** (1985), 66–75.

[5] R. B. Davies, Asymptotic inference in stationary Gaussian time-series, *Adv. Appl. Prob.* **5** (1973), 469–497.

[6] W. Feller, *An Introduction to Probability Theory and its Applications*, Vol. II, second edition, J. Wiley, New York, 1966.

[7] P. Hall and C. C. Heyde, *Martingale Limit Theory and its Application*, Academic Press, New York, 1980.

[8] I. V. Nikiforov, Modification and analysis of the cumulative sum procedure, *Automatica i Telemekanika* **41**(9) (1980), 74–80.

[9] G. G. Roussas, *Contiguity of Probability Measures: Some Applications in Statistics*, Cambridge University Press, Cambridge, 1972.

[10] P. Stoica and T. Söderström, Optimal instrumental variable estimation and approximate implementations, *IEEE Trans. on Automatic Control* **AC-28**(7) (1983), 757–772.

[11] A. S. Willsky, A survey of design methods for failure detection in dynamic systems, *Automatica* **12** (1976), 601–611.