Normality of a Non-linear Transformation of AR parameters: Application to Reflection and Cepstrum Coefficients.

Jean-Yves Tourneret

ENSEEIHT/GAPSE

2 rue Camichel

31071 Toulouse, France

Tel: (33) 5 61 58 83 14

Fax: (33) 5 61 58 82 37

e-mail: tournere@len7.enseeiht.fr

June 5, 1997

ABSTRACT

Two sets of random vectors cannot both be Gaussian if they are nonlinearly related. Thus, Autoregressive (AR) parameters and reflection coefficient (resp. cepstrum coefficient) estimators cannot both be Gaussian for a finite number of

samples. However, most estimators of AR parameters and reflection coefficients (resp. cepstrum coefficients) are Gaussian asymptotically. Thus, the distribution of AR parameter and reflection coefficient (resp. cepstrum coefficient) estimates are close to Gaussian for large samples.

This paper studies the "closeness" between the Gaussian distribution and the "non-linear transformation of Gaussian AR parameters" distribution. A new distance is defined which is based on the Taylor expansion of the non-linear transformation. This "Taylor" distance called M-distance is used to measure the deviations from the Gaussian distribution of reflection coefficient and cepstrum coefficient statistics. A comparison is presented between this distance and Kullback's divergence. The main advantage of the M-distance with respect to other distances is a very simple closed form expression of the deviations from normality. This closed form expression shows that the convergence of the reflection and cepstrum coefficient distribution to its asymptotic Gaussian distribution (when the number of samples tends to infinity) depends on the position of AR model poles in the unit circle.

RESUME

Deux vecteurs aléatoires liés par des relations non-linéaires ne peuvent être simultanément Gaussiens. Cette propriété implique que les estimateurs des paramètres AutoRégressifs (AR) et des coefficients de réflexion (resp. coefficients cepstraux) ne peuvent être simultanément Gaussiens pour un nombre fini d'échantillons. Cependant, la plupart de ces estimateurs sont asymptotiquement

Gaussiens. Donc, pour un “grand” nombre d’échantillons, la loi des estimateurs des paramètres AR et des coefficients de réflexion (resp. coefficients cepstraux) est proche de la loi normale.

Cet article étudie la distance entre la loi normale et la loi de coefficients liés aux paramètres AR par une transformée non-linéaire. Une nouvelle distance basée sur un développement de Taylor de la non-linéarité est étudiée. Cette distance appelée M-distance permet de déterminer les écarts entre la loi des coefficients de réflexion (resp. coefficients cepstraux) et la loi normale. Une comparison entre la M-distance et la divergence de Kullback est présentée. Le principal avantage de la M-distance par rapport aux autres distances est qu’elle permet d’obtenir une expression analytique très simple des écarts entre une loi et la loi normale. Cette expression analytique permet de montrer que la convergence de la loi des coefficients de réflexion et des coefficients cepstraux vers la loi normale (lorsque le nombre d’échantillons tend vers l’infini) dépend de la position des poles du modèle AR dans le cercle unité.

KEYWORDS

Non-Gaussian Processes, AR parameters, Reflection Coefficients, Cepstrum Coefficients, Pattern Recognition, Estimation

I. Introduction

Reflection coefficients and cepstrum coefficients are nonlinearly related to Autoregressive (AR) parameters. These non-linear relations define a diffeomorphism (it is one-to-one onto appropriate subsets, and the function and its inverse are continuously differentiable). The jacobian of this transformation (as well as higher-order derivatives) can be constructed explicitly. This observation has two consequences in **Estimation Theory (**where the AR parameters have to be estimated since they are deterministic):

- *asymptotic viewpoint*: for a large class of methods based on second-order and/or higher-order statistics (autocorrelation method, covariance method, Burg method, ...), the estimated AR parameter vector $\widehat{a}$ can be shown to form an asymptotically normal sequence of estimators of the "true" AR parameter vector a (under appropriate moment conditions for the innovation):

$$\sqrt{n}\left(\widehat{a}-a\right) \underset{n\to+\infty}{\overset{d}{\to}} \mathcal{N}\left(0,\sigma_e^2 R\left(a\right)^{-1}\right) \tag{1}$$

n is the number of samples, σ_e^2 is the driving noise variance, $R\left(a\right)$ is the AR process covariance matrix and $\overset{d}{\to}$ denotes convergence in distribution. Moreover, if the innovation has a finite fourth-order moment, it is possible to demonstrate that [2]:

$$\lim_{n\to+\infty} nE\left[\left(\widehat{a}-a\right)\left(\widehat{a}-a\right)^t\right] = \sigma_e^2 R\left(a\right)^{-1} \tag{2}$$

Applying a regular (differentiable with a non-singular Jacobian) mapping T to $\widehat{a}$ yields an asymptotically normal sequence $T\left(\widehat{a}\right)$ of $T\left(a\right)$:

$$\sqrt{n}\left[T\left(\widehat{a}\right)-T\left(a\right)\right] \underset{n\to+\infty}{\overset{d}{\to}} \mathcal{N}\left(0,\sigma_e^2 DR\left(a\right)^{-1}D^t\right) \tag{3}$$

such that:

$$\lim_{n \to +\infty} nE\left(\left[T\left(\widehat{a}\right) - T\left(a\right)\right]\left[T\left(\widehat{a}\right) - T\left(a\right)\right]^t\right) = \sigma_e^2 DR\left(a\right)^{-1} D^t \quad (4)$$

where D is a matrix defined in ([4], p. 211). This property can be heuristically derived by linearizing $T\left(\widehat{a}\right)$ about $T\left(a\right)$ [7].

Consequently, reflection coefficients and cepstrum coefficients are asymptotically Gaussian.

• *non-asymptotic viewpoint*: Suppose the map is locally invertible. Then, the probability density function (pdf) of $T\left(\widehat{a}\right)$ can be deduced from the pdf of $\widehat{a}$. Since they are linked by non-linear relations, $\widehat{a}$ and $T\left(\widehat{a}\right)$ cannot both be Gaussian. Nevertheless, the pdf's of $\widehat{a}$ and $T\left(\widehat{a}\right)$ are very close to Gaussian for a large number of samples because of the properties (1) and (3). Similarly, the covariance matrices of $\widehat{a}$ and $T\left(\widehat{a}\right)$ can be approximated by $\frac{\sigma_e^2}{n}R\left(a\right)^{-1}$ and $\sigma_e^2 DR\left(a\right)^{-1} D^t$, respectively, for a large number of samples because of the properties (2) and (4).

All the properties described above can be generalized to **Pattern Recognition.** The AR parameters are random in pattern recognition applications. To derive the optimal Bayesian Classifier, the pdf of "a" has to be known, conditioned on each class . According to ([6], p. 22), *the multivariate normal density is an appropriate model for an important situation, viz., the case where the features vectors for a given class are continuous valued, mildly corrupted versions of a single typical or prototype mean vector.* When the features are statistically independent and have the same variance σ^2, the covariance matrix of each class is of the form $\sigma^2 I$ (I being the identity matrix). When the features are not independent, the covariance matrix of each class is no longer diagonal. It can then be expressed as $C_a = \sigma^2 C$. C is a unit norm matrix and σ^2 is the variance of the class (sometimes called within-class scatter)

([6], p. 26). The following observations can then be made:

• *asymptotic viewpoint*:

Applying a regular mapping T to "a" yields a sequence $T(a)$ such that:

$$\frac{1}{\sigma}\left[T(a)-T(m_a)\right] \underset{\sigma\to 0}{\overset{d}{\to}} \mathcal{N}\left(0, DCD^t\right) \tag{5}$$

and:

$$\lim_{\sigma\to 0}\frac{1}{\sigma^2}E\left(\left[T(a)-T(m_a)\right]\left[T(a)-T(m_a)\right]^t\right) = DCD^t \tag{6}$$

where $m_a = E(a)$ and D is the matrix defined in ([4], p. 211).

• *non-asymptotic viewpoint*:

$T(a)$ is related to a by a non-linear transformation. For a non-zero variance σ^2, $T(a)$ cannot be Gaussian. However, for a small value of σ^2, the pdf and covariance matrix of $T(a)$ are very close to Gaussian and DCD^t respectively because of the properties (5) and (6).

For brevity in what follows, the generic terms "Estimation Case" are used whenever AR parameters are deterministic and "Pattern Recognition Case" whenever AR parameters are random. However, a unified notation is used for the AR parameter vector:

• in estimation theory, the AR process is a Gaussian vector a with mean $m_a = E(a)$ and covariance matrix $C_a = \frac{\sigma_e^2}{n}R^{-1}$,

• in pattern recognition, a is a Gaussian vector with $m_a = E(a)$ and $C_a = \sigma^2 C$, C being a unit norm matrix.

The Gaussian assumption for a (for which the reflection coefficients and cepstrum coefficients cannot be Gaussian) can be justified by some of the previous properties. In many practical applications moreover, AR parameter histograms have a Gaussian shape contrary to reflection and cepstrum coefficients ([10], p. 166),([12], p. 146), [15].

This paper studies the "closeness" between the distribution of $T(a)$ and the Gaussian (for a fixed number of samples n or a fixed variance σ^2) as a function of m_a and C_a. Section II presents a recursive way to determine the pdf of reflection coefficients and cepstrum coefficients as functions of the pdf of the AR parameter vector. Section III formulates the problem and defines a new distance, denoted as M-distance, between the statistics of $T(a)$ and Gaussian. In sections IV and V, the M-distance is used to derive a closed form expression of the "closeness" between reflection and cepstrum coefficient distribution and Gaussian.

II. Reflection and Cepstrum Coefficient pdf

The first part reminds the reader of a recursive way of determining the reflection and cepstrum coefficient pdf as a function of the AR parameter vector pdf. The results of this section (which were first derived in [15]) will be used in sections IV and V to compute the Kullback divergence between the reflection and cepstrum coefficient pdf and Gaussian.

A. Reflection Coefficient pdf

Reflection coefficients k_j are linked to AR parameters a_j according to the following relations:

$$\begin{array}{ll} j=1,...,p & a_j^{(p)}=a_j \\ 1\leq j\leq i-1 & a_j^{(i-1)}=\frac{a_j^{(i)}-a_i^{(i)}a_{i-j}^{(i)}}{1-\left[a_i^{(i)}\right]^2} \\ j=1,...,p & a_j^{(j)}=k_j \end{array} \quad (7)$$

The parameters $a_j^{(i)}, j=1,...,i$ are the ith order linear predictor coefficients. For $i=p$ (p is the AR model order), these parameters are the AR parameters. For $i=j$, these parameters are the reflection coefficients. The vectors $\left[a_1^{(i)},...,a_i^{(i)}\right]$ can then be computed recursively for

$i = p-1, ..., 1$ from AR parameters and Eq. (7). Each step allows the determination of one reflection coefficient $k_i = a_i^{(i)}$. Note that, in many algorithms such as Burg's algorithm, AR parameters are computed from reflection coefficients. However, this does not prevent Eq. (7) to be satisfied.

Denote

$$V_i = \left[a_p^{(p)}, a_{p-1}^{(p-1)}, ..., a_i^{(i)}, a_{i-1}^{(i)}, a_{i-2}^{(i)}, ..., a_1^{(i)}\right] \tag{8}$$

This vector can be split into two parts:

- the first one has $p-i+1$ components equal to the $p-i+1$ reflection coefficients:

$$\left[a_p^{(p)}, a_{p-1}^{(p-1)}, ..., a_i^{(i)}\right] = [k_p, ..., k_i] \tag{9}$$

- the second one has $i-1$ components:

$$\left[a_{i-1}^{(i)}, a_{i-2}^{(i)}, ..., a_1^{(i)}\right] \tag{10}$$

In particular, for $i = p$, V_i is the AR parameter vector V_p, whose pdf is assumed to be known. For $i = 1$, V_i is the reflection coefficient vector V_1, whose pdf is unknown. The $p-i+1$ first components of vectors V_i and V_{i-1} are equal and the $i-1$ last ones are linked by relations (7) which can be inverted. The jacobian of the one-to-one transformation between V_i and V_{i-1} can then be computed (for more details see [15]). If $f_i\left(x_p, x_{p-1,...,}x_1\right)$ denotes the pdf of V_i , the pdf of V_{i-1} can be computed with the following relations:

- i odd

$$f_{i-1}\left(x_p, ..., x_1\right) = \left(1 - x_i^2\right)^{(i-1)/2} f_i\left(x^{'}\right) \tag{11}$$

with $x^{'} = (x_p, ..., x_i, x_{i-1} + x_i x_1, ..., x_{1+} x_i x_{i-1})$.

- i even

$$f_{i-1}(x_p, ..., x_1) = (1 + x_i)\left(1 - x_i^2\right)^{(i-2)/2} f_i\left(x''\right) \tag{12}$$

with $x'' = \left(x_p, ..., x_i, x_{i-1} + x_i x_1, ..., (1 + x_i)\, x_{i/2}, ..., x_{1+} x_i x_{i-1}\right)$

By means of p iterations, using (11) and (12), the reflection coefficient vector pdf can be computed from the AR parameter vector pdf.

B. Cepstrum Coefficient pdf

Cepstrum coefficients are linked to AR parameters according to the following non-linear relations:

$$\begin{array}{ll} 2 \leq k \leq p & kc_k = -ka_k - \sum_{i=1}^{k-1} ic_i a_{k-i} \\ k > p & kc_k = -\sum_{i=1}^{p} (k-i)\, c_{k-i} a_i \end{array} \tag{13}$$

with $c_1 = -a_1$. It is obvious from Eq. (13) that all the information contained in AR parameters is in the p first cepstrum coefficients, mainly because of the one-to-one transformation between these two finite dimension sets of parameters. In most applications, only the p first cepstrum coefficients are then considered. The p first relations (13) can be written in the following form:

$$\begin{aligned} c_k &= -a_k + f_{k-1}(a_1, ..., a_{k-1}) \\ a_k &= -c_k + g_{k-1}(c_1, ..., c_{k-1}) \end{aligned} \tag{14}$$

f_{k-1} and g_{k-1} being two polynomial functions with $k-1$ variables $a_1, ..., a_{k-1}$ and $c_1, ..., c_{k-1}$ respectively. From the form of equations (14), the jacobian matrix of the one-to-one transformation between the two vectors $a = (a_1, ..., a_p)$ and $c = (c_1, ..., c_p)$ is an upper triangular matrix, whose diagonal terms are equal to -1. The determinant of this matrix is consequently equal to $(-1)^p$. The cepstrum coefficient pdf $c(x_1, ..., x_p)$ can then be determined from the AR parameter pdf $a(x_1, ..., x_p)$:

$$c(x_1, ..., x_p) = a(-x_1, ..., -x_p + g_{p-1}(x_1, ..., x_{p-1})) \tag{15}$$

The functions f_k and g_k can be determined recursively using equations (13) (for more details see [15],[17]).

III. Distance between $T(a)$ pdf and Gaussian

Many distances between random variables, such as the Kullback divergence or the Bhattacharyya distance, can be used to measure the closeness between $T(a)$ pdf (when it can be determined) and the Gaussian pdf [1]. Using the results of the previous section, a "closeness" measure between the reflection and cepstrum coefficient distribution and Gaussian can then be derived. However, these distances depend on m_a and C_a through non-linear (in general integral) relations which are difficult to study. In this part, a new distance, which depends on m_a and C_a by very simple relations, is presented. For simplicity, the pattern recognition case for which $C_a = \sigma^2 C$ is considered. However, the results can be extended easily to the estimation case for which $C_a = \frac{\sigma_e^2}{n} R^{-1}$. The study is restricted to the set S of variables X satisfying the two following conditions:

$$\forall k \in \mathbb{N} \quad M_X^k = E\left(X^k\right) < +\infty \tag{16}$$

$$\limsup_{k \to +\infty} \left(\frac{\left| M_X^k \right|}{k!} \right)^{1/k} < +\infty \tag{17}$$

Variables satisfying conditions (16) and (17) are characterized by their moments ([13], p. 290). In this section, T is a transformation from $\mathbb{R}^p$ into $\mathbb{R}$ (p is the AR model order) and the first two derivatives of $T(a)$, denoted by $T'(a)$ and $T''(a)$, exist in a neighborhood of m_a. In the case of a transformation $T(a) = [T_1(a), ..., T_m(a)]^t$ from $\mathbb{R}^p$ into $\mathbb{R}^m$, the m transformations $T_i(a)$, $i \in \{1, ..., m\}$ can be considered separately. In what follows, the sequence of variables $\{X_\sigma\}$ converges in probability to zero, written $X_\sigma = o_p(1)$, if

$$\forall \varepsilon > 0 \qquad \lim_{\sigma \to 0} P[|X_\sigma| > \varepsilon] = 0 \tag{18}$$

The sequence $\{X_\sigma\}$ is bounded in probability, written $X_\sigma = O_p(1)$, if

$$\forall \varepsilon > 0 \qquad \exists \delta(\varepsilon) > 0 \text{ such that } P[|X_\sigma| > \delta(\varepsilon)] < \varepsilon \text{ for all } \sigma \tag{19}$$

In the estimation case, the covariance matrix of the AR process a is of the form $C_a = \sigma^2 C$ such that

$$a - m_a = O_p(\sigma) \tag{20}$$

which means that $\frac{1}{\sigma}(a - m_a)$ is bounded in probability. According to ([4], p. 210), $T(a)$ admits the following Taylor expansions:

$$T(a) = T(m_a) + T'(m_a)(a - m_a) + o_p(\sigma) \tag{21}$$

$$T(a) = T(m_a) + T'(m_a)(a - m_a) + \frac{1}{2}T''(m_a)(a - m_a, a - m_a) + o_p\left(\sigma^2\right) \tag{22}$$

The computation of the second order derivative $T''(m_a)(a - m_a, a - m_a)$ can be made from the Hessian $H(a)$ of the transformation T (see [3] p. 143). For a transformation from $\mathbb{R}^p$ into $\mathbb{R}$, $T''(m_a)$ is defined by:

$$T''(m_a)(u, v) = u^t H(a) v \tag{23}$$

with $[H(a)]_{ij} = \frac{\partial^2 T}{\partial a_i \partial a_j}$. Eq.'s (22) and (21) mean that the random vectors

$$\frac{1}{\sigma}\left[T(a) - T(m_a) - T'(m_a)(a - m_a)\right]$$

$$\frac{1}{\sigma^2}\left[T(a) - T(m_a) - T'(m_a)(a - m_a) - \frac{1}{2}T''(m_a)(a - m_a, a - m_a)\right]$$

converge in probability to zero. For small values of σ, the variable $o_p(\sigma^2)$ is negligible with respect to the three first terms in (22). Under these conditions, the Gaussian or non-Gaussian nature of the random vector $T(a)$ is due to the two following terms:

$$G_\sigma = T'(m_a)(a - m_a) \tag{24}$$

$$NG_\sigma = \frac{1}{2}T''(m_a)(a - m_a, a - m_a) \tag{25}$$

In what follows, to make shorter, we will write G and NG, instead of G_σ and NG_σ. If the development (21) reduces to $T(a) - T(m_a) \simeq G$, there exists a linear relation between $T(a) - T(m_a)$ and $a - m_a$. These two vectors are then both Gaussian. On the other hand, when the second order term NG is not negligible, the vector $T(a) - T(m_a)$ is no longer Gaussian.

Define the following distance between variables X and Y [16]:

$$d(X, Y) = \sum_{k=1}^{+\infty} \frac{\left|M_X^k - M_Y^k\right|}{k!} \tag{26}$$

This distance comes from the l_1 norm applied to infinite sequences of the form

$$1, \frac{iM_X^1}{1!}, \ldots, \frac{(i)^n M_X^n}{n!}, \ldots \tag{27}$$

which appears naturally in the development of the characteristic function in terms of its moments. In general, d is not a distance because two different random variables can have the same moments ([9] p. 12). But variables satisfying conditions (16) and (17) are characterized by their moments such that d is a distance on S. For a Gaussian vector a, variables G and $G+NG$ belong to the set S with $R=+\infty$. Appendix A then shows that:

$$\lim_{\sigma^2 \to 0} d(G, G+NG) = \lim_{\sigma^2 \to 0} \sum_{k=1}^{+\infty} \frac{1}{k!} \left| M_G^k - M_{G+NG}^k \right| = 0 \tag{28}$$

Eq. (28) is in accordance with the asymptotic normality of $T(a)$ defined in Eq. (5). Since a is Gaussian, $d(G, G+NG)$ can be expressed as

$$d(G, G+NG) = \sum_{k=1}^{M} d_k + o\left(\sigma^{2M}\right) \tag{29}$$

with $d_k = O\left(\sigma^{2k}\right)$. For small values of σ, $d(G, G+NG)$ can be approximated by the first few terms in (29) such that

$$d(G, G+NG) \simeq \sum_{k=1}^{M} d_k \tag{30}$$

In (30), d_M is the first d_k depending on the AR model poles (see appendices B and C for the expression of d_M in the case of reflection and cepstrum coefficients). Similarly, in estimation, $d(G, G+NG)$ can be expressed as:

$$d(G, G+NG) = \sum_{k=1}^{M} d_k + o\left(\frac{1}{n^M}\right) \simeq \sum_{k=1}^{M} d_k \tag{31}$$

with $d_k = O\left(\frac{1}{n^k}\right)$. In both cases, the lower $d(G, G+NG)$, the lower the distance between the distribution of $T(a)$ and the Gaussian. Next sections use Eqs. (30) and (31) to study the convergence of the reflection coefficient and cepstrum coefficient distribution to the Gaussian as a function of the position of the AR model poles in the unit circle.

IV. Application to Reflection Coefficients

For simplicity, a Gaussian real second order AR parameter vector a is considered. $E(a) = m_a$ can be expressed as a function of two conjugated complex poles $p_1 = \rho e^{j\varphi}$ and $p_2 = \rho e^{-j\varphi}$ such that $m_a = (-2\rho\cos\varphi, \rho^2)^t$. The reflection coefficient vector k is linked to the AR parameter vector with the non-linear transformation $k = \left(\frac{a_1}{1+a_2}, a_2\right)^t$. The second reflection coefficient being Gaussian, only the first one is considered such that

$$T(a) = \frac{a_1}{1+a_2} \tag{32}$$

The first terms d_k in Eqs. (30) and (31) (which correspond to the first reflection coefficient) can then be determined as a function of ρ and φ (see Appendix B):

Pattern Recognition Case

$$d_1 = \sigma^2 \left| \frac{-2\rho c_{22}\cos\varphi}{(1+\rho^2)^3} - \frac{c_{12}}{(1+\rho^2)^2} \right| \tag{33}$$

$$d_2 = \frac{\sigma^4}{2(1+\rho^2)^6} \left| 12c_{22}^2\rho^2\cos^2\varphi + \left(c_{11}c_{22} + 2c_{12}^2\right)\left(1+\rho^2\right)^2 - 12\rho\cos\varphi\left(1+\rho^2\right)c_{12}c_{22} \right| \tag{34}$$

Estimation Case

$$d_1 = 0 \tag{35}$$

$$d_2 = \frac{(1-\rho^2)^2}{2n^2(1+\rho^2)^4} \left[\left(1+\rho^2\right)^2 - 4\rho^2\cos^2\varphi \right] \tag{36}$$

Higher order terms can be derived in a similar way but, for large number of samples, $d(G, G+NG)$ can be approximated by the first d_k. This property is illustrated in Fig. 1

a) and b), representing the variations of d_1 (continuous line) and $d_1 + d_2$ (dotted line) as a function of ρ and φ for the pattern recognition case. In this case, the distance $d(G, G+NG)$ can be approximated by the first term d_1.

The variations of $d(G, G + NG)$ as a function of ρ and φ for the estimation case are plotted in Fig. 2 a) and b). In both cases, the distance between the reflection coefficient distribution and the Gaussian is very small when the AR model poles are close to the unit circle.

To show the efficiency of our distance, the variations of $d(G, G+NG)$ are compared with the Kullback divergence between the first reflection coefficient pdf (determined in the first part of the paper) and the Gaussian. The Kullback divergence between two variables X and Y, with respective pdf's p_1 and p_2, is defined by:

$$d_K(p_1, p_2) = \int_{\mathbf{R}} [p_2(x) - p_1(x)] \ln \frac{p_2(x)}{p_1(x)} dx \tag{37}$$

Figs. 1 and 3 show that the qualitative behavior of the two approaches is very similar.

The lower the $d(G, G + NG)$, the lower the distance between the reflection coefficient distribution and the Gaussian. Thus, this distance is very low when the AR model poles are close to the unit circle. The three dimensional curves plotted in Fig 4, describing the variations of $d(G, G + NG)$ as a function of ρ and φ, confirm this result.

This section shows that the convergence of the reflection coefficient pdf to the Gaussian (when the number of samples tends to infinity in the Estimation case or when the variance tends to zero in the Pattern Recognition case) is related to the AR model pole position in the unit circle. Consequently, the analysis of the AR model pole position in the unit circle allows us to determine whether the reflection coefficient distribution can be approximated by the Gaussian or not.

V. Application to Cepstrum Coefficients

The second order cepstrum coefficient vector c is linked to AR parameters with the non-linear transformation $c = \left(-a_1, -a_2 + \frac{1}{2}a_1^2\right)^t$. The first cepstrum coefficient being Gaussian, only the second one is considered such that:

$$T(a) = -a_2 + \frac{1}{2}a_1^2 \tag{38}$$

The first terms in Eqs. (30) and (31) (which correspond to the second cepstrum coefficient) can then be determined as a function of ρ and φ (see Appendix C):

Pattern Recognition Case

$$d_1 = \frac{\sigma^2 c_{11}}{2} \tag{39}$$

$$d_2 = \sigma^4 \left[\frac{3c_{11}^2}{8} + \frac{1}{2}c_{12}^2 + 3c_{11}c_{12}\rho\cos\varphi + 3c_{11}^2\rho^2\cos^2\varphi + \frac{1}{4}c_{11}c_{22}\right] \tag{40}$$

Estimation Case

$$d_1 = \frac{1}{2n}\left(1 - \rho^4\right) \tag{41}$$

$$\begin{aligned} d_2 \quad = \quad & \frac{3\left(1 - m_2^2\right)^2}{8n^2} + \frac{\left(1 - m_2^2\right)^2}{n^2}\left[3\rho^2\cos^2\varphi + \frac{1}{4}\right] \\ & + \frac{m_1\left(1 - m_2\right)^2}{n^2}\left[\frac{1}{2}m_1 + 3\left(1 - m_2\right)\left(1 + m_2\right)^2\rho\cos\varphi\right] \end{aligned} \tag{42}$$

Higher order terms can be derived in a similar way but, for large number of samples, $d(G, G + NG)$ can be approximated by the first terms in (30) and (31). This property is illustrated in Fig. 5 a) and b), representing the variations of d_1 (dotted line), $d_1 + d_2$ (continuous line) and $\sum_{i=1}^{3} d_i$ (cross line) as a function of ρ and φ for the pattern recognition case. The first term d_1 does not depend on ρ and φ. The distance $d(G, G + NG)$ can be approximated by the two first terms in (30).

In the estimation case, from Eq. (41) and (42), d_1 (dotted line) depends on ρ but not on φ. One term in (31) is then sufficient to describe the evolution of $d(G, G + NG)$ as a function of the modulus, as it is illustrated in Fig. 6 a). For the variations of $d(G, G+NG)$ as a function of the phase, Fig. 6 b) shows that two terms in (26) have to be considered.

To show the efficiency of our method, the variations of $d(G, G + NG)$ are compared (in the pattern recognition case) with the Kullback divergence between the second cepstrum coefficient pdf and the Gaussian. Figs. 5 and 7 show that the two approaches lead to similar results. The convergence of the cepstrum coefficient distribution to the Gaussian is different in the two cases:

- in the pattern recognition case, the closer to the origin the AR model poles, the lower the distance between the cepstrum coefficient distribution and the Gaussian.
- in the estimation case, the closer to the unit circle the AR model poles, the lower the distance between the cepstrum coefficient distribution and the Gaussian

The three dimensional curves plotted in Fig 8, describing the variations of $d(G, G+NG)$ as a function of ρ and φ, confirm these results.

This section shows that the convergence of the cepstrum coefficient pdf to the Gaussian (when the number of samples tends to infinity in the Estimation case or when the variance tends to zero in the Pattern Recognition case) is related to the AR model pole position in the unit circle. Consequently, the analysis of the AR model pole position in the unit circle allows us to determine whether the cepstrum coefficient distribution can be approximated by the Gaussian or not.

VI. Conclusion

The distribution of a non-linear transformation of Gaussian AR parameters cannot be Gaussian. The M-distance between this distribution and Gaussian, based on the Taylor expansion of the non-linearity, is defined. The M-distance is used to measure the deviations from the normal distribution of reflection coefficient and cepstrum coefficient statistics. Other methods such as the expansions in Edgeworth or Gram-Charlier series could be used as well. However, the main advantage of the M-distance is a very simple closed form expression of these deviations as a function of the AR parameter vector mean and covariance matrix. This closed form expression shows that the distance between reflection or cepstrum coefficient statistics and Gaussian depends, by very simple relations, on the position of the AR model poles in the unit circle. Such a result is of great interest in estimation theory and in pattern recognition:

In estimation theory, it shows that the number of samples necessary to approximate the reflection coefficient and cepstrum coefficient distribution by the Gaussian is directly related to the position of the AR model poles in the unit circle.

In pattern recognition, the M-distance depends on the class variance but on the position of the AR model poles in the unit circle as well. This allows us to determine in which cases the Gaussian approximation can be used for reflection and cepstrum coefficients. This is very important, since a fine knowledge of the parameter pdf is not necessary for the optimal Bayesian classifier.

VII. Acknowledgements

Thanks to our visitor from the US Professor Neil Bershad. His help significantly improved the readability of the paper. Thanks to F. Castanié, B. Lacaze, C. Mailhes and to the two anonymous referee's for their helpful comments and suggestions.

APPENDIX A

PROOF OF EQUATION (28)

This appendix shows that variables G and $G+NG$ satisfy the following equation:

$$\lim_{\sigma\to 0} d(G, G+NG) = \lim_{\sigma\to 0}\sum_{k=1}^{+\infty} \frac{1}{k!}\left|M_G^k - M_{G+NG}^k\right| = 0 \tag{43}$$

For simplicity, the study is restricted to the transformations T from $\mathbb{R}^p$ into $\mathbb{R}$. According to the form of the covariance matrix $C_a = \sigma^2 C$, $G = T'(m_a)(a - m_a)$ and $NG = \frac{1}{2}T''(m_a)(a - m_a, a - m_a)$ converge in mean square and then in distribution to 0. The characterization of the convergence in distribution due to Levy leads to:

$$\lim_{\sigma\to 0} E\left[e^{itG}\right] = 1 \qquad \forall t \in \mathbb{R} \tag{44}$$

$$\lim_{\sigma\to 0} E\left[e^{itNG}\right] = 1 \qquad \forall t \in \mathbb{R} \tag{45}$$

Since a is Gaussian, variables G and NG belong to the set S with $R = +\infty$ and ([13], p. 290):

$$E\left[e^{itG}\right] = \sum_{k=1}^{+\infty} \frac{M_G^k (it)^k}{k!} \qquad t \in \mathbb{R} \tag{46}$$

$$E\left[e^{it(G+NG)}\right] = \sum_{k=1}^{+\infty} \frac{M_{G+NG}^k (it)^k}{k!} \qquad t \in \mathbb{R} \tag{47}$$

hence:

$$\lim_{\sigma\to 0}\sum_{k=1}^{+\infty} \frac{(it)^k}{k!}\left(M_G^k - M_{G+NG}^k\right) = 0 \qquad t \in \mathbb{R} \tag{48}$$

Eq. (48) leads to:

$$\lim_{\sigma\to 0}\sum_{k=1}^{+\infty} \frac{1}{k!}\left|M_G^k - M_{G+NG}^k\right| = 0 \tag{49}$$

APPENDIX B

COMPUTATION OF d_k FOR REFLECTION COEFFICIENTS

In this appendix, the first moments of variables G and $G + NG$ are determined for reflection coefficients in the case of a second order AR model. Similar results can be obtained for higher order AR models. AR parameters and reflection coefficients are linked by the following relations:

$$\begin{pmatrix} k_1 \\ \\ k_2 \end{pmatrix} = \begin{pmatrix} \frac{a_1}{1+a_2} \\ \\ a_2 \end{pmatrix} \tag{50}$$

The second reflection coefficient being Gaussian, only the first one is considered such that:

$$T(a) = \frac{a_1}{1 + a_2} \tag{51}$$

$$T(m_a) = \frac{m_1}{1 + m_2} \tag{52}$$

with $m_a = \begin{pmatrix} m_1 \\ \\ m_2 \end{pmatrix} = E(a)$. The first and second order derivatives of T can then be computed:

$$T'(m_a) = \begin{pmatrix} \frac{1}{1+m_2} \\ \\ \frac{-m_1}{(1+m_2)^2} \end{pmatrix} \tag{53}$$

$$T_1''(m_a) = \begin{pmatrix} 0 & \frac{-1}{(1+m_2)^2} \\ \\ \frac{-1}{(1+m_2)^2} & \frac{2m_1}{(1+m_2)^3} \end{pmatrix} \tag{54}$$

Equations (24) and (25) lead to:

$$G = T'(m_a)(a - m_a) = \frac{a_1 - m_1}{1 + m_2} - \frac{m_1(a_2 - m_2)}{(1 + m_2)^2} \tag{55}$$

$$NG = \frac{1}{2}T''(m_a)(a - m_a, a - m_a) = \frac{m_1(a_2 - m_2)^2 - (a_1 - m_1)(a_2 - m_2)(1 + m_2)}{(1 + m_2)^3} \tag{56}$$

The moments of NG can be determined as a function of the AR parameter covariance matrix C_a.

- Pattern Recognition case $C_a = \sigma^2 \begin{pmatrix} c_{11} & c_{12} \\ c_{12} & c_{22} \end{pmatrix}$

  The following relations

$$E\left[(a_2 - m_2)^2\right] = \sigma^2 c_{22} \qquad E\left[(a_1 - m_1)(a_2 - m_2)\right] = \sigma^2 c_{12} \tag{57}$$

  lead to:

$$E(NG) = d_1 = \sigma^2 \frac{m_1 c_{22} - c_{12}(1 + m_2)}{(1 + m_2)^3} \tag{58}$$

  which can be expressed as a function of the AR model poles:

$$E(NG) = \sigma^2 \frac{-2\rho c_{22} \cos\varphi - c_{12}(1 + \rho^2)}{(1 + \rho^2)^3} \tag{59}$$

  It is well known that higher order moments of the Gaussian distribution can be determined as a function of its mean and covariance matrix . In particular, for $i \in \{1, 2\}$ and $j \in \mathbb{N}$:

$$\begin{aligned} E\left[(a_i - m_i)^{2j+1}\right] &= 0 \\ E\left[(a_i - m_i)^{2j+2}\right] &= \frac{(2j+1)!}{2^j j!}\sigma^{2j+2} \end{aligned}$$

  Higher order moments of NG can then be computed. For instance, the three following relations

$$\begin{aligned} &E\left[(a_2 - m_2)^4\right] = 3\sigma^4 c_{22}^2 \\ &E\left[(a_1 - m_1)^2 (a_2 - m_2)^2\right] = \sigma^4 (c_{11}c_{22} + 2c_{12}^2) \\ &E\left[(a_1 - m_1)(a_2 - m_2)^3\right] = 3\sigma^4 c_{12}c_{22} \end{aligned} \tag{60}$$

lead to:

$$d_2 = \frac{\sigma^4}{2\left(1+m_2\right)^6}\left(\left|3c_{22}^2 m_1^2 + \left(c_{11}c_{22} + 2c_{12}^2\right)\left(1+m_2\right)^2 - 6m_1\left(1+m_2\right)c_{12}c_{22}\right|\right) \quad (61)$$

- Estimation Case $C_a = \frac{\sigma_e^2}{n} R^{-1}$

For large number of samples, the Cramer-Rao Bound of the AR process can be used for C_a:

$$C_a = \frac{1}{n}\left[A_1 A_1^T - A_2 A_2^T\right]^{-1} \quad (62)$$

A_1 and A_2 are two matrices which depend on the AR parameter vector m_a[8]. In the case of a second order AR process, Eq. (62) leads to:

$$C_a = \frac{1}{n}\begin{pmatrix} 1-m_2^2 & m_1\left(1-m_2\right) \\ m_1\left(1-m_2\right) & 1-m_2^2 \end{pmatrix} \quad (63)$$

Thus:

$$d_1 = \frac{m_1\left(1-m_2^2\right) - m_1\left(1-m_2\right)\left(1+m_2\right)}{n\left(1+m_2\right)^3} = 0 \quad (64)$$

$$d_2 = \frac{\left(1-m_2\right)^2\left[\left(1+m_2\right)^2 - m_1^2\right]}{2n^2\left(1+m_2\right)^4} \quad (65)$$

APPENDIX C

COMPUTATION OF d_k FOR CEPSTRUM COEFFICIENTS

This appendix determines the first moments of variables G and $G + NG$ for cepstrum coefficients in the case of a second order AR model. AR parameters and cepstrum coefficients are linked by the following relations:

$$\begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = \begin{pmatrix} -a_1 \\ -a_2 + \frac{1}{2}a_1^2 \end{pmatrix} \tag{66}$$

This leads to:

$$T(a) = -a_2 + \frac{1}{2}a_1^2 \tag{67}$$

$$T\left(m_a\right) = -m_2 + \frac{1}{2}m_1^2 \tag{68}$$

with $m_a = \begin{pmatrix} m_1 \\ m_2 \end{pmatrix} = E\left(a\right)$. The first and second order derivatives of T can then be computed:

$$T'\left(m_a\right) = \begin{pmatrix} m_1 \\ -1 \end{pmatrix} \tag{69}$$

$$T''\left(m_a\right) = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \tag{70}$$

Equations (24) and (25) lead to:

$$G = T'\left(m_a\right)\left(a - m_a\right) = -\left(a_2 - m_2\right) + m_1\left(a_1 - m_1\right) \tag{71}$$

$$NG = \frac{1}{2}T''\left(m_a\right)\left(a - m_a, a - m_a\right) = \frac{1}{2}\left(a_1 - m_1\right)^2 \tag{72}$$

In this particular case, the third order derivative of T is zero such that c is exactly equal to $T(a) + G + NG$ (the third and higher order terms in (22) are equal to zero). The moments of $\acute{G}$ and $G + NG$ can be determined as a function of the AR parameter covariance matrix C_a.

- Pattern Recognition case $C_a = \sigma^2 \begin{pmatrix} c_{11} & c_{12} \\ c_{12} & c_{22} \end{pmatrix}$

  Eq. (71) and (72) lead to:

$$d_1 = E(NG) = \frac{\sigma^2 c_{11}}{2} \tag{73}$$

$$\frac{1}{2}\left|E\left(NG^2\right) + 2E(NG.G)\right| = \frac{3\sigma^4 c_{11}^2}{8} \tag{74}$$

  The two first term in $d(G, G+NG)$ do not depend on the AR model poles. The third term in $d(G, G+NG)$ depends on the third moments of variables G and NG:

$$\frac{1}{6}\left|E(NG^3) + 3E(NG^2.G) + 3E\left(NG.G^2\right)\right| \tag{75}$$

  It can be expressed as:

$$\frac{5\sigma^6 c_{11}^3}{16} + \sigma^4\left[3c_{11}^2\rho^2\cos^2\varphi + \frac{1}{4}c_{11}c_{22}\right] + \sigma^4\left[\frac{1}{2}c_{12}^2 + 3c_{11}c_{12}\rho\cos\varphi\right] \tag{76}$$

  hence

$$d_2 = \sigma^4\left[\frac{3c_{11}^2}{8} + 3c_{11}^2\rho^2\cos^2\varphi + \frac{1}{4}c_{11}c_{22} + \frac{1}{2}c_{12}^2 + 3c_{11}c_{12}\rho\cos\varphi\right] \tag{77}$$

- Estimation Case $C_a = \frac{\sigma_e^2}{n}R^{-1}$

  Eq. (73), (74) and (75) yield:

$$d_1 = \frac{1 - m_2^2}{2n} \tag{78}$$

$$
\begin{aligned}
d_2 &= \frac{3\left(1-m_2^2\right)^2}{8n^2} + \frac{\left(1-m_2^2\right)^2}{n^2}\left[3\rho^2\cos^2\varphi + \frac{1}{4}\right] \\
&\quad + \frac{m_1\left(1-m_2\right)^2}{n^2}\left[\frac{1}{2}m_1 + 3\left(1-m_2\right)\left(1+m_2\right)^2\rho\cos\varphi\right] \qquad (79)
\end{aligned}
$$

References

[1] M. Basseville, "*Distance Measures for Signal Processing and Pattern Recognition,*" Signal Processing, vol. 18, n°6, pp. 349-369, December 1989.

[2] K. Berk, *"Consistent Autoregressive Spectral Estimates,"* The Annals of Statistics, vol. 2, n°3, pp. 489-502, 1974.

[3] P. G. Ciarlet, *Introduction à l'Analyse Numérique Matricielle et à l'Optimisation*, Masson, 1982.

[4] P. J. Brockwell and R. Davis, *Time Series: Theory and Methods.* New York: Springer-Verlag, 1991, 2nd Ed.

[5] H. Cramer, *Mathematical Methods of Statistics,* Princeton Univ. Press, 1946.

[6] R. O. Duda, *Pattern Classification and Scene Analysis*, Wiley, New York, 1973.

[7] S. Kay and J. Makhoul, "On the Statistics of the Estimated Reflection Coefficients of an Autoregressive Process," *IEEE Trans. Acoust., Speech, Sig. Proc.*, vol. 31, n°6, pp. 1447-1455, December 1984.

[8] B. Friedlander, "On the Computation of the Cramer-Rao Bound for ARMA Parameter Estimation," *IEEE Trans. Acoust., Speech, Sig. Proc.*, vol. 32, n°6, pp. 721-727, August 1984.

[9] E. Lukacs, *Characteristic functions*, Griffin, London, 1970.

[10] C. Mailhes, "Modélisation et Compression d'Interférogrammes," Ph. D. dissertation, n°310, INP Toulouse, France, 1990.

[11] J. M. Mendel, "Tutorial on Higher-Order Statistics (Spectra) in Signal Processing and

System Theory: Theoretical Results and Some Applications," *Proc. of the IEEE,* vol. 79, n°3, pp. 278-305, March 1991.

[12] H. Poublan, "Reconnaissance Automatique de Signaux à Evolution Continue," Ph. D. dissertation, n°518, INP Toulouse, France, 1991.

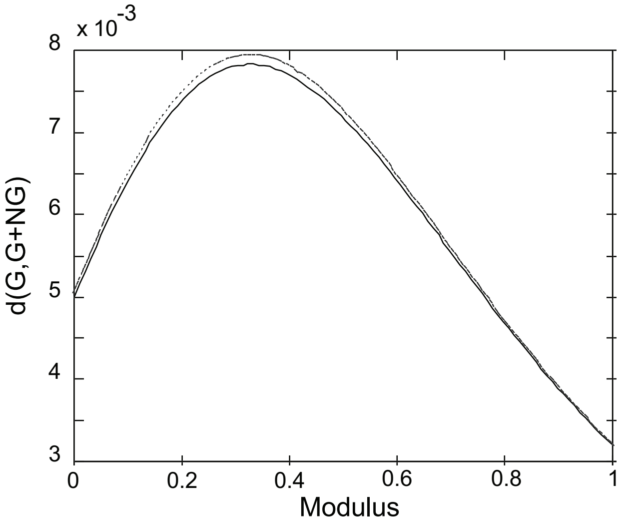
[13] A. Rényi, *Calcul des Probabilités,* Dunod, France, 1966.

[14] J.Y. Tourneret, and B. Lacaze, "Study of the couple (Cepstrum Coefficient, k-NN Rule)," *Proc. of 7th IEEE Workshop on Stat. and Array Proc. (SSAP'94)*, pp. 357-360, Quebec, June 26-29, 1994.

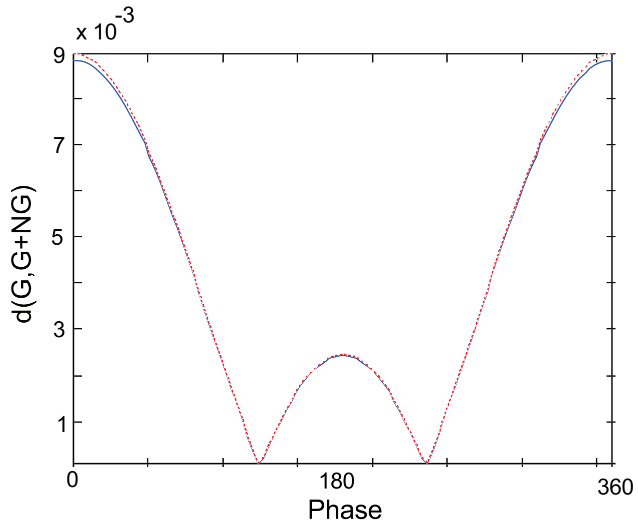
[15] J.Y. Tourneret, and B. Lacaze, "On the statistics of estimated cepstrum and reflection coefficients of an autoregressive process," Signal processing, Vol. 43, No 3, May 1995.

[16] J.Y. Tourneret and B. Lacaze, "Statistical study of reflection coefficients derived from a random AR process," Proc. of ICASSP'95, pp. 1709-1712, Detroit, May 8-12, 1995.

[17] J.Y. Tourneret and B. Lacaze, "Convergence Rate of the Cepstrum Coefficient Distribution to the Gaussian," Proc. of the IEEE Work. on HOS, pp. 200-203, Girona, Spain, June 12-14, 1995.



(a)

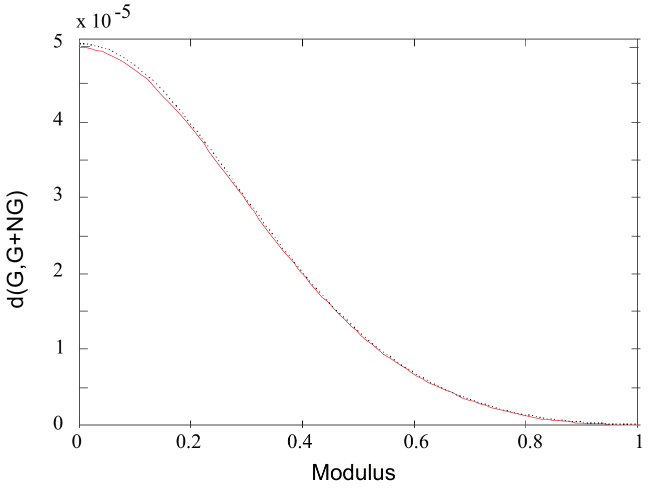


(b)

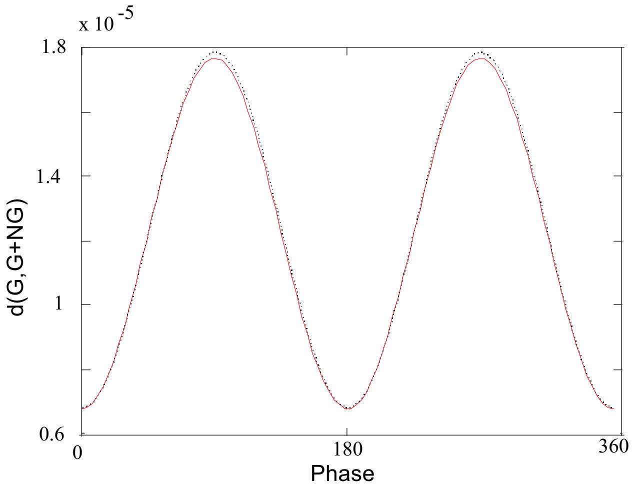
Pattern Recognition Case

$$C_a = \sigma^2 C = 10^{-2} \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$$

Fig. 1. $d\,(G, G+NG)$ *for the first reflection coefficient (a) as a function of* ρ $\left(\varphi = \frac{\pi}{4}\right)$ *(b) as a function of* φ $\left(\rho = \frac{1}{2}\right)$ *(continuous line :* d_1 *, dotted line :* $d_1 + d_2$*)*



(a)



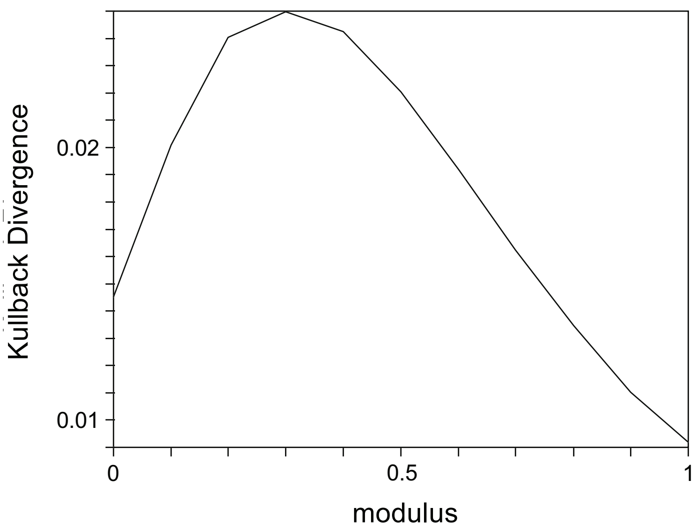
(b)

Estimation Case

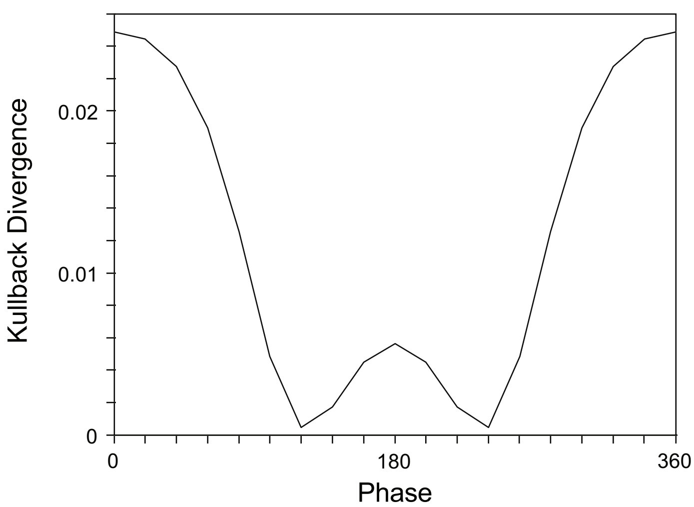
$$C_a = \frac{\sigma_e^2}{n} R^{-1}, n = 100, \sigma^2 = 1$$

Fig. 2. $d\left(G, G+NG\right)$ *for the first reflection coefficient* (a) *as a function of* ρ $\left(\varphi = \frac{\pi}{4}\right)$

(b) *as a function of* φ $\left(\rho = \frac{1}{2}\right)$*(continuous line :* d_2 *, dotted line :* $d_2 + d_3$*)*



(a)



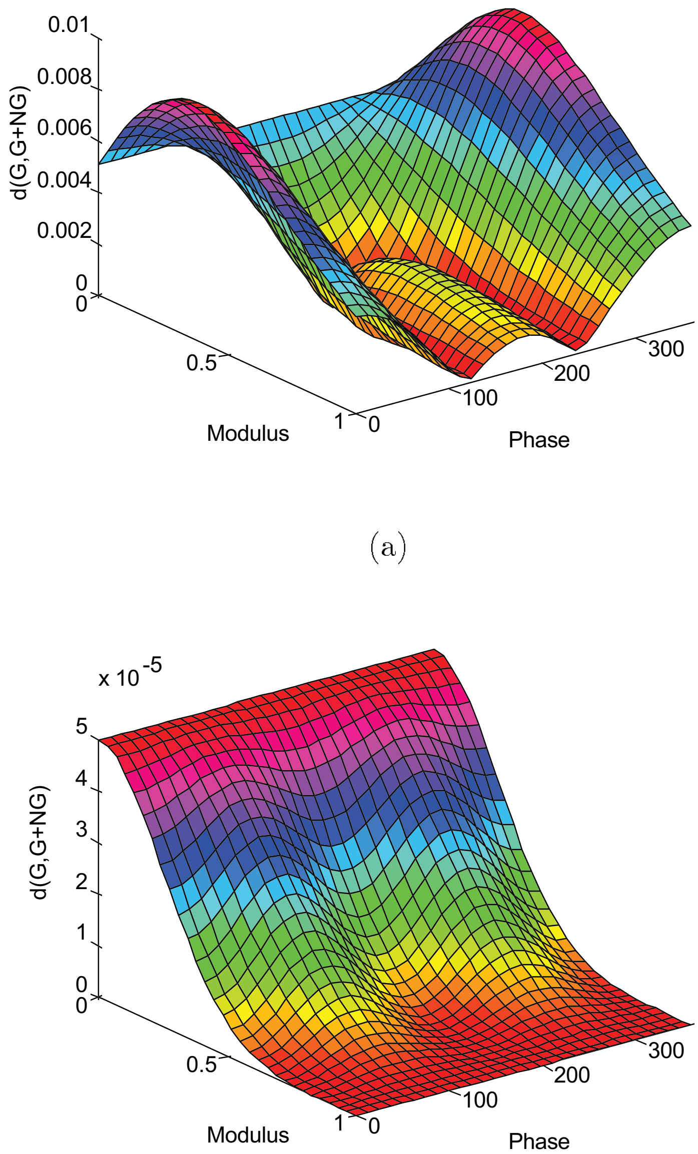
(b)

Pattern Recognition Case

$$C_a = \sigma^2 C = 10^{-2} \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$$

Fig. 3. Divergence of Kullback between the first reflection coefficient pdf and Gaussian

(a) as a function of ρ $\left(\varphi = \frac{\pi}{4}\right)$ *(b) as a function of* φ $\left(\rho = \frac{1}{2}\right)$

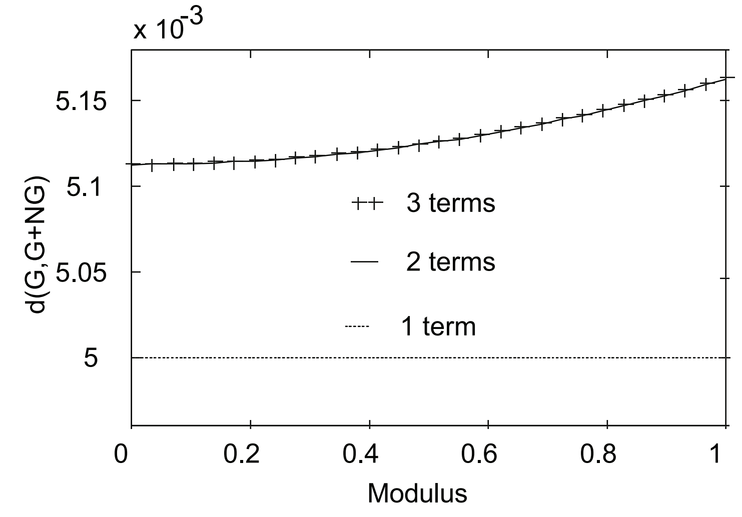


(a)

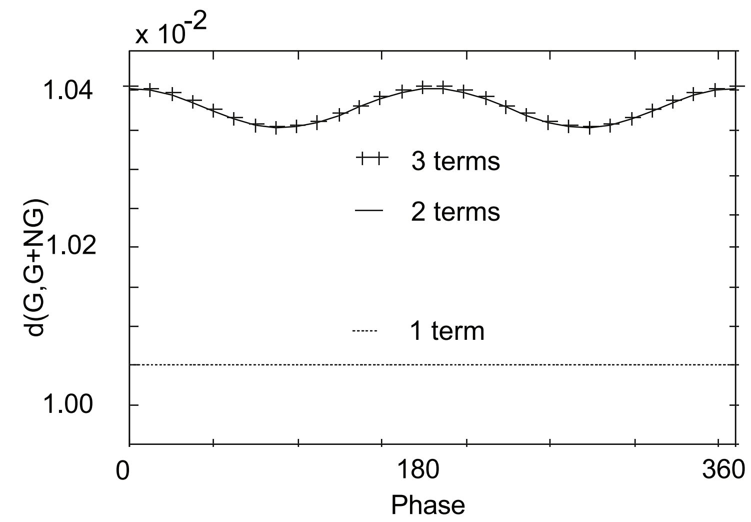
(b)

Fig. 4. $d(G, G + NG)$ *for the first reflection coefficient as a function of* ρ *and* φ

(a) Pattern Recognition Case (b) Estimation case



(a)



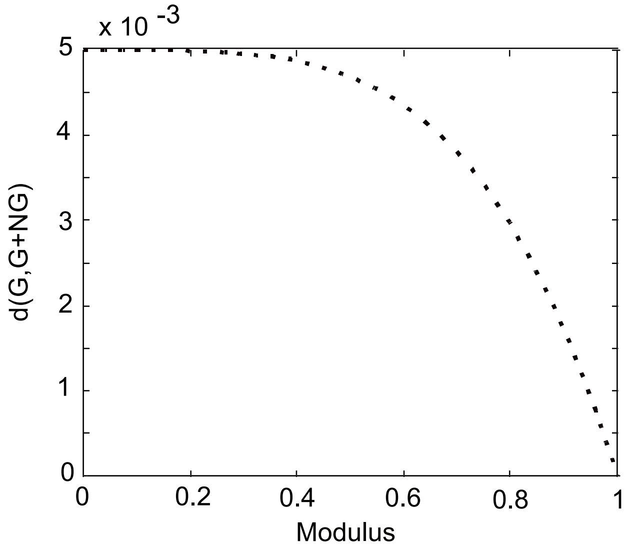
(b)

Pattern Recognition Case

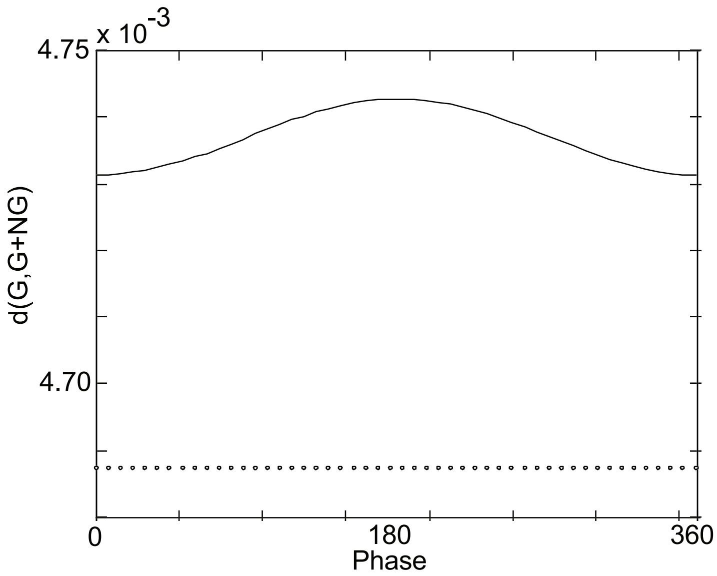
$$C_a = \sigma^2 C = 10^{-2} \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$$

Fig. 5. $d(G, G+NG)$ *for the second cepstrum coefficient (a) as a function of* ρ $\left(\varphi = \frac{\pi}{4}\right)$

(b) as a function of φ $\left(\rho = \frac{1}{2}\right)$



(a)

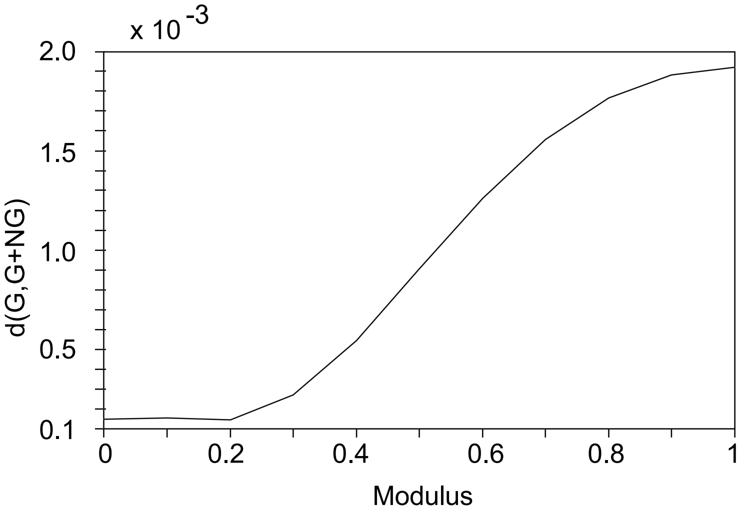


(b)

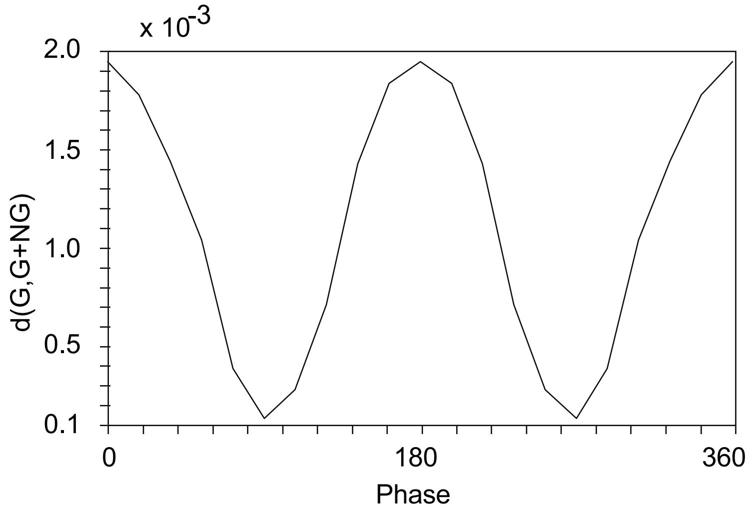
Estimation Case

$$C_a = \frac{\sigma_e^2}{n} R^{-1}, n = 100, \sigma_e^2 = 1$$

Fig. 6. $d(G, G+NG)$ for the second cepstrum coefficient (a) as a function of ρ $\left(\varphi = \frac{\pi}{4}\right)$ (b) as a function of φ $\left(\rho = \frac{1}{2}\right)$ (dashed line: d_1, continuous line: $d_1 + d_2$)



(a)

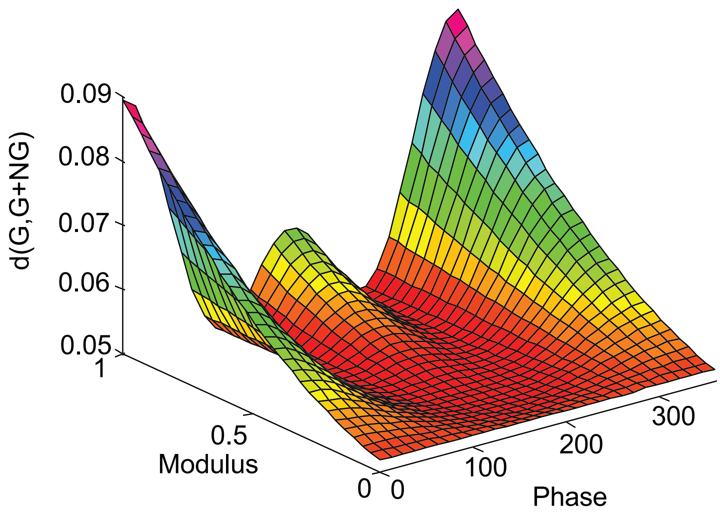


(b)

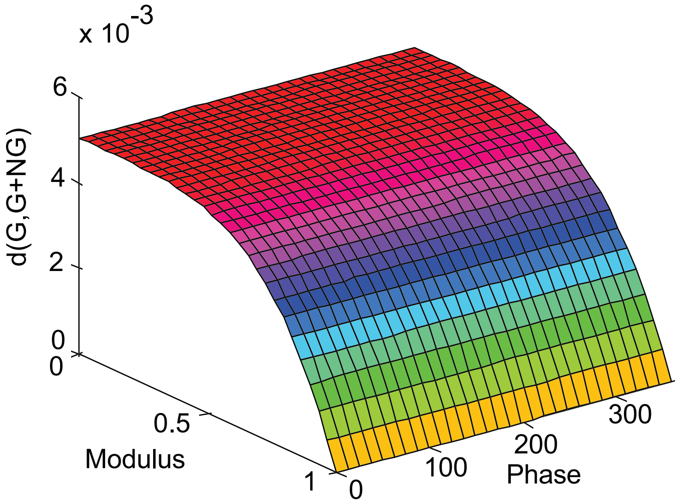
Pattern Recognition Case

$$C_a = \sigma^2 C = 10^{-2} \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$$

Fig. 7. Divergence of Kullback between the second cepstrum coefficient pdf and the Gaussian (a) as a function of ρ $\left(\varphi = \frac{\pi}{4}\right)$ *(b) as a function of* φ $\left(\rho = \frac{1}{2}\right)$



(a)



(b)

Fig. 8. $d(G, G + NG)$ *for the second cepstrum coefficient as a function of* ρ and φ

(a) Pattern Recognition Case (b) Estimation case