ADAPTIVE AUTOREGRESSIVE MODELING FOR TIME-FREQUENCY ANALYSIS

Stephan Hengstler$^1$, Stephan Sand$^2$, and Antonio H. Costa$^2$

$^1$Agilent Technologies, Inc.
Santa Clara, California 95052-8059, USA
stephan.hengstler@ieee.org

$^2$Department of Electrical and Computer Engineering
University of Massachusetts Dartmouth
North Dartmouth, Massachusetts 02747-2300, USA
acosta@umassd.edu

ABSTRACT

A new adaptive method for discrete time-frequency analysis based on autoregressive (AR) modeling is proposed. The performance of adaptive AR modeling depends heavily upon a good selection of two parameters — the forgetting factor and the model order. A previously introduced method of AR modeling with variable forgetting factors employs an extended prediction error criterion that accounts for the nonstationarity of the signal. The predictive least squares (PLS) principle of Rissanen has been found to be a good criterion for model order estimation in the case of stationary processes. A modified formulation of the PLS criterion is suggested for nonstationary processes. Efficient lattice filters based on the covariance method are used to estimate the model parameters of all model orders less than some maximum order $M$. The resulting complexity is no larger than that of the maximum model order $M$. The combination of variable forgetting factors and the modified PLS criterion allows the AR model parameters to effectively adapt to nonstationary processes. Examples of time-frequency analysis are provided for two synthetic signals to demonstrate the feasibility of the novel method of adaptive AR modeling.

1. INTRODUCTION

In recent years, there has been some interest in developing time-frequency representations (TFRs) with an underlying parametric model. For stationary signals, the parametric AR modeling approach yields high resolution power spectral density (PSD) estimates even for short data records. Furthermore, researchers have developed very efficient adaptive time- and order-recursive algorithms to compute the AR coefficients of nonstationary signals ([1, 2]).

Based on an AR model, Griffiths [3] derived an adaptive gradient descent algorithm for computing the instantaneous frequency of nonstationary signals. Hodgkiss [4] investigated the behavior of the gradient transversal filter, the gradient lattice and the least squares lattice when used to track time-varying sinusoidal components. He demonstrated that the recursive least squares approach with a forgetting factor performs as well as the gradient methods or better. In [5] a recursive least squares algorithm with a sliding window was presented to estimate the PSD of nonstationary signals. However, the author did not implement criteria to estimate the order, “because they considerably” increased “the computation time” [5]. Cho et al. [6] introduced a variable forgetting factor that adapted to the time-varying signal by an extended prediction error criterion accounting for the nonstationarity of the signal. Using a Kalman filter to obtain the time-varying AR coefficients, the authors of [7] proposed an equivalent method to the least squares lattice in [4] for computing an instantaneous power spectral density.

None of the AR based methods cited above estimate the model order. However, the performance of an AR model based TFR greatly depends upon a good selection of the model order. In this paper, we present a strategy for adaptively choosing the forgetting factor in conjunction with a modified predictive least squares criterion to determine a good estimate for the model order. We then compute the TFR of nonstationary signals using lattice filters based on the covariance assumption.

2. TIME-VARYING AR MODELING

A discrete-time nonstationary process can be modeled by a time-varying autoregressive (TVAR) process of model order $m$ represented by the linear difference equation

$$x[n] = -\sum_{k=1}^{m} a_n^{-1}[k]x[n-k] + u[n]$$

(1)

where $u[n]$ is a stationary white noise process with zero mean and variance $\sigma^2_u$ [8]. The TVAR parameters $a_n^{-1}[k], k = 1, \ldots, m$ are commonly found as the least squares solution of the accumulated squared prediction error. The exact definition of this error criterion depends on the specific approach chosen but, in general, it is a properly weighted sum of the error between the true sample value $x[n]$ and the predicted sample value $\hat{x}[n]$, i.e.,
\[ I_\lambda(x, \hat{x}) = \sum_{i=1}^{t} \lambda^{i-1} \left| x_i - \hat{x}_i \right|^2 \]  \hspace{1cm} (2)

where the limits of summation depend on the type of data windowing employed. One common approach, namely the covariance method, will be considered below. The forgetting factor \( \lambda \), where \( 0 < \lambda \leq 1 \), may be used to weigh recent data more heavily in the least squares computation and thus forget about past prediction errors in an exponentially decaying manner.

At each time instant \( n \), the time dependent model (1) admits a stationary model [8], which allows the computation of the instantaneous power spectral density (IPSD) [7]. This can be attained by taking the z-transform of (1) and evaluating it along the unit circle. Hence, the resulting spectral estimate at time index \( n \) is given by

\[ P(f) = \frac{\sigma^2}{1 + \sum_{k=1}^{m} a_k \sigma^2} \].  \hspace{1cm} (3)

However, note that all information about the process is contained in the parameter estimates \( a_k \) \( k = 1, \ldots, m \) and \( \sigma^2 \). Any other means to represent this information is redundant and serves only as a tool to visualize and interpret the information.

### 3. FORGETTING FACTOR ADAPTATION

A properly chosen forgetting factor \( \lambda \) enables the AR modeling to grasp the local statistics of the process under consideration, i.e., to obtain a local estimate of its AR parameters. If it is known a priori that the process is stationary \( \lambda \) is set equal to 1, i.e., all data is weighed equally. In the case of nonstationary processes, a forgetting factor \( \lambda < 1 \) ensures a good local convergence of the estimated AR parameters to the optimal ones. This is accomplished by weighing past prediction errors less than more recent ones.

Although \( \lambda \) can be set to a fixed value throughout the computation, an adaptive implementation of the forgetting factor may account for changing degrees of nonstationarity throughout a process. With a small value of \( \lambda \), the estimated AR parameters reflect the global trend quickly at the expense of higher variance due to a smaller amount of available data. On the other hand, a value of \( \lambda \) close to unity delays the estimation of the true AR parameters but the parameters are estimated with higher accuracy. The speed of adaptation may be expressed by the asymptotic memory length

\[ L_{\text{eff}} = \frac{1}{1-\lambda}. \]  \hspace{1cm} (4)

To implement a variable forgetting factor \( \lambda[n] \) into the AR parameter estimation, Cho et al. [6] suggested determining the actual forgetting factor \( \lambda[n] \) based on the local degree of nonstationarity of the process. For this purpose, the extended prediction error

\[ q[n] = \frac{1}{L_{\text{avg}}} \sum_{i=1}^{L_{\text{avg}}-1} \left| e[n-i] \right|^2 \]  \hspace{1cm} (5)

provides a measure of the noise variance at time index \( n \) where \( e[n] \) denotes the true prediction error between actual and predicted data samples. The averaging length \( L_{\text{avg}} \) is used to cancel out the periodicity of the prediction error and to average the prediction noise.

The strategy for adapting the forgetting factor \( \lambda[n] \) can now be expressed by

\[ \lambda[n] = 1 - \rho^{-1}[n] \]  \hspace{1cm} (6)

where

\[ \rho[n] = \frac{\hat{\sigma}^2 L_0}{Q[n]} \]  \hspace{1cm} (7)

may be interpreted as the degree of nonstationarity at time index \( n \). The AR parameter computation also provides an estimate of the noise variance \( \hat{\sigma}^2 \). Note that in contrast to the strategy proposed in [6], this approach does not require real knowledge of the process noise variance \( \sigma^2 \). The asymptotic memory length \( L_0 \) determines the speed of adaptation and is usually chosen much larger than the averaging length \( L_{\text{avg}} \). The strategy given by (6) and (7) does not guarantee \( \lambda[n] \) to be positive for all \( n \). Thus, in practice, the adaptive forgetting factor \( \lambda[n] \) is bounded by some reasonable lower limit \( \lambda_{\text{min}} \) where \( 0 < \lambda_{\text{min}} < 1 \).

### 4. MODEL ORDER SELECTION

In addition to the estimation of the AR model parameters with a variable forgetting factor, the selection of an appropriate model order \( m \) in (1) is an important issue in time series modeling. The model order determines the amount of memory required to model the process. Higher model orders may capture more process details but essentially require more data to accurately estimate the parameters.

One criterion for model order estimation proposed by Rissanen [9] is usually referred to as the predictive least squares (PLS) principle. Wax [2] proposed an efficient implementation of this principle based on a prewindowed type least squares lattice algorithm. Chambers and Nandi [10] extended the PLS principle to the covariance assumption. Moreover, they have shown by extensive simulation that the PLS principle based on either windowing assumption is a valid criterion for model order estimation.

The PLS criterion makes use of the true prediction error [10], which for an AR model is defined as the error in the prediction of \( x[n] \) based only upon past observed data. In contrast to recursive a posteriori predictions, this prediction error is sometimes referred to as the a priori error and it is calculated from [9]
\[ \hat{e}_n[n] = x[n] + \sum_{k=1}^{n} a_n^{-k} [k] x[n-k]. \]  

The conventional PLS order selection criterion [9] chooses the model order \( \hat{m} \) with the smallest mean squared a priori error, i.e.,
\[ \hat{m}[n] = \min_{l \in \{1,M\}} \{PLS_{\alpha}[n]\} \]  

where
\[ PLS_{\alpha}[n] = \frac{1}{n-t_0} \sum_{l=t_0}^{n} \|e^l[n]\|^2. \]

The lower summation index \( t_0 \) is determined by the windowing assumption chosen, and \( M \) is the maximum possible model order. Specifically, for the prewindowed scheme \( t_0 \) is set to zero and it equals \( 2M \) for the covariance scheme.

A modification to this conventional PLS criterion is proposed here to take the nonstationarity of processes properly into account. This modified PLS (MPLS) criterion can analogously to (9) and (10) be denoted by
\[ \hat{m}[n] = \min_{l \in \{1,M\}} MPLS_{\alpha}[n] \]

where
\[ MPLS_{\alpha}[n] = \frac{1}{M - m + 1} \sum_{l=m+1}^{n} \{l|l-1 \cdots \lambda[n-1]e^l[n]\|^2. \]

The important modification is to incorporate the forgetting factor \( \lambda[n] \) in the order selection criterion. Again, this weighs past prediction errors less and thus allows the order to adapt to the process as time advances. Since the processes of interest are not restricted to be of the AR type only but any process that can be appropriately modeled by an AR model, the penalty term \((M-m+1)^{-1}\) is added to account for selection of higher order models.

5. ALGORITHMIC ISSUES

The application of the MPLS criterion requires the computation of the a priori prediction error of all model orders up to some maximum order \( M \). Although a large class of least squares (LS) algorithms can be used to solve this task, recursive least squares (RLS) algorithms based on lattice structures offer a very efficient implementation. This paper considers one of the RLS algorithms from [1], namely the growing memory covariance least squares (GMCLS) lattice algorithm. Other implementations may also be possible.

In the case of the GMCLS algorithm with variable forgetting factor, the prediction error energy — the error criterion to be minimized — is defined as
\[ E_{\alpha}[n] = \sum_{l=m}^{n} \{l|l-1 \cdots \lambda[n-1]e^l[n]\|^2 + \sum_{k=1}^{n} a_n^{-k} [k] x[n-k]\|^2. \]

Employing this definition, the covariance method makes no assumptions about the unknown data, i.e., the data outside the interval \([l,n]\).

The algorithm mentioned above can be viewed as operating \( M \) separate adaptive filters and using the AR parameter estimates of that filter selected by the MPLS criterion at each time instant. Additionally, the forgetting factor is updated according to the strategy described previously.

At each iteration \( n > M \), the algorithm requires \( O(M^{-1}) \) arithmetic operations to compute the AR coefficients up to order \( M \). Then the local PSD estimate of \( N_{PSD} \) frequency bins is computed using an algorithm similar in structure to the FFT. With \( N_{PSD} \) being a power of two, it requires arithmetic operations in the order of \( N_{PSD} \cdot \log_2 N_{PSD} \). Therefore, the total computational complexity of the algorithm applied to a signal with \( N \) samples is \( O(N(M^{-2} + N_{PSD} \cdot \log_2 N_{PSD})) \). This facilitates the application of this algorithm in real-time, since \( M \) and \( N_{PSD} \) are usually much smaller than \( N \).

6. SIMULATION RESULTS

To show the performance of the proposed time-frequency analysis based on adaptive AR modeling, we applied the GMCLS algorithm to several synthetic signals. Simulation results for two synthetic signals are given below. Intensity plots in the joint time-frequency plane are used to visualize the results. In addition, the estimated model order and the variable forgetting factor are included.

The first synthetic signal is generated by two complex sinusoids and one complex linear chirp in white Gaussian noise (WGN). Specifically, the signal is 1024 samples long and its components are arranged as follows:

1. Complex sinusoid at \( f = +0.15 \) switched on at time index \( n = 1 \) and off at time index \( n = 512 \);
2. Complex sinusoid at \( f = -0.20 \) beginning at \( n = 1 \) and ending at \( n = 512 \); and
3. Complex linear chirp from index \( n = 513 \) to index \( n = 1024 \) starting at a normalized frequency of \( f = -0.10 \) and ascending at a chirp rate of \( 0.08 \cdot 512^{-1} \).

Next, the signal is degraded by white Gaussian noise resulting in an SNR of 10 dB.

Fig. 1 shows the result of the time-frequency analysis using the novel adaptive GMCLS algorithm in logarithmic scale. We chose a maximum order of \( M = 8 \), an averaging length \( L_{avg} = 2 \), an asymptotic memory length \( L_0 = 80 \), and a minimum forgetting factor of \( \lambda_{min} = 0.96 \). In the bottom plot the estimated model order (dashed line) and the variable forgetting factor (solid line) are plotted over the same time scale. To the right of the time-frequency plot, a color bar indicates the scale of the power density in decibels (dB). Note that we used a short segment of the process considered to initialize the proposed adaptive RLS algorithm, which is not shown in Fig. 1.

Between \( n = 1 \) and \( n = 512 \), the forgetting factor oscillates around a relatively large value of \( \lambda[n] = 0.985 \), thus capturing the stationary sinusoidal components. As the sinusoids are switched off and the linear chirp starts causing
a high degree of nonstationarity, the forgetting factor instantaneously decreases. Even this short segment with a significantly reduced forgetting factor allows the GMCLS algorithm to “forget” about the sinusoidal components and to adapt to the chirp component. Due to the presence of the linear chirp — a slowly varying process — $\lambda[n]$ varies around a smaller value than before. The same trend is visible for the estimated model order. The MPLS criterion determines a model order of three as the result of two sinusoidal components in fairly high noise. But as one reduces the number of signal components within the process by one, the estimated model order also decreases by one almost instantly. This is mainly made possible by the adaptive forgetting factor. With a fixed forgetting factor during the entire computation, the TFR and the model order estimate would exhibit either a considerably higher variance or an increased amount of memory length.

In a second simulation example, we further investigated the performance of the novel adaptive AR method for time-frequency analysis. The signal considered is composed of one complex sinusoid, one complex linear chirp, one complex quadratic chirp, and a complex-valued Gaussian component. As for the previous example, the signal is degraded by white Gaussian noise resulting in an SNR of 20 dB. Specifically, the signal with a length of 1024 samples contains the following components:

1. Complex quadratic chirp ranging from time index $n = 1$ to time index $n = 512$ with center at $n = 512$, $f = -0.20$ and a second order coefficient of $0.4 \cdot 512^{-2}$;
2. Complex Gaussian component between $n = 257$ and $n = 448$ centered at $n = 352$, $f = +0.05$ and with a “standard deviation” of 32 samples;
3. Complex sinusoid at $f = -0.20$ beginning at $n = 513$ and ending at $n = 1024$;
4. Complex linear chirp from index $n = 513$ to index $n = 1024$ starting at a normalized frequency of $f = +0.15$ and descending at a chirp rate of 0.1·512$^{-1}$.

Subsequently, we added white Gaussian noise to this signal resulting in an SNR of 20 dB.

To analyze the signal in the joint time-frequency domain, we applied the novel adaptive GMCLS algorithm with a maximum order of $M = 8$, an averaging length $L_{av} = 8$, an asymptotic memory length $L_n = 118$, and a minimum forgetting factor of $\lambda_{\text{min}} = 0.96$. The result is shown in Fig. 2 including the estimated model order (dashed line) and the variable forgetting factor (solid line) in the bottom plot. Again, a short segment of the process under consideration was used to initialize the adaptive RLS algorithm. This segment has been omitted in Fig. 2.

Initially, a forgetting factor around $\lambda[n] = 0.993$ and a model order of two capture the nonstationary quadratic chirp in white Gaussian noise. Notice that in comparison to the previous simulation example, this example generally exhibits forgetting factors that are less in variation and closer to unity. This is mainly due to the higher SNR of 20 dB. Between $n = 257$ and $n = 448$, the Gaussian component is correspondingly reflected in the bell-shaped trough of the forgetting factor with a minimum of $\lambda[n] = 0.982$. However, no increase in estimated model
order seems to be necessary in this situation. At time index \( n = 513 \), the linear chirp is switched on and thus introduces a high degree of nonstationarity. This causes the forgetting factor to immediately drop down to \( \lambda[n] = 0.984 \) and the model order estimate to increase by one after a short delay. To account for one sinusoidal and one linear frequency component, the variable forgetting factor converges towards a value of \( \lambda[n] = 0.990 \) while the model order remains fixed at three.

7. CONCLUSIONS

In this paper, an adaptive AR method to estimate time-varying spectra of nonstationary signals is proposed. Its advantage is the adaptability with respect to forgetting factor and to model order. The MPLS criterion allows for estimation of the time-varying model order and the adaptive forgetting factor ensures good tracking performance of nonstationary signals and processes. We implemented this approach using an efficient GMCLS lattice type algorithm. The simulation results confirm that this joint estimation of forgetting factor and model order shows good adaptability to time-varying signals. Moreover, the resulting time-frequency analysis exhibits high time and frequency resolution without any interference. Basically, this remains true even for relatively short segments of data and in high noise environments.

REFERENCES