ON THE ROLE OF THE PARTIAL TRIGONOMETRIC MOMENT PROBLEM IN AR SPEECH MODELLING

P. Delsarte, Y. Genin, Y. Kamp, P. Van Dooren

Philips Research Laboratory Brussels
Av. Van Becelaere 2, Box 8, B1170 Brussels, Belgium

ABSTRACT

The partial trigonometric moment problem is shown to provide a unifying framework for several speech modelling techniques, such as the classical LPC autoregressive model, the line spectral pairs and composite sinusoidal waves models, and the Toeplitz eigenvector model for formant extraction. From a mathematical viewpoint, this moment problem can be identified to an extension problem in the class of impedance functions or equivalently in the class of nonnegative definite Toeplitz matrices.

INTRODUCTION

The linear prediction or autoregressive (AR) model is one of the most popular techniques for the modelling of discrete time signals by rational functions, especially in the field of speech processing [1]. In particular, the reflection or partial correlation (PARCOR) coefficients have become a classical characterization of an AR model [2]. Itakura and Sugamura have given an alternative efficient parametrization of AR models in terms of line spectral pairs [3]. Sagayama and Itakura have proposed a new speech synthesis method based on a representation of the time signal as a sum of sinusoidal waves [4]. Carayannis and Gueguen have described another speech model, useful for formant extraction, which is built from the eigenvector relative to the smallest eigenvalue of the autocorrelation matrix [5]. The location of the zeros of the corresponding eigenfilter has been examined by Makhoul [6]. As pointed out by Gueguen [7] and Kung [8], this eigenmodel is closely related to a method developed by Pisarenko in the field of geophysics [9].

This contribution aims at showing how the partial trigonometric moment problem [10] provides a unifying theoretical framework for the methods above. It is explained how the set of extensions of a given nonnegative definite Toeplitz matrix can be parametrized in terms of an arbitrary reflectance function (also called S-function). It turns out that the models above can be obtained from this general approach via appropriate choices of the reflectance function [11].

PARTIAL TRIGONOMETRIC MOMENT PROBLEM

A sequence of p+1 complex numbers $c_0, c_1, \ldots, c_p$ being given, the partial trigonometric moment problem [10] requires to find the whole set of positive measures $\mu(\theta)$ having the c_k as first Fourier coefficients :

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} e^{-ik\theta}\, d\mu(\theta) = c_k\,, \quad 0 \leq k \leq p\,. \qquad (1)$$

Of course, a prerequisite consists in obtaining a criterion for the existence of a solution. This problem was solved by Carathéodory and Toeplitz [10]. There exists a solution if and only if the $(p+1) \times (p+1)$ Toeplitz matrix

$$C_p = [\, c_{k-\ell} : 0 \leq k,\ \ell \leq p\,]\,, \qquad (2)$$

with $c_{-k} = \bar{c}_k$ is nonnegative definite. The solution is unique if and only if C_p is singular. In case C_p is nonsingular, there are infinitely many solutions.

Let us now give an equivalent formulation of the problem. To any positive measure $\mu(\theta)$ let us associate the function

$$f(z) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{e^{i\theta}+z}{e^{i\theta}-z}\, d\mu(\theta)\,. \qquad (3)$$

This appears to be a C-function (where C stands for Carathéodory), which means that it is analytic and has nonnegative real part in the unit disk $|z| < 1$. (In circuit theory this is called a "positive function".) Note that any C-function corresponds via (3) to a well-defined positive measure. Assume now $\mu(\theta)$ to be a solution of (1). In view of (3) the Maclaurin expansion of f(z) has the form

$$f(z) = c_0 + 2\sum_{k=1}^{p} c_k z^k + 2\sum_{k=p+1}^{\infty} c'_k z^k\,, \qquad (4)$$

for some numbers c'_k. As a result, the set of solutions to the partial trigonometric moment problem can be represented by the set of C-functions having $c_0, 2c_1, \ldots, 2c_p$ as first coefficients, or, equivalently, by the set of nonnegative definite Toeplitz

17.4

matrices $[c'_{k-\ell}: 0\leqslant k, \ell<\infty]$ of infinite order having C_p as first principal submatrix.

Assuming C_p to be positive definite we now parametrize the set of C-functions f(z) solving the problem. From the solution vector $A_p = [1, A_{p,1}, \ldots, A_{p,p}]^T$ to the Toeplitz system

$$C_p A_p = [\sigma_p, 0, \ldots, 0]^T \qquad (5)$$

construct the polynomial $A_p(z) = [1, z, \ldots, z^p] A_p$. This can be computed from the Levinson recurrence relations

$$A_j(z) = A_{j-1}(z) + k_j z \hat{A}_{j-1}(z), \quad j=1,\ldots,p, \qquad (6)$$

with $A_0(z)=1$, where $\hat{A}_j(z) = z^j \bar{A}_j(1/\bar{z})$ denotes the reciprocal polynomial of $A_j(z)$. The numbers k_j in (6), called reflection coefficients, are derived from $\sigma_{j-1} k_j = -[c_j, c_{j-1}, \ldots, c_1]\, A_{j-1}$, together with $\sigma_j = (1-|k_j|^2)\sigma_{j-1}$ and $\sigma_0 = c_0$. Their main property is $|k_j| < 1$.

Substituting $-k_j$ for k_j in (6) yields a sequence of "dual polynomials" $R_j(z)$; they are defined by

$$R_j(z) = R_{j-1}(z) - k_j z \hat{R}_{j-1}(z), \quad j=1,\ldots,p, \qquad (7)$$

with $R_0(z)=c_0$. The polynomials $A_j(z)$ and $R_j(z)$ are closely related to Szegö's first and second kind orthogonal polynomials relative to the Toeplitz matrix C_p [12]. They can be shown to be devoid of zeros in $|z| \leqslant 1$. Moreover, they are proved from (5) to satisfy

$$A_p(z)\hat{R}_p(z) + \hat{A}_p(z)R_p(z) = 2\sigma_p z^p, \qquad (8)$$

$$R_p(z) = f(z)A_p(z) + z^{p+1} u_p(z), \qquad (9)$$

$$\hat{R}_p(z) = -f(z)\hat{A}_p(z) + z^p v_p(z), \qquad (10)$$

for any given C-solution f(z), where $u_p(z)$ and $v_p(z)$ are analytic functions in $|z|<1$. The function $s_p(z)/v_p(z)$ has the form

$$s_p(z) = [R_p(z) - f(z)A_p(z)]/z[\hat{R}_p(z) + f(z)\hat{A}_p(z)]. \qquad (11)$$

This turns out to be an S-function (where S stands for Schur), which means that it is analytic and satisfies $|s(z)| \leqslant 1$ in the unit disk $|z| < 1$. This property is proved by induction as follows. Let $s_j(z)$ denote the function obtained from replacing p by j in (11). Using (6) and (7) yields

$$z\, s_j(z) = [s_{j-1}(z) - k_j]/[1 - \bar{k}_j s_{j-1}(z)], \qquad (12)$$

hence $1-|z\, s_j|^2 = (1-|k_j|^2)(1-|s_{j-1}|^2)\,|1-\bar{k}_j s_{j-1}|^{-2}$, for $j \leqslant p$. Since $s_{j-1}(0)=k_j$ (in view of $v_{j-1}(0) = 2\sigma_{j-1}$ and $u_{j-1}(0) = 2\sigma_{j-1}k_j$) this implies that $s_j(z)$ is an S-function if such is the case for $s_{j-1}(z)$. The fact that $s_0(z) = [c_0 - f(z)]/z[c_0 + f(z)]$ is an S-function completes the proof. Next, solving (11) for f(z) yields

$$f(z) = [R_p(z) - z\, s_p(z)\, \hat{R}_p(z)]/[A_p(z) + z s_p(z)\, \hat{A}_p(z)]. \qquad (13)$$

Applying the argument above backwards proves that (13) is a C-function whenever $s_p(z)$ is an S-function. Furthermore (8) and (13) yield $f(z)A_p(z) = R_p(z) + 0(z^{p+1})$, so that the desired condition (4) is satisfied. As a conclusion we have the following characterization.

<u>THEOREM</u> (regular case). If the Toeplitz matrix C_p is positive definite the whole set of C-solutions f(z) to the partial trigonometric problem is given by (13) where $s_p(z)$ is an arbitrary S-function.

Let us now consider the case where C_p is singular and nonnegative definite. If q+1 denotes the rank of C_p the result (13) applies for p replaced by q. However, there is only one S-function $s_q(z)$ for which the Maclaurin expansion of $f(z)^q$ has the required form (4) up to the order p. Indeed, the fact that the Toeplitz matrix C_{q+1} is singular can be expressed by $|k_{q+1}|=1$. Since $s_q(0)=k_{q+1}$, this forces the S-function $s_q(z)$ to be a constant [10], namely $s_q(z) = k_{q+1}$, which yields the following characterization.

<u>THEOREM</u> (singular case). If the Toeplitz matrix C_p is nonnegative definite and has rank q+1 with $q < p$, the C-solution f(z) to the partial trigonometric moment problem is unique and is given by

$$f(z) = [R_q(z) - \varepsilon\, z\, \hat{R}_q(z)]/[A_q(z) + \varepsilon\, z\, \hat{A}_q(z)], \qquad (14)$$

where ε is a constant of unit modulus, equal to the (q+1)-th reflection coefficient relative to C_p.

Note that (14) is a rational C-function of degree q+1 satisfying $f(z) = -\bar{f}(1/\bar{z})$. A function of this type, called "lossless positive function" in circuit theory, admits a decomposition

$$f(z) = \sum_{u=0}^{q} \rho_u \frac{e^{i\theta_u} + z}{e^{i\theta_u} - z}, \qquad (15)$$

with $-\pi < \theta_0 < \theta_1 < \ldots < \theta_q \leqslant \pi$ and $\rho_u > 0$ for all u.

Let us finally mention the following circuit theoretical interpretation of the formulas (13) and (14). The function f(z) can be viewed as the input impedance of a passive network consisting of a well-defined lossless two-port closed on a one-part of reflectance s(z) in the regular case and ε in the singular case [13].

APPLICATIONS TO AR SPEECH MODELLING

A standard modelling technique in speech processing is provided by the AR model [1],[2]. It consists of a stable all-pole filter $1/A_p(z)$ excited by a periodic pulse train in the case of voiced speech and by white noise in the case of unvoiced speech.

Let $P(\theta)$ denote the power density function of a given sequence $(x_n)_{n=-\infty}^{+\infty}$ of speech samples x_n (real numbers). The Fourier expansion of $P(\theta)$ has the form $P(\theta) = \sum_{-\infty}^{+\infty} c_k e^{ik\theta}$ where the c_k are the autocorrelation lags :

$$c_k = c_{-k} = \sum_{n=-\infty}^{+\infty} x_n x_{n-k} . \qquad (16)$$

The output sequence (x'_n) of the AR filter has power density

$$P'(\theta) = \sigma_p / |A_p(e^{i\theta})|^2 , \qquad (17)$$

with σ_p the input energy. The polynomial $A_p(z) = 1+A_{p,1}z + \dots + A_{p,p} z^p$ is determined so that the Fourier coefficients c'_k of $P'(\theta)$ match the corresponding c_k for $0 \leq |k| \leq p$. Thus it is required to find a particular solution to the partial trigonometric moment problem (1), namely a solution with $d\mu(\theta)=P'(\theta)d\theta$ where $P'(\theta)$ has the form (17). It was shown by Whittle that this solution is obtained from defining the polynomial $A_p(z)$ as in (5), where C_p is the partial autocorrelation matrix given by (2) and (16). Note that (8) and (17) yield $P'(\theta)=\mathrm{Re}\,[R_p(e^{i\theta})/A_p(e^{i\theta})]$. Hence the C-function (3) appears to be $f(z) = R_p(z)/A_p(z)$. In other words, the appropriate choice for the S-function $s_p(z)$ in the general representation (13) simply is $s_p(z)=0$.

The next simple choices for $s_p(z)$ in (13) are $s_p(z)=1$ and $s_p(z)= -1$, yielding the C-functions

$$f(z)= [R_p(z) \mp z \hat{R}_p(z)] / [A_p(z) \pm z \hat{A}_p(z)] . \qquad (18)$$

These can be viewed as the real "singular solutions" of smallest degree to the regular problem since they have the form (14) but with $q=p$. Thus either function (18) can be decomposed as in (15) with $q=p$. The power density function $P^*(\theta)=d\mu(\theta)/d\theta$ corresponding to $f(z)$ via (3) is found to be $P^*(\theta) = 2\pi \sum_0^p \rho_u \,\delta(\theta-\theta_u)$ and thus consists of $p+1$ spectral rays, located at the points θ_u, with energy ρ_u (for $u=0,1,\dots,p$). As a result, the first $p+1$ autocorrelation lags are given by

$$c_k = \sum_{u=0}^{p} \rho_u e^{-ik\theta_u} , \quad k=0,1,\dots,p, \qquad (19)$$

and the signal is modelled by $x'_n = \sum_0^p \sqrt{\rho_u}\, e^{in\theta_u}$ (within arbitrary phases). For p odd and $s_p(z)=1$ all poles $e^{i\theta_u}$ of (18) occur in conjugate pairs, so that one obtains Sagayama's composite sinusoidal waves model [4], namely $x'_n = 2 \sum_0^t \sqrt{\rho_u} \cos n\theta_u$ with $t=(p-1)/2$.

The formants of the speech signal can be shown to be closely related to the zeros of the denominator polynomials $F(z)=A_p(z)+z\hat{A}_p(z)$ and $G(z)=A_p(z)-z\hat{A}_p(z)$ of (18). In fact, these zeros form the line spectral pairs in Itakura's approach to the classical AR model. It turns out that the line spectral pairs have better discretization and sensitivity properties than the PARCOR coefficients from the viewpoint of bit rate reduction. Note that the filter denominator $A_p(z)$ is retrieved from the line spectral pairs via the relation $2A_p(z)=F(z)+G(z)$, which constitutes the basis of Itakura's line spectral pairs implementation of the AR filter.

Let us finally describe the model introduced by Pisarenko [9]. The largest constant λ_0 is first subtracted from the signal mean energy c_0 under the constraint of nonnegative definiteness of $C_p-\lambda_0 I_{p+1}$. Thus λ_0 is the smallest eigenvalue of the Toeplitz matrix C_p. Consider now the partial trigonometric moment problem relative to the data $c_0-\lambda_0, c_1,\dots,c_p$, which belongs to the singular case. As shown above, the solution is uniquely given by (14), with $\varepsilon = \pm 1$, where $q+1$ is the rank of $C_p-\lambda_0 I_{p+1}$. Adding up the contribution λ_0 to the singular case solution (15) one obtains the particular C-solution

$$f_0(z) = \lambda_0 + \sum_{u=0}^{q} \rho_u \frac{e^{i\theta_u}+z}{e^{i\theta_u}-z} \qquad (20)$$

to the original problem. Note that (20) can in principle be obtained from selecting the appropriate S-function $s_p(z)$ in the regular case solution (13). The power spectral density function $P_0(\theta)$ relative to (20) is given by $P_0(\theta) = \lambda_0 + 2\pi\sum_0^q \rho_u \delta(\theta-\theta_u)$. Thus the resulting model of the signal is a linear combination of white noise of energy λ_0 and of deterministic sinusoidal waves of the Sagayama type.

REFERENCES

[1] J.D. Markel, A.H. Gray, "Linear Prediction of Speech", Springer Verlag, Berlin, 1976.

17.4

[2] J. Makhoul, "Linear prediction : a tutorial review", Proc. IEEE, 63(1975), pp. 561-580.

[3] F. Itakura, N. Sugamura, "LSP speech synthesizer, its principle and implementation", Proc. Speech Study Group of the Acoust. Soc. of Japan, S79-46, Nov. 1979.

[4] S. Sagayama, F. Itakura, "Speech synthesis by composite sinusoidal waves", Proc. Speech Study Group of the Acoust. Soc. of Japan, S79-39, Oct. 1979.

[5] G. Carayannis, C. Gueguen, "The factorial linear modelling : a Karhunen-Loeve approach to speech analysis", Proc. IEEE Int. Conf. Acoust. Speech Signal Process., 1976, pp. 489-492.

[6] J. Makhoul, "On the eigenvectors of symmetric Toeplitz matrices", IEEE Trans. ASSP-29 (1981), pp. 868-872.

[7] C. Gueguen, "Linear prediction in the singular case and the stability of eigenmodels", Proc. IEEE Int. Conf. Acoust. Speech Signal Process., 1981, pp. 881-885.

[8] S.Y. Kung, "A Toeplitz approximation method and some applications", Proc. Int. Symp. Math. Th. Networks and Systems, Santa Monica, 1981, pp. 262-266.

[9] V.P. Pisarenko, "The retrieval of harmonics from a covariance function", Geophys. J.R. Astr. Soc., 33(1973), pp. 347-366.

[10] N.I. Akhiezer, "The Classical Moment Problem", Oliver and Boyd, London, 1965.

[11] P. Delsarte, Y. Genin, Y. Kamp, P. Van Dooren, "Speech modelling and the trigonometric moment problem", Philips J. of Research, to appear.

[12] L. Ya. Geronimus, "Orthogonal Polynomials", Consultants Bureau, New York, 1961.

[13] P. Dewilde, A. Vieira, T. Kailath, "On a generalized Szegö-Levinson realization algorithm for optimal linear predictors based on a network synthesis approach", IEEE Trans. CAS-25 (1978), pp. 663-675.

[14] P. Whittle, "On the fitting of multivariate autoregressions and the approximate canonical factorization of a spectral density matrix", Biometrika, 50(1963), pp. 129-134.