MAXIMUM-LIKELIHOOD AUTOREGRESSIVE ESTIMATION ON INCOMPLETE SPECTRA

Luis Weruaga

VISKOM, Austrian Academy of Sciences
Donau-City Strasse, 1090 Vienna, AUSTRIA
weruaga@ieee.org

Marián Képesi

SPSC Lab, Graz University of Technology
Inffeldgasse 12, 8010 Graz, AUSTRIA
kepesi@tugraz.at

ABSTRACT

Frequency-selective autoregressive (AR) estimation is arousing increasing interest. We propose herein a new method to estimate the AR model from a reduced set of spectral samples. The proposed method is founded on the maximum likelihood criterion over the logarithmic spectral residue, and it is implemented efficiently with a multivariate Newton–Raphson algorithm. Results over deterministic and stochastic scenarios show its excellent performance.

Index Terms— Autoregressive model, frequency domain, incomplete spectrum, maximum likelihood.

1. INTRODUCTION

Autoregressive (AR) modeling is a popular parametric method for spectral analysis and prediction of stationary processes [1]. An AR model is described by the difference equation

$$x[n] = \sum_{k=1}^{P} a_k\, x[n-k] + e[n] \tag{1}$$

where $e[n]$ is usually considered a random stationary process, a_k are the AR coefficients (ARC) and P is the model order.

AR modeling exclusively in the spectral domain [2] is not a common practice in time series analysis. However, in many scenarios only the spectrum (or a subband) is available. In most of the cases, the methodology has been the conversion of the spectral samples to the time domain so that the efficient autocorrelation method can be applied. As alternative to that methodology, recent works [3] suggest more complex procedures. However, all these approaches are not appropriate when dealing with missing or corrupted spectral samples, a problem that has not aroused yet enough interest.

This paper presents a novel method to estimate an AR model directly from spectral samples associated to scattered regions of the frequency domain. The proposed frequency selective estimation method is based on a risk functional derived following a maximum likelihood criterion over the logarithmic residue. This solution is achieved by a multivariate Newton–Raphson algorithm, which here results in a cost-efficient implementation. The proposed method is validated on stochastic and deterministic scenarios with simulations.

2. ANTECEDENT

The transfer function of the AR model in terms of its ARC corresponds to the well-known expression

$$H(e^{j\omega}) = \left(1 - \sum_{k=1}^{P} a_k\, e^{-j\omega k}\right)^{-1} . \tag{2}$$

Let $\hat{X}(e^{j\omega})$ be the Fourier transform of a realization of the AR process $\hat{x}[n] = w[n]\, x[n]$, where $w[n]$ is the analysis window. Classically [2], the AR estimation from the spectral samples $\hat{X}(e^{j\omega})$ corresponds to the minimization of functional [1]

$$\mathcal{J}_{\text{lin}} \triangleq \int_{-\pi}^{\pi} \Lambda(\omega)\, \big|\varepsilon(\omega)\big|^2\, d\omega \tag{3}$$

where $\varepsilon(\omega)$ is the spectral residue

$$\varepsilon(\omega) \triangleq \frac{\hat{X}(e^{j\omega})}{H(e^{j\omega})} . \tag{4}$$

The positive-valued spectral mask $\Lambda(\omega)$ is introduced here as tentative way to weight/disregard untrustful spectral samples (for instance, because corrupted by an additive interference).

The minimum of the functional $\mathcal{J}_{\text{lin}}$ (3) corresponds to the zero of its gradient [2]

$$\frac{\partial \mathcal{J}_{\text{lin}}}{\partial a_k} = -\Re \int_{\langle\Lambda\rangle} \big|\hat{X}(\omega)\big|^2 \frac{e^{-j\omega k}}{H(e^{j\omega})^*}\, d\omega \tag{5}$$

where * denotes complex conjugate and $\Re\, z = \frac{1}{2}(z + z^*)$. A well-known property of this functional is its quadratic nature, as the Hessian matrix reveals

$$\frac{\partial^2 \mathcal{J}_{\text{lin}}}{\partial a_\ell \partial a_k} = \Re \int_{\langle\Lambda\rangle} e^{j\omega(\ell-k)} \big|\hat{X}(e^{j\omega})\big|^2\, d\omega\, . \tag{6}$$

Therefore, with the gradient vector and Hessian matrix this quadratic problem can be solved in one step as follows

$$\mathbf{a}_{\text{lin}} = -\big(\nabla^2_{\mathbf{a}} \mathcal{J}_{\text{lin}}\big)^{-1} \nabla_{\mathbf{a}} \mathcal{J}_{\text{lin}} \Big|_{\mathbf{a}=\mathbf{0}} \tag{7}$$

[1] This scenario does not take into account the spectral leakage or bias caused by the windowing, an effect addressed carefully in [2].

[2] For the sake of simplicity in the notation, thereafter $\int_{\langle\Lambda\rangle} F(\omega)\, d\omega = \int_{-\pi}^{\pi} \Lambda(\omega)\, F(\omega)\, d\omega$, where $F(\omega)$ is a generic function of variable ω.

where $\mathbf{a} = [a_1, \ldots, a_P]$. Note that (7) is equivalent to the solution obtained by the autocorrelation (Yule–Walker) method. [3] Thus, since the Hessian (6) is a Toeplitz matrix, the solution (7) can be efficiently obtained.

However, a major problem arises when analyzing the role of the proposed spectral mask within the functional (3): if $\Lambda(\omega')$ is set to zero for a given frequency ω' as mean to disregard the corresponding spectral sample $\hat{X}(e^{j\omega'})$, the resulting effect is equivalent to setting that sample zero, i.e., $\hat{X}(e^{j\omega'}) \equiv 0$. Given that absence of spectral information does not imply the spectral energy to be equal to zero, the solution provided by the minimization of $\mathcal{J}_{\text{lin}}$ in case of frequency selective scenarios is clearly not appropriate.

3. MAXIMUM-LIKELIHOOD FUNCTIONAL

The maximum-likelihood (ML) dependence between $H(e^{j\omega})$ and the samples $\hat{X}(e^{j\omega})$ is given by

$$\mathcal{L} \triangleq -\int_{\langle\Lambda\rangle} \log \mathrm{P}\Big[f\big(\varepsilon(\omega)\big)\Big]\, d\omega \tag{8}$$

where $f(x)$ is a continuous monotonously increasing function that maps the spectral residue $\varepsilon(\omega)$ into another scale, and $\mathrm{P}[\,]$ denotes probability. In consequence, the statistics of the residue $f\big(\varepsilon(\omega)\big)$ define the ML estimator, that is, the ML estimation does not depend on the choice of the mapping $f(\varepsilon)$ as long as (8) is observed [6].

In case of the linear mapping $f(x) = x$, the transformed residue corresponds to $\varepsilon(\omega)$, which is Gaussian (both real and complex parts). According to (8), the ML functional should be thus built with a squared loss (as $\mathcal{J}_{\text{lin}}$ (3) is in fact). However, since the presence of zero values in the spectral mask, $\Lambda(\omega') = 0$ yields $\varepsilon(\omega') \equiv 0$, the equivalent residue is not strictly Gaussian, and thus that functional does not result in the ML estimation with incomplete spectra.

Let us consider the logarithmic transformation

$$f(\varepsilon) \triangleq \log \frac{|\varepsilon|^2}{\eta} \tag{9}$$

where η is an additional parameter that accounts for the mean energy of the residue. In order to build the ML functional (8) based on the new logarithmic residue, its probability density function (pdf) is required. This pdf results in [4]

$$\mathrm{P}\big[x\big] = \exp\left(x - \exp x\right). \tag{10}$$

This pdf is depicted in Fig. 1. By replacing (10) into (8), the ML functional results in

$$\mathcal{L}_{\log} \triangleq \int_{\langle\Lambda\rangle} \frac{|\varepsilon(\omega)|^2}{\eta} - \log \frac{|\varepsilon(\omega)|^2}{\eta}\, d\omega \tag{11}$$

[3] The solution deviates from that of Yule–Walker method due to the mentioned spectral leakage caused by the windowing.

[4] The squared residue $|\varepsilon(\omega)|^2/\eta$ is a chi-square random variable with two degrees of freedom and variance one, whose pdf is exponential, $P[x] = e^{-x}u(x)$. The logarithmic transformation (9) gives rise to the pdf in (10).

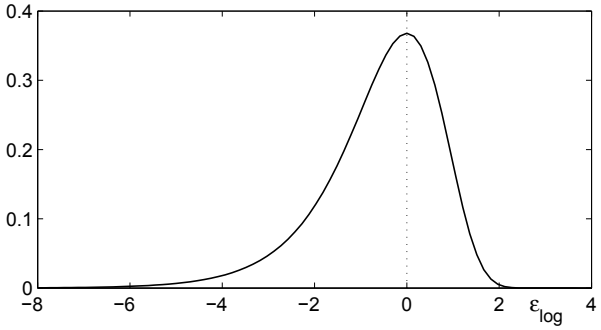


Fig. 1. Probability density function of the log residue.

which after simplifications turns out equivalent to

$$\mathcal{L}_{\log} \equiv \int_{\langle\Lambda\rangle} \frac{\big|\hat{X}(e^{j\omega})\big|^2}{\eta\big|H(e^{j\omega})\big|^2} + \log\Big(\eta\big|H(e^{j\omega})\big|^2\Big)\, d\omega\,. \tag{12}$$

This functional (12) is equal to the Whittle likelihood function [4]. The work [5] has recently proposed that likelihood for similar purposes. Unfortunately, that work does not provide the arguments on the choice of the Whittle likelihood versus the ML linear residue (3) for a frequency selective scenario.

In case of missing spectral samples, setting $\Lambda(\omega') = 0$ into the new functional (11) is also equivalent to setting the term in the integral to zero, and thus

$$\epsilon = \log \epsilon \tag{13}$$

where $\epsilon = |\varepsilon(\omega')|^2/\eta > 0$. However, equation (13) has no solution. Thus, a zero in the spectral mask does not imply constraining the value of the residue $\varepsilon(\omega')$ nor of the spectral sample $\hat{X}(e^{j\omega'})$. We can then conclude that, contrary to the linear residual, the use of the spectral mask in (11) is suited to represent missing spectral samples, or even to weight the importance of each spectral sample in the final estimation. This fact will turn out essential for delivering accurate frequency selective AR estimates.

3.1. Minimization Mechanism

It is clear that the proposed functional (11), or (12), is not quadratic with respect to a_k. Furthermore, the energy parameter η is an additional degree of freedom in the problem. The following iterative multivariate Newton–Raphson algorithm is proposed as mechanism to reach the minimum of $\mathcal{L}_{\log}$ [5]

$$\mathbf{a}^{(\xi+1)} = \mathbf{a}^{(\xi)} - \left(\nabla^2_{\mathbf{a}}\mathcal{L}_{\log}\right)^{-1} \nabla_{\mathbf{a}}\mathcal{L}_{\log}\Big|^{(\xi)} \tag{14a}$$

$$\eta^{(\xi+1)} = \eta^{(\xi)} - \left(\nabla^2_{\eta}\mathcal{L}_{\log}\right)^{-1} \nabla_{\eta}\mathcal{L}_{\log}\Big|^{(\xi)} \tag{14b}$$

[5] In fact, the solution of the quadratic functional $\mathcal{J}_{\text{lin}}$, given by (7), corresponds to a multivariate Newton–Raphson mechanism as well. In that case, one single iteration suffices to reach the solution.

where ξ is iteration of the algorithm, and the numerator and denominator of the update correspond to the gradient vector and Hessian matrix respectively. Note that the ARC $\mathbf{a}$ and the energy η are updated independently from each other. Although a global multivariate Newton–Raphson algorithm could be also an option, as proven later the proposed update (14) turns out to have interesting computational advantages. The gradient vector results in

$$\frac{\partial \mathcal{L}_{\log}}{\partial a_k} = -\Re \int_{\langle \Lambda \rangle} \left(|\varepsilon(\omega)|^2 - \eta \right) e^{-j\omega k} H(e^{j\omega})\, d\omega \qquad (15)$$

and the Hessian matrix in [6]

$$\frac{\partial^2 \mathcal{L}_{\log}}{\partial a_\ell \partial a_k} = \frac{\partial^2 \mathcal{J}_{\text{lin}}}{\partial a_\ell \partial a_k} + \eta\, \Re \int_{\langle \Lambda \rangle} e^{-j\omega(\ell+k)} H(e^{j\omega})^2\, d\omega\,. \quad (16)$$

The update of the energy (14b) turns out

$$\eta^{(\xi+1)} = \eta^{(\xi)} \left(1 + \frac{\int_{\langle \Lambda \rangle} |\varepsilon(\omega)|^2 - \eta^{(\xi)}\, d\omega}{\int_{\langle \Lambda \rangle} 2|\varepsilon(\omega)|^2 - \eta^{(\xi)}\, d\omega} \right). \qquad (17)$$

3.2. Efficient Implementation

The Hessian (16) has been written accordingly so as to reveal the differences with that of the linear residue minimization. By inspecting the additional term, we can state

$$\int_{\langle \Lambda \rangle} e^{-j\omega(k+\ell)} H(e^{j\omega})^2\, d\omega \equiv \left. h[n] * h[n] * \lambda[n] \right|_{n=-k-\ell} \qquad (18)$$

where $*$ denotes convolution, and $\lambda[n]$ and $h[n]$ are the time-domain counterpart of $\Lambda(\omega)$ and $H(e^{j\omega})$ respectively. Since $h[n]$ corresponds to a real causal system, its value at negative time instants, $n = -(k+\ell)$, is negligible, and thus so is (18). The spectral mask $\Lambda(\omega)$ may contradict the previous statement since $\lambda[n]$ may be non-causal, but in any case the term (18) is clearly much lower than the correlation-based term (6). Thus, we can simplify the Hessian matrix as

$$\nabla^2_{\mathbf{a}} \mathcal{L}_{\log} \simeq \nabla^2_{\mathbf{a}} \mathcal{J}_{\text{lin}}\,. \qquad (19)$$

This simplification brings outstanding computational advantages with respect to the initial mechanism: since the new Hessian matrix (19) is a Toeplitz matrix, its inverse can be very efficiently computed.

The initialization of the algorithm plays an important role for guaranteing functional convexity and thus converging to a minimum. Let us not forget that although the applied Hessian (19) is positive (semi)definite, the convergence depends on the actual Hessian and especially on the update of the energy term η. Equations (16) and (17) clearly point out to the energy term η as responsible to guarantee convexity in the problem. In fact, $\eta \simeq 0$ clearly makes the problem convex; on the other hand very large values may compromise the stability. We observed heuristically that the following initialization

$$\eta^{(0)} = \int_{\langle \Lambda \rangle} \left| \hat{X}(e^{j\omega}) \right|^2 d\omega \Big/ \int_{\langle \Lambda \rangle} d\omega \qquad (20)$$

yields fast convergence. Given the convexity of Hessian (16), the initialization of a_k was observed not being important. We simply recommend initial zero-valued ARC.

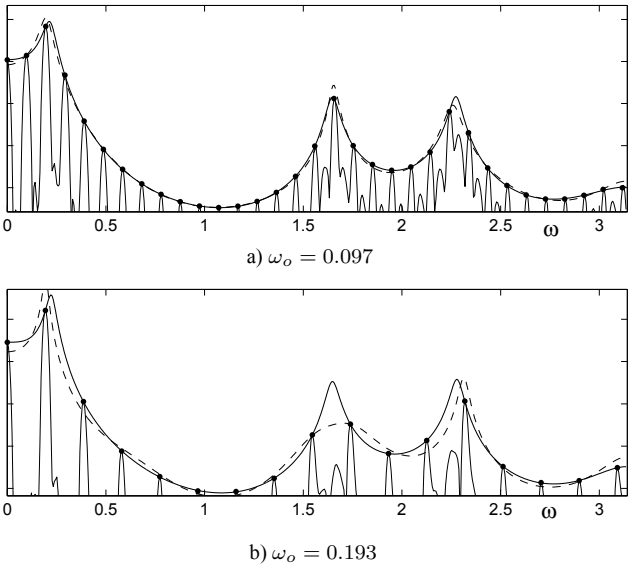
[6] For the sake of clarity, both gradient (15) and Hessian (16) have been multiplied by η, which clearly has no consequences in the update (14a).

4. SIMULATION RESULTS

The first example addresses the AR estimation on a synthetic harmonic deterministic signal. The signal was generated as the output of an 8-order all-pole filter excited with a train of periodic pulses. A short segment (384 samples) of the signal was Hamming windowed, zero-padded to $N = 512$ samples, Digital Fourier transformed, and thereof the energy of the harmonics obtained. This last step is equivalent to setting

$$\Lambda(\omega) = \sum_k \delta(\omega - k\omega_o) \qquad (21)$$

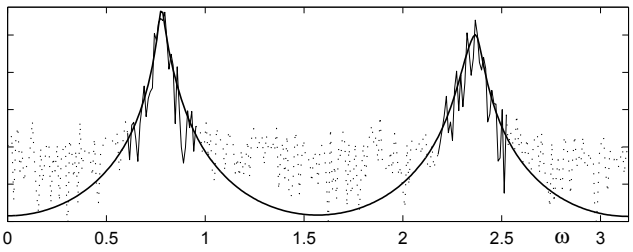
where $\delta(\omega)$ is the Dirac delta, and ω_o is the fundamental frequency (estimated as shown in [7]). This scenario corresponds clearly to frequency selective estimation, in which the spectral energy only at frequencies multiple of the fundamental $\omega = k\omega_o$ is valuable. The analysis order was set to $P = 14$, as is commonly used in speech coding. Fig. 2 contains the results obtained with the proposed $\mathcal{L}_{\log}$ minimization and the classical approach ($\mathcal{J}_{\text{lin}}$ minimization) over two



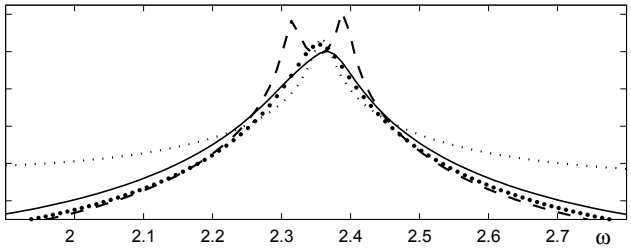
a) $\omega_o = 0.097$

b) $\omega_o = 0.193$

Fig. 2. Estimation on synthetic speech: spectrum (fine line), envelope of $\mathcal{L}_{\log}$ (solid) and $\mathcal{J}_{\text{lin}}$ (dashed) minimization. Y-axis in log scale. At 8 kHz, it sounds as the /i/ utterance.



a)



b) Theoretical $\cdots$, $\mathcal{L}_{\log}$ —, $\mathcal{J}_{\lin}$ - - -, and $\Lambda(\omega) = 1$ $\cdots$.

Fig. 3. Frequency selective estimation: a) valid spectral samples (thin) and noise-corrupted ones (dotted); b) comparative by zooming on the second resonance. Y-axis in log scale.

different fundamental frequency cases. In both situations, the classical method is unable to deliver an appropriate spectral envelope (dashed). In case of wide-spaced harmonics, that solution is largely inaccurate in terms of frequency and energy of the resonance. On the contrary, the proposed $\mathcal{L}_{\log}$ minimization delivers an accurate "all-pole interpolation" (solid) of the spectral harmonic energy. Furthermore, the result on the wide-spaced harmonic case is outstanding: frequency and level of each resonance are precisely identified, in spite of the critically undersampled spectral envelope.

The second example deals with the estimation of a four-order AR process corrupted with additive noise. The spectral mask $\Lambda(\omega)$ plays an important role since it needs to be set to one for clean spectral values and to zero in case of noisy samples. The following simple adaptive procedure is used

$$\Lambda^{(\xi)}(\omega) \triangleq \begin{cases} 1, & \text{if } N(e^{j\omega}) < \eta^{(\xi)} \left|H^{(\xi)}(e^{j\omega})\right|^2 \\ 0, & \text{otherwise} \end{cases} . \quad (22)$$

The previous procedure requires the knowledge of the power spectral density $N(e^{j\omega})$ of the additive noise, which is compared against the current estimated spectral envelope. Fig. 3 shows the results of the method on this scenario. The solid line in Fig. 3.a corresponds to the spectral envelope resulting from the proposed $\mathcal{L}_{\log}$ minimization. In the same figure, the spectral samples that turn out noise-free and the noise-corrupted ones according to (22) are shown in thin-solid and dotted line respectively.

The interpolation capability of the proposed method is compared with other methods in Fig. 3.b. The $\mathcal{J}_{\lin}$ minimization (dashed) yields two poles (two peaks of the spectral envelope) apart from each other and with a too high radius (closer to unit circle) than the actual ones. This inaccurate result is due to the noise-corrupted spectral regions which are actually treated in the functional as zero-valued spectral samples. By considering all samples as valid, i.e. $\Lambda(\omega) = 1$, the resulting spectral envelope (dotted) yields a one-pole resonance (noticeable in the narrow bandwidth of the resonance) instead of the actual two poles. On the contrary, the frequency selective estimation resulting from the proposed $\mathcal{L}_{\log}$ minimization is close to the theoretical envelope. The number of iterations in the $\mathcal{L}_{\log}$ minimization for this example is twenty.

5. CONCLUSIONS

The use of the logarithmic spectral residue is essential for achieving maximum-likelihood frequency-selective autoregressive (AR) estimation. This paper has proven that fact and proposed an iterative Newton mechanism that converges to the solution. Since the Hessian matrix is Toeplitz, the algorithm can be implemented efficiently as to demand at each iteration as low computational load as in the Yule–Walker mechanism. The method is validated on two ill-posed scenarios in which a reduced set of the spectral samples is available: the AR estimation from the deterministic energy of harmonics, and from a noise-corrupted stochastic process. The comparison with the classical spectral AR analysis reveals the excellent estimation accuracy of the proposed method.

6. REFERENCES

[1] B. Rust and D. Donelly, "The fast Fourier transform for experimentalists, part IV: Autoregressive spectral analysis," *Comp. Sci. & Eng.*, vol. 7, pp. 85–90, 2005.

[2] R. Pintelon and J. Schoukens, "Time series analysis in the frequency domain," *IEEE Trans. Signal Processing*, vol. 47, pp. 206–210, 1999.

[3] P. M. T. Broersen and S. de Vaele, "Time series analysis in a frequency subband," *IEEE Trans. Instrum. Meas.*, vol. 52, pp. 1054–1060, 2003.

[4] P. Whittle, "Gaussian estimation in stationary time series," *Bull. Intl. Stat. Instit.*, 39, pp. 105–130, 1961.

[5] J. Gillberg and F. Gustafsson, "Frequency-domain continuous-time AR modeling using non-uniformly sampled measurem.," *IEEE ICASSP 2005*, pp. 105–108.

[6] B. Schölkopf and A. J. Smola. *Learning with Kernels*. MIT Press, 2002.

[7] M. Képesi and L. Weruaga, "Adaptive chirp-based time–frequency analysis of speech signals," *Speech Commun.*, vol. 48, pp. 474–492, 2006.