WEIGHTED MAXIMUM LIKELIHOOD AUTOREGRESSIVE AND MOVING AVERAGE SPECTRUM MODELING

Roland BADEAU, Bertrand DAVID

Télécom Paris (ENST) / CNRS LTCI, TSI Department
roland.badeau@enst.fr

ABSTRACT

We propose new algorithms for estimating autoregressive (AR), moving average (MA), and ARMA models in the spectral domain. These algorithms are derived from a maximum likelihood approach, where spectral weights are introduced in order to selectively enhance the accuracy on a predefined set of frequencies, while ignoring the other ones. This is of particular interest for modeling the spectral envelope of harmonic signals, whose spectrum only contains a discrete set of relevant coefficients. In the context of speech processing, our simulation results show that the proposed method provides a more accurate ARMA modeling of nasal vowels than the Durbin method.

Index Terms— Autoregressive moving average processes, Spectral domain analysis, Maximum likelihood estimation.

1. INTRODUCTION

Parametric spectrum modeling is a prominent tool in time series analysis [1]. The popular autoregressive moving average models are particularly useful for predicting the future values of a time series from its past samples. This is why the estimation task is naturally performed in the time domain. However, performing this task in the spectral domain may be relevant to process signals whose spectrum only partly satisfies the parametric model. For instance, a given region of the spectrum can be corrupted by an extraneous signal. In the case of harmonic spectra, the parametric model is appropriate for representing the envelope of the observed spectrum, which only contains a discrete set of relevant coefficients. The problem of partial spectrum modeling was first addressed by A. El-Jaroudi and J. Makhoul [2], who introduced the Discrete All-Pole (DAP) method for estimating AR models. The DAP method included a frequency-dependent weighting of the whole frequency range, so that the spectral accuracy could be enhanced in certain frequency regions compared to others. A similar approach was later presented in [3]. In [4], the True Envelope Linear Prediction Coding (TELPC) method was proposed for estimating the envelope of harmonic speech signals. Finally, the problem of ARMA modeling in a frequency subband was addressed in [5]. In this paper, we generalize the DAP approach to the estimation of MA and ARMA models. Our estimator involves the iterative maximization of a weighted maximum likelihood function, which is performed by using a new algorithm, whose convergence is proved in the case of AR modeling.

The research leading to this paper was supported by the French GIP ANR under contract ANR-06-JCJC-0027-01, Décompositions en Éléments Sonores et Applications Musicales - DESAM, and by the Groupe des Écoles des Télécommunications (GET), under contract TAM-TAM. We also acknowledge Gaël Richard, Professor at ENST, who kindly lent his voice for the signal analyzed in section 4, and Jean-Louis Durrieu, Ph.D. student at ENST, for its feedback about this work.

2. PRINCIPLE

Given an observed power spectrum $Y(\nu)$, where the normalized frequency ν belongs to a subset of $[0,1)$, we propose a Weighted Maximum Likelihood (WML) approach for estimating a general ARMA model (which includes AR and MA models). The Power Spectral Density (PSD) of an ARMA process of order (P,Q) is of the form

$$S(\nu) = \sigma^2 \frac{\left|\sum_{q=0}^{Q} b_q e^{-i2\pi\nu q}\right|^2}{\left|\sum_{p=0}^{P} a_p e^{-i2\pi\nu p}\right|^2} = \sigma^2 \frac{B(\nu)}{A(\nu)}, \tag{1}$$

where $\sigma > 0$, $a_0 = 1$, and $b_0 = 1$. As usually assumed in the literature, we suppose that $\forall \nu$, $Y(\nu)$ follows an exponential distribution of parameter $S(\nu)$: its probability density function p_ν satisfies $\log(p_\nu(Y(\nu))) = \log\left(\frac{1}{S(\nu)}\right) - \frac{Y(\nu)}{S(\nu)}$. Assuming that the values $Y(\nu)$ for all frequencies are independent, we define the weighted log-likelihood function of the whole observed spectrum as a function of σ^2 and of the vectors $a = [a_0, \ldots, a_P]^T$ and $b = [b_0, \ldots, b_Q]^T$:

$$L(\sigma^2, a, b) = \int_0^1 \left(1 + \log\left(\frac{Y(\nu)}{S(\nu)}\right) - \frac{Y(\nu)}{S(\nu)}\right) d\mu(\nu), \tag{2}$$

where we added the constant $\int_0^1 (1 + \log(Y(\nu)))d\mu(\nu)$, so that this expression is similar to that of the Itakura-Saito distance [2]. Spectral weights are introduced via measure μ, which can be any positive measure such that $\int_0^1 \, d\mu(\nu) = 1$ (for instance, μ can typically be a discrete measure whose support uniformly samples the interval $[0,1)$). Substituting equation (1) into equation (2), and maximizing the log-likelihood with respect to parameter σ, yields the estimate

$$(\sigma^\star)^2 = \int_0^1 \frac{Y(\nu)A(\nu)}{B(\nu)} d\mu(\nu). \tag{3}$$

Then substituting equations (1) and (3) into equation (2) yields

$$L(a,b) = \int_0^1 \log\left(\frac{Y(\nu)A(\nu)}{B(\nu)}\right) d\mu(\nu) - \log\left(\int_0^1 \frac{Y(\nu)A(\nu)}{B(\nu)} d\mu(\nu)\right). \tag{4}$$

The concavity of function log shows that L is non-positive, and attains the value 0 if and only if $\frac{Y(\nu)A(\nu)}{B(\nu)}$ is constant in the support of measure μ. Thus maximizing the log-likelihood is equivalent to maximizing the flatness of the ratio $\frac{Y(\nu)A(\nu)}{B(\nu)}$. Since L is unchanged by inverting and conjugating any roots of the polynomials $A(z) = \sum_{p=0}^{P} a_p z^p$ and $B(z) = \sum_{q=0}^{Q} b_q z^q$, we can look for vectors a and b which belong to the set $\mathscr{S}$ of vectors of first coefficient 1, and such that the polynomial roots are strictly outside the unit circle[1].

[1] For simplicity, the notation $\mathscr{S}$ does not account for the vector dimension, which can be P or Q when appropriate. When viewed as the coefficients of

3. MAXIMIZATION OF THE LOG-LIKELIHOOD

3.1. Autoregressive models

In this section, we investigate the case of AR models, for which function L can be straightforwardly optimized[2]. For all vectors $a \in \mathscr{S}$, we parameterize the continuous spectrum in the form

$$A(\nu) = \left| \sum_{p=0}^{P} a_p e^{-i2\pi\nu p} \right|^2 = a^T T(\nu) a, \tag{5}$$

where $\forall \nu \in [0,1)$, $T(\nu)$ is the Toeplitz matrix whose coefficient of index (p,q) is $\cos(2\pi\nu(p-q))$.

Proposition 3.1 (Characterization of the WML AR estimator)**.** *If the $\mathscr{C}^1$ function L admits a critical point $a^\star$ in $\mathscr{S}$, then this point is the global maximum of L, characterized by the equation*

$$R(a^\star, \mu) = R(a^\star, \mu_{YA^\star}) \tag{6}$$

which involves the $(P+1)\times(P+1)$ Toeplitz matrices

$$R(a,\mu) = \int_0^1 \frac{T(\nu)}{A(\nu)}\, d\mu(\nu) \tag{7}$$

$$R(a,\mu_{YA}) = \frac{\int_0^1 Y(\nu)\, T(\nu)\, d\mu(\nu)}{\int_0^1 Y(\nu) A(\nu)\, d\mu(\nu)}. \tag{8}$$

The proof of proposition 3.1 can be found in Appendix A, and relies on the gradient of L: $\frac{1}{2}\nabla L(a) = (R(a,\mu) - R(a,\mu_{YA}))\, a$. In proposition 3.1, notation $R(a,\mu_{YA})$ is to be understood as follows: for any measure μ on the interval $[0,1)$, and any positive function W of period 1, we define the measure $d\mu_W(\nu) = \frac{W(\nu)}{\int_0^1 W(\xi)\, d\mu(\xi)} d\mu(\nu)$. In the case $W = YA$, the notation $R(a,\mu_{YA})$ is compatible with $R(a,\mu)$. Similar notations will be used throughout the paper.

In practice, L always admits a global maximum $a^\star$ in $\mathscr{S}$, except in the singular case where the polynomial $A^\star(z)$ has at least one root on the unit circle (see [2] for more details). Below, we propose an iterative algorithm which solves equation (6) in $\mathscr{S}$. Since $\nabla L(a^\star) = 0$, it can be noted that $a^\star$ is a fixed point in $\mathscr{S}$ of the function

$$\phi(a) = R(a,\mu_{YA})^{-1} R(a,\mu)\, a. \tag{9}$$

Then a first order expansion yields

$$\begin{aligned} L(\phi(a)) - L(a) &= (\phi(a)-a)^T \nabla L(a) + o(\|\phi(a)-a\|) \\ &= \tfrac{1}{2} \nabla L(a)^T R(a,\mu_{YA})^{-1} \nabla L(a) + o(\|\phi(a)-a\|) \\ &\geq 0 \text{ in a neighborhood of } a^\star. \end{aligned} \tag{10}$$

This suggests an ascent method for maximizing function L, which consists in recursively applying function ϕ to an initial point in $\mathscr{S}$. In practice however, the polynomial defined by vector $\phi(a)$ may have some roots inside the unit circle. Thus $\phi(a)$ should be remapped into $\mathscr{S}$ by replacing those roots by their inverse conjugate. This remapping, which is denoted $\mathbb{P}_{\mathscr{S}}$ below, does not modify the value of function L. Moreover, we show in Appendix B that forcing vector a in $\mathscr{S}$ dramatically improves the convergence rate of the recursion. Finally, our algorithm consists of the following steps:

- Initialization : $a = [1, 0 \ldots 0]^T$
- Repeat until convergence: $a \leftarrow \mathbb{P}_{\mathscr{S}}\left(R(a,\mu_{YA})^{-1} R(a,\mu)\, a\right)$

The convergence of this algorithm is analyzed in the following proposition, which is proved in Appendix B.

Proposition 3.2 (Convergence of the algorithm for AR estimation)**.**

1. *A global maximum $a^\star$ of function L in $\mathscr{S}$ is a locally stable equilibrium point of the discrete dynamical system formed by the recursion $a \leftarrow \mathbb{P}_{\mathscr{S}}(\phi(a))$.*
2. *If μ is the Lebesgue measure, the recursion converges in one iteration only.*

In practice, we observed that this algorithm globally converges to the global maximum (whatever the initial point is). Moreover, if μ is a measure close enough to the Lebesgue measure[3], the algorithm converges in a few iterations, as shown in Appendix B.

3.2. Moving average models

Proposition 3.3 (Characterization of the WML MA estimator)**.** *If L admits a critical point $b^\star$ in $\mathscr{S}$, then this point satisfies the equation*

$$R(b^\star, \mu) = R(b^\star, \mu_{Y/B^\star}) \tag{11}$$

which involves the $(Q+1)\times(Q+1)$ Toeplitz matrices

$$R(b,\mu) = \int_0^1 \frac{T(\nu)}{B(\nu)}\, d\mu(\nu) \tag{12}$$

$$R(b,\mu_{Y/B}) = \frac{\int_0^1 \frac{Y(\nu)\, T(\nu)}{B(\nu)^2}\, d\mu(\nu)}{\int_0^1 \frac{Y(\nu)}{B(\nu)}\, d\mu(\nu)}. \tag{13}$$

Proposition 3.3 can be proved in the same way as proposition 3.1. Note however that this proposition does not guarantee that the solutions of equation (11) are global maxima. In order to maximize function L, we propose the following iterative algorithm:

- Initialization : $b = [1, 0 \ldots 0]^T$
- Repeat until convergence: $b \leftarrow \mathbb{P}_{\mathscr{S}}\left(R(b,\mu)^{-1} R(b,\mu_{Y/B})\, b\right)$

As in section 3.1, it can be verified that this recursion constitutes an ascent method for maximizing function L. In practice, we observed that this algorithm globally converges to the global maximum of function L; however our proof of convergence, which could not be included in this paper, relies on additional assumptions regarding the observed spectrum. Again, if μ is a measure close enough to the Lebesgue measure, the convergence rate is enhanced.

3.3. Autoregressive moving average models

We propose the following iterative algorithm for maximizing function L, which consists in interlacing the two updates introduced above:

- Initialization : $a = b = [1, 0 \ldots 0]^T$
- Repeat until convergence:

Finite Impulse Response (FIR) filters, the vectors in $\mathscr{S}$ correspond to minimum phase filters.

[2]Note that the proposed WML estimator is equivalent to the DAP method in the case of AR modeling. However, our new algorithm for computing the optimal solution has an enhanced convergence rate, as proved in Appendix B.

[3]For instance, we say that a discrete measure, which uniformly samples the interval $[0,1)$ with constant weights, is "close to" the Lebesgue measure, in the sense that the integral of any continuous function with respect to this discrete measure approximates the integral of the same function with respect to the Lebesgue measure, when the number of samples is high enough.

– Repeat: $a \leftarrow \mathbb{P}_{\mathscr{S}}\left(R(a,\mu_{YA/B})^{-1}R(a,\mu)a\right)$

– Repeat: $b \leftarrow \mathbb{P}_{\mathscr{S}}\left(R(b,\mu)^{-1}R(b,\mu_{YA/B})\,b\right)$

where $R(a,\mu)$ and $R(b,\mu)$ are defined in equations (7) and (12), and

$$R(a,\mu_{YA/B})) = \frac{\int_0^1 \frac{Y(\nu)\ T(\nu)}{B(\nu)}\ d\mu(\nu)}{\int_0^1 \frac{Y(\nu)A(\nu)}{B(\nu)}\ d\mu(\nu)}$$

$$R(b,\mu_{YA/B}) = \frac{\int_0^1 \frac{Y(\nu)A(\nu)\ T(\nu)}{B(\nu)^2}\ d\mu(\nu)}{\int_0^1 \frac{Y(\nu)A(\nu)}{B(\nu)}\,d\mu(\nu)}.$$

The convergence was not proved, but observed in practical cases[4].

4. SIMULATION RESULTS

We address the problem of estimating the spectral envelope of a voiced speech signal. In speech processing, AR modeling is a prominent tool, since it constitutes a simplified model of the speech production system. However, it is not well adapted to fit the spectra of nasal sounds, which contain zeros due to the coupling between the vocal and nasal cavities. Below we focus on a French nasal vowel, which is denoted /ɛ̃/ in the International Phonetic Alphabet. This vowel was pronounced at a fundamental frequency $F_0 = 120$ Hz, and recorded at a sampling frequency of 8000 Hz. The spectrum of this original sound was then obtained by successively extracting a signal of length 60 ms, applying a Hann window, and computing a Digital Fourier Transform (DFT), zero-padded to $N = 1024$ samples. Then an ARMA model was estimated by applying the proposed WML method with weights equal to 1 for the DFT samples corresponding to multiples of F_0, and 0 elsewhere[5]. A good fit was obtained with $P = 8$ and $Q = 7$. Finally, the vowel was resynthesized at $F_0 = 100$ Hz and 300 Hz respectively, by filtering a train of periodic pulses, and the spectra of those two synthetic signals were computed in the same way as that of the original one. The first experiment illustrated in Figure 1-a shows the inadequacy of AR models for modeling nasal vowels. The spectrum of the synthetic signal at $F_0 = 100$ Hz, represented by the solid gray line, is modeled by an AR model of order $P = 15$, represented by the solid black line[6]. The AR model was estimated with the DAP method [2]. The formants are well represented by this AR model; however the model does not fit the zeros present at 1250, 2250, and 3900 Hz. The second experiment illustrated in Figure 1-b confirms that conversely, an ARMA model of orders $P = 8$ and $Q = 7$ can efficiently represent the spectral envelope. The dashed black line represents the result obtained with the classical Durbin method [6], and the solid black line represents the result obtained with the WML method. Both methods provide an accurate estimate of the original ARMA model. In the last experiment illustrated in Figure 1-c, both methods are applied to the synthetic signal at $F_0 = 300$ Hz, with $P = 8$ and $Q = 7$. The WML method still correctly estimates the positions of the zeros, even though one of them falls between two harmonics, whereas Durbin's method only provides a smooth approximation of the spectral envelope.

[4] Note that within one iteration of this algorithm, we perform several updates of a and b, because we observed that if only one update was performed for each vector, the algorithm could alternate between two sub-optimal solutions. Computing a few updates within each iteration fixed the problem.

[5] Note that if F_0 is greater than the spectral width of the Hann window, the values of the smoothed periodogram at multiples of F_0 can still be considered independent, as assumed in section 2.

[6] We chose $P = 15$, so that the total number of parameters used for modeling the observed spectrum is the same as that of the original ARMA model.

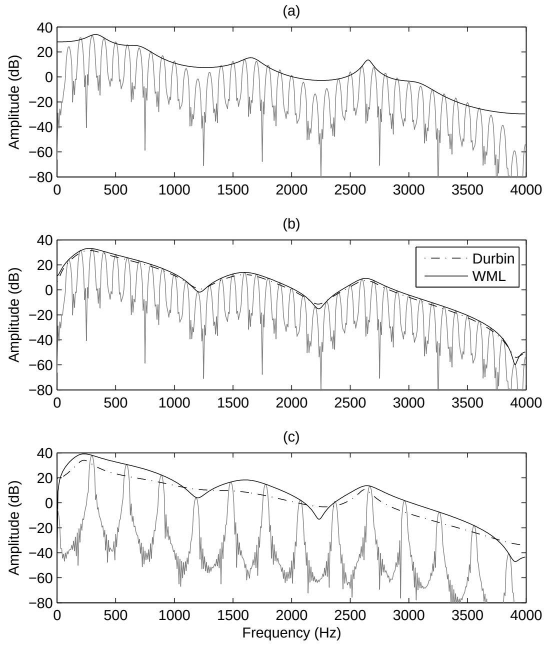

Fig. 1. Estimation of the spectral envelope of a nasal vowel

5. CONCLUSIONS

We proposed new algorithms for estimating AR, MA, and ARMA models in the spectral domain, when only a subset of the original spectrum is observed. This technique is derived from a Weighted Maximum Likelihood (WML) approach, and can be seen as a generalization of the DAP method, which was initially designed for AR models only [2]. In the simple case of AR modeling, we proved that the proposed algorithm converges to the optimal solution, and that the convergence rate is enhanced by remapping the poles at each iteration. In the case of MA modeling, the proposed algorithm still converges to the optimal solution, although the proof could not be included in the paper. The case of ARMA models seems more complex, and we noted that the practical convergence requires a more careful implementation. In all cases, the convergence toward the optimal solution requires that the number of distinct parameters remains lower than the number of non-zero spectral weights. Finally, the algorithm for estimating ARMA models was successfully applied in the context of speech analysis. Our results first confirmed that ARMA models are more appropriate for representing nasal vowels than AR models, and the WML method proved to provide a much more accurate estimation of the envelope of high-pitched harmonic spectra than the classical Durbin method.

APPENDIX

A. CHARACTERIZATION OF THE AR ESTIMATOR

Proof of proposition 3.1. Any critical point $a^\star$ of the $\mathscr{C}^1$ function L is such that the gradient $\nabla L(a^\star)$ is zero. By substituting equation (5) into equation (4) where $B(\nu) \equiv 1$, and by differentiating the resulting expression, we obtain $\frac{1}{2}\nabla L(a) = (R(a,\mu) - R(a,\mu_{YA}))\,a$, where matrices $R(a,\mu)$ and $R(a,\mu_{YA})$ were defined in equations (7) and (8). Note that $R(a,\mu) - R(a,\mu_{YA})$ is a symmetric Toeplitz matrix. Let $\{r_k(a)\}_{k\in\{-P\ldots P\}}$ be the series formed by the coefficients in the first column and the first row of this matrix; this series is symmetric.

Since $\frac{1}{2}\nabla L(a^\star)=0$, matrix $R(a^\star,\mu)-R(a^\star,\mu_{YA})$ is singular. However classical results regarding singular Toeplitz matrices [7] show that the series $\{r_k(a^\star)\}_{k\in\{-P\dots P\}}$ is a linear combination of complex exponentials (possibly modulated by polynomials), whose poles are the roots of the polynomial $A^\star(z)$. If $a^\star\in\mathscr{S}$, all these roots are strictly outside the unit circle, and the symmetry of this series yields only one solution: $r_k(a^\star)=0\ \forall k\in\{-P\dots P\}$, which proves equation (6). Now let show that any solution of equation (6) is a global maximum of function L. For any vector a, substituting equations (6) to (8) in to equation (4) yields

$$\begin{aligned} L(a)-L(a^\star) &= \int_0^1 \log\left(\frac{A(\nu)}{A^\star(\nu)}\right) d\mu(\nu) - \log\left(\int_0^1 \frac{A(\nu)}{A^\star(\nu)}\, d\mu(\nu)\right) \\ &\leq 0 \text{ because of the concavity of function } \log. \end{aligned}$$

□

B. CONVERGENCE ANALYSIS FOR THE AR MODEL

Proof of assertion 1. in Prop. 3.2. Any equilibrium point of the dynamical system $a\leftarrow\mathbb{P}_{\mathscr{S}}(\phi(a))$ in $\mathscr{S}$ is the global maximum of function L. Now let prove that this equilibrium point is locally stable. According to the general theory of discrete dynamical systems, the local stability of an equilibrium point $a^\star$ of a recursion of the form $a\leftarrow f(a)$ is guaranteed if the Jacobian of f at point $a^\star$ has a spectral radius strictly smaller than one. Here, we will prove that the Jacobian $J_\phi(a^\star)$ of function ϕ has all its eigenvalues of magnitude strictly lower than 1, except that associated to the eigenvector $a^\star$, which is equal to 1. This shows that the *direction* of the dynamical system locally converges to the *direction* of $a^\star$. Since in the neighborhood of $a^\star\in\mathscr{S}$, the remapping $\mathbb{P}_{\mathscr{S}}$ consists in normalizing a vector so that its first coefficient becomes 1, we will then conclude that the dynamical system locally converges to $a^\star$. First, differentiating equation (9) shows that the Jacobian of ϕ satisfies

$$J_\phi(a)=R(a,\mu_{YA})^{-1}\left(R(a,\mu)-P(a,\mu)+2u(a,\mu)u(a,\mu_{YA})^T\right) \tag{14}$$

where

$$\begin{aligned} u(a,\nu) &= \frac{T(\nu)a}{A(\nu)} &(15)\\ P(a,\mu) &= 2\int_0^1 u(a,\nu)u(a,\nu)^T d\mu(\nu) &(16)\\ u(a,\mu) &= R(a,\mu)\,a=\int_0^1 u(a,\nu)d\mu(\nu) &(17)\\ u(a,\mu_{YA}) &= R(a,\mu_{YA})\,a. &(18) \end{aligned}$$

Then noting that $T(\nu)$ can be written in the form $T(\nu)=\Re\left(e(\nu)e(\nu)^H\right)$ where $e(\nu)=[1,e^{i2\pi\nu},\dots,e^{i2\pi\nu P}]^T$, it can be verified that

$$P(a,\mu)=R(a,\mu)+H(a,\mu), \tag{19}$$

where the Toeplitz matrix $R(a,\mu)$ was defined in equation (7), and $H(a,\mu)$ is a Hankel matrix:

$$\begin{aligned} R(a,\mu) &= \Re\left(\int_0^1 \frac{e(\nu)e(\nu)^H}{|a^Te(\nu)|^2}d\mu(\nu)\right) &(20)\\ H(a,\mu) &= \Re\left(\int_0^1 \frac{e(\nu)e(\nu)^T}{(a^Te(\nu))^2}d\mu(\nu)\right). &(21) \end{aligned}$$

Then substituting equation (19) into equation (14), we obtain

$$J_\phi(a)=-R(a,\mu_{YA})^{-1}H(a,\mu)+2\phi(a)u(a,\mu_{YA})^T.$$

Applying this equality to the equilibrium point $a^\star$ yields

$$J_\phi(a^\star)=a^\star u(a^\star,\mu)^T-\varepsilon(a^\star), \tag{22}$$

where

$$\varepsilon(a^\star)=R(a^\star,\mu)^{-1}\left(H(a^\star,\mu)-u(a^\star,\mu)u(a^\star,\mu)^T\right). \tag{23}$$

Then it can be verified that $\varepsilon(a^\star)a^\star=0$, and $u(a^\star,\mu)^T\varepsilon(a^\star)=0^T$. Thus zero is an eigenvalue of $\varepsilon(a^\star)$, associated to the right and left eigenvectors $a^\star$ and $u(a^\star,\mu)$, which additionally satisfy $u(a^\star,\mu)^Ta^\star=1$. In other respects, noting that matrix $R(a^\star,\mu)\pm H(a^\star,\mu)$ is positive semi-definite, it can be proved that all the other eigenvalues of $\varepsilon(a^\star)$ have a magnitude strictly lower than 1. This finally proves the local stability of the dynamical system. □

Proof of assertion 2. in Prop. 3.2. If $\mu(\nu)$ is the Lebesgue measure,

- Substituting equation (20) into equation (17) yields
  $$u(a,\mu)=\Re\left\{\int_0^1 \frac{e(\nu)}{a^Te(\nu)}\,d\nu\right\}=\Re\left\{\frac{1}{2i\pi}\int_\Gamma \frac{e(z)}{a^Te(z)}\frac{1}{z}dz\right\}$$
  where Γ is the unit circle and $z=e^{2i\pi\nu}$. The k-th coefficient is $u_k(a,\mu)=\Re\left\{\frac{1}{2i\pi}\int_\Gamma f_k(z)dz\right\}$, where $f_k(z)=\frac{z^{k-1}}{\sum_{p=0}^{P}a_pz^p}$. If $a\in\mathscr{S}$ and $k>0$, the holomorphic function f_k has no singularity inside Γ, thus $u_k(a,\mu)=0$. If $a\in\mathscr{S}$, f_0 has one singularity inside Γ ($z=0$), thus the residue theorem yields $u_0(a,\mu)=1$.
- Equation (21) shows that $H(a,\mu)$ can be written in the form
  $$H(a,\mu)=\Re\left(\frac{1}{2i\pi}\int_\Gamma \frac{e(z)e(z)^T}{(a^Te(z))^2}\frac{1}{z}dz\right).$$
  Similar considerations yield $H(a,\mu)=u(a,\mu)u(a,\mu)^T$.

Consequently, matrix $\varepsilon(a^\star)$ defined in equation (23) is zero. Thus the algorithm converges in one iteration only, which was expected, since $u(a,\mu)$ does not depend on $a\in\mathscr{S}$. □

If μ is a measure close to the Lebesgue measure, the spectral radius of $\varepsilon(a^\star)$ remains much lower than 1. Thus the discrete dynamical system locally converges to $a^\star$ at a high convergence rate.

C. REFERENCES

[1] S. M. Kay, *Modern Spectral Estimation, Theory and Application*, Prentice-Hall, Englewood Cliffs, NJ, USA, 1988.

[2] A. El-Jaroudi and J. Makhoul, "Discrete all-pole modeling," *IEEE Trans. Signal Processing*, vol. 39, no. 2, pp. 411–423, Feb. 1991.

[3] L. Weruaga, "All-Pole Estimation in Spectral Domain," *IEEE Trans. Signal Processing*, vol. 55, no. 10, pp. 4821–4830, 2007.

[4] F. Villavicencio, A. Robel, and X. Rodet, "Improving LPC spectral envelope extraction of voiced speech by true-envelope estimation," in *Proc. of ICASSP'06*, Toulouse, France, May 2006, vol. 1, pp. 869–872, IEEE.

[5] P.M.T. Broersen and S. de Waele, "Time series analysis in a frequency subband," *IEEE Trans. Instrum. Meas.*, vol. 52, no. 4, pp. 1054–1060, Aug. 2003.

[6] J. Durbin, "The Fitting of Time Series Models," *Review of the Int. Statistical Institute*, vol. 28, no. 3, pp. 233–243, 1960.

[7] R. Badeau, B. David, and G. Richard, "High resolution spectral analysis of mixtures of complex exponentials modulated by polynomials," *IEEE Trans. Signal Processing*, vol. 54, no. 4, pp. 1341–1350, Apr. 2006.