KYBERNETIKA — VOLUME 23 (1987), NUMBER 6

SPECTRUM-ORIENTED SOURCE CODING THEORY

OTAKAR ŠEFL, IGOR VAJDA

In this paper we present rigorous mathematical foundations of a source coding theory with spectral distances at the place of distortion measures. Such a theory is applicable e.g. in speech coding. We prove a new general joint source/channel coding theorem which is of theoretical as well as of practical importance. We also establish correctness of some frequently used conclusions and procedures for which semirigorous "proofs" are known to the authors only, or the rigorous proofs are too scattered in the literature. The theory presented in this paper is oriented not only to recipients of speech signals but to all users concerned about the spectral structure of signals rather than about signals as such.

1. SPECTRAL DISTORTION FUNCTIONS

This paper is a direct continuation of the paper [5], including the terminology and notation. That paper is assumed to be available to the reader of the present paper. In the present paper we restrict ourselves to continuous stationary ergodic information sources $(\mathbb{R}^\infty, \mathscr{A}^\infty, \nu)$ satisfying the condition (25) of [5].

It follows from Example 5 and Proposition 3 in [5] that there is a category of users interested in *spectral distortion functions*

$$d_n(x, \mu);\ x \in \mathbb{R}^n, \quad \mu \in \mathfrak{M}. \tag{1}$$

Here $\mathfrak{M}$ is the set of all wide-sense stationary sources $(\mathbb{R}^\infty, \mathscr{A}^\infty, \mu)$ with positively definite covariance functions r_μ: $\{\ldots, -1, 0, 1, \ldots\} \mapsto \mathbb{R}$ (i.e. all $s \times s$ matrices $[r_{\mu,k-j}]$, $s \geqq 1$, are positively definite) satisfying the condition

$$\sum_{k=-\infty}^{\infty} |r_{\mu,k}| < \infty\,, \tag{2}$$

and with spectral densities φ_μ: $[-\pi, \pi] \mapsto [0, \infty]$ defined by

$$\varphi_\mu(\omega) = \sum_{k=-\infty}^{\infty} r_{\mu,k}\, \mathrm{e}^{-\mathrm{i}k\omega} \tag{3}$$

It is further assumed in (1) that

(4) $$d_?(x, \mu) = D(\varphi_x, \varphi_\mu)\,, \quad x \in \mathbb{R}^n\,, \quad \mu \in \mathfrak{M}\,,$$

where φ_x is the spectral density of x defined by (28) in [5] and D is a *spectral distance*, i.e. any mapping $D: \Phi \times \Phi \mapsto [0, \infty)$, for

(5) $$\Phi = \{\varphi_\mu : \mu \in \mathfrak{M}\}\,,$$

such that, for any $\varphi, \psi, \tilde{\varphi} \in \Phi$,

(6) $$D(\varphi, \psi) \geqq 0\,, \quad D(\varphi, \psi) \leqq D(\varphi, \tilde{\varphi}) + D(\tilde{\varphi}, \psi)$$

Correctness of the condition (4) follows from the next statement.

Proposition 1. It holds $\varphi_x \in \Phi$ for every $x \in \mathbb{R}^n$, $n \geqq 1$.

Proof. It follows from (28) in [5] that φ_x is defined by (3) with μ formally replaced by x for r_x given by (26) or (27). Clearly, (26) as well as (27) are positively definite and satisfy (2). □

Example 1. Let L_α, $1 \leqq \alpha \leqq \infty$, be the linear set of all functions $\varphi: [-\pi, \pi] \mapsto \mapsto [-\infty, \infty]$ with finite norm

$$\|\varphi\|_\alpha = \begin{cases} (1/2\pi \int_{-\pi}^{\pi} |\varphi(\omega)|^\alpha \, d\omega)^{1/\alpha} & \text{if} \quad 1 \leqq \alpha < \infty \\ \text{ess sup}\, |\varphi| \quad \text{if} \quad \alpha = \infty\,, \end{cases}$$

where ess sup is taken with respect to the rectangular probability distribution on $[-\pi, \pi]$. It follows from (2), (3) that, for every $\varphi \in \Phi$, $|\varphi(\omega)|$ is bounded above by the left-hand term of (2) and, consequently,

$$\Phi \subset L_\alpha\,, \quad 1 \leqq \alpha \leqq \infty\,.$$

Therefore the mapping $\mathbb{D}_\alpha: \Phi \times \Phi \mapsto [0, \infty)$ defined for $1 \leqq \alpha \leqq \infty$ by

(7) $$\mathbb{D}_\alpha(\varphi, \psi) = \|\varphi - \psi\|_\alpha$$

is a spectral distance. In fact (7) is a metric on Φ for it is symmetric and $\|\varphi - \psi\|_\alpha = 0$ iff $\varphi = \psi$ a.s. which, in view of (2), (3), is equivalent to $\varphi(\omega) = \psi(\omega)$ on $[-\pi, \pi]$.

Example 2. By what precedes Example 5 in [5], every $\varphi \in \Phi$ is nonnegative. Let us consider a function $f: [0, \infty] \mapsto [0, 8]$ which satisfies the condition

(8) $$f(x) = 0 \quad \text{iff} \quad x = 1$$

and which is nonincreasing on $[0, 1]$ and nondecreasing on $[1, \infty)$. Let for every $\varphi, \psi \in \Phi$

(9) $$D_f(\varphi, \psi) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f\left(\frac{\varphi(\omega)}{\psi(\omega)}\right) d\omega$$

where

(10) $$\frac{x}{0} = \begin{cases} \infty & \text{if} \quad x > 0 \\ 1 & \text{if} \quad x = 0\,. \end{cases}$$

It is easy to see that D_f is a spectral distance provided

$$(11) \qquad f(xy) \leqq f(x) + f(y) \quad \text{for every} \quad 0 \leqq x, y \leqq \infty .$$

Indeed, the nonnegativity of f together with (8) implies

$$D_f(\varphi, \psi) \geqq 0 , \quad \varphi, \psi \in \Phi .$$

with the equality iff $\varphi = \psi$, and (11) implies the triangle inequality considered in (6). Since, further, the condition

$$(12) \quad f(1/x) = f(x) \quad \text{for every} \quad 0 \leqq x \leqq \infty \quad (\text{with } 1/0 = \infty , \; 1/\infty = 0)$$

obviously implies the symmetry

$$(13) \qquad D_f(\varphi, \psi) = D_f(\psi, \varphi) , \quad \varphi, \psi \in \Phi ,$$

it follows from here that if, in addition to (8), (9), f satisfies (12) then D_f is a metric on Φ. An example of a function f satisfying (8), (9), (12) is

$$f_c(x) = c|\ln x| , \quad c > 0 \quad (\text{with} \quad |\ln 0| = |\ln \infty| = \infty) .$$

More generally, if G is any subgroup of the multiplicative group $(0, \infty)$ then (8), (9), (12) are satisfied by

$$f_{c,G}(x) = \infty \; 1_{G^c}(x) + 1_G(x) f_c(x)$$

where the products and the sum are assumed to be defined in accordance with the arithmetic of the extended real line (cf. e.g. § 0 in [1]).

2. SOURCE CODING THEOREM FOR SPECTRAL DISTORTION

The above introduced concepts and examples specify the basic conceptual framework of the source coding theory presented in this paper. The basic step done above was the assumption that there are users who are satisfied when, instead of a message $\boldsymbol{x} \in \mathbb{R}^n$, they obtain a wide-sense stationary information source $\mu \in \mathfrak{M}$ with small distortion $d_n(\boldsymbol{x}, \mu) = D(\varphi_{\boldsymbol{x}}, \varphi_\mu)$. This point deserves to be classified in more detail. In view of what was said in [5], we can restrict ourselves to $\boldsymbol{x} \in \mathbb{B}_n \subset \mathbb{R}^n$ for which the covariance function $r_{\boldsymbol{x}}$ is defined by (26) in [5]. It follows from the next proposition that, under the standard assumptions about source, and for arbitrary $1 \leqq \alpha \leqq \infty$, the distortion $d_n(\boldsymbol{x}, \mu) = \mathbb{D}_\alpha(\varphi_{\boldsymbol{x}}, \varphi_\mu)$ can be arbitrarily closely approximated by $\mathbb{D}_\alpha(\varphi_{\boldsymbol{x}}, \varphi_{\boldsymbol{y}})$, where $\boldsymbol{y} = (y_1, \ldots, y_N)$ is a message from the N-source $(\mathbb{R}^N, \mathscr{A}^N, \mu_N)$ and $\varphi_{\boldsymbol{y}}$ is defined by (26)–(28) in [5] with n and $\boldsymbol{x}$ replaced by N and $\boldsymbol{y}$. In other words, in spite of that a user's ear cannot hear the probability measure as such (if the user is a recipient of the speech signal $\boldsymbol{x}$), it can hear N-messages $\boldsymbol{y}$ produced by the N-sources $(\mathbb{R}^N, \mathscr{A}^N, \mu_N)$, $N = 1, 2, \ldots$, and, if N is large enough, then the user faces distortions $\mathbb{D}_\alpha(\varphi_{\boldsymbol{x}}, \varphi_{\boldsymbol{y}})$ close to $\mathbb{D}_\alpha(\varphi_{\boldsymbol{x}}, \varphi_\mu)$.

Proposition 2. Let for $\mu \in \mathfrak{M}$ the source $(\mathbb{R}^\infty, \mathscr{A}^\infty, \mu)$ be stationary ergodic, and let $\varepsilon > 0$. Then there exists $m \geqq 1$, $N_0 > m$ such that for every $n > m$, $\boldsymbol{x} \in \mathbb{B}_n$,

$N > N_0, 1 \leqq \alpha \leqq \infty$,

$$\mu_N(\{y \in \mathbb{R}^N \mid \mathbb{D}_\alpha(\varphi_x, \varphi_y) < \mathbb{D}_\alpha(\varphi_x, \mu) + \varepsilon\} > 1 - \varepsilon .$$

Proof. Let $\varepsilon > 0$, let $m_0 \geqq 1$ be such that for $m > m_0$

$$\sum_{|k|>m} |r\mu_{,k}| < \tfrac{1}{2}\varepsilon$$

and let us consider arbitrary $n > m > m_0$ and $x \in \mathbb{B}_n$. It follows from Proposition 3 in [5] and from the definition of φ_y that there exists $N_1 \geqq m$ such that for every $N > N_1$

$$\mu_N(\mathbb{B}_N) > 1 - \tfrac{1}{2}\varepsilon$$

and that, for $y \in \mathbb{B}_N \subset \mathbb{R}^N$, $\varphi_y(\omega)$ is defined by (28) and (26) in [5] with n, x replaced by N, y. Further, by the ergodic theorem it follows from (26) that there exists $N_0 \geqq N_1$ such that for every $N > N_0$ there exists $\mathbb{C}_N \in \mathscr{A}^N$ such that

$$\mu_N(\mathbb{C}_N) > 1 - \tfrac{1}{2}\varepsilon$$

and, for every $y \in \mathbb{C}_n$, the function r_y defined by (26) satisfies the inequality

$$\sum_{k=-m}^{m} |r_{y,k} - r_{\mu,k}| < \tfrac{1}{2}\varepsilon .$$

It follows from here and from the definitions of $\varphi_y(\omega)$, $\varphi_\mu(\omega)$ that if $N > N_0$ and $y \in \mathbb{C}_N \cap \mathbb{B}_N$, then for every $-\pi \leqq \omega \leqq \pi$

$$|\varphi_y(\omega) - \varphi_\mu(\omega)| \leqq \tfrac{1}{2}\varepsilon + \sum_{|k|>m} |r_{\mu,k}| < \varepsilon .$$

On the other hand, if $N > N_0$ then

$$\mu_N(\mathbb{C}_N \cap \mathbb{B}_N) > 1 - \varepsilon .$$

The rest of the proof is clear from (7). □

Thus the distortion level $\mathbb{D}_\alpha(\varphi_x, \varphi_\mu)$ is practically achievable by the user provided he is equipped with a computer able to evaluate sufficiently long random messages $y = (y_1, \ldots, y_N)$ from sources $\mu \in \mathfrak{M}$. In this area there is a potential source of user's discontent with the present source coding project. Namely, the question is whether the computer able to evaluate sufficiently long messages y in the real time covered by the signal $x \in \mathbb{R}^n$ is not too expensive. Let t (in second) be the speech segment described by signals $x \in \mathbb{R}^n$ (according to Markel and Gray Jr. [3], t is of order 10^{-2} second and n is of order 10^2; further, m is of order 10 and N is of order 10^2 too), and let us analyze the computer's performances required by the task above, under the assumption that μ is stationary Gaussian regular (then the assumptions of Proposition 1 hold). Let us assume that the computer's memory contains an N-tuple of numbers — a realization of N i.i.d. random variables with zero means and variances $\sigma^2 > 0$. Then the computer is able to generate messages $y \in \mathbb{R}^N$ from the source μ_N at time t if its memory contains an $N \times N$ matrix and if it is able to perform N^2 multiplications and additions at time t (multiplication of an N-vector by the $N \times N$

matrix). These terms are at least problematic. If however μ is an autoregressive source (AR-source, see the definitions between Examples 4 and 5 in [5]) with parameters $\sigma^2 > 0$, $\boldsymbol{a} \in \mathbb{A}_m$, then it suffices to keep in the computer memory two m-tuples of numbers (one is a realization of the random input m vector and the other is $\boldsymbol{a}$), and to perform $(m + 1) N$ operations of multiplication and addition at time, t, as required by (19) in [5]. These terms are fully acceptable for most users.

Hence the above considered source coding model, described by a stationary ergodic source $(\mathbb{R}^\infty, \mathscr{A}^\infty, \nu)$, by a fixed natural number m, and by distortion functions d_n, $n > m$, defined by (4) using a spectral distance D, is completed by the subset $\mathfrak{A} \subset \mathfrak{M}$ of *available codebook elements*, where $\mathfrak{A}$ is the set of all AR-sources $\mu \in \mathfrak{M}$ with parameters $(\sigma^2, \boldsymbol{a}) \in (0, \infty) \times \mathbb{A}_m$ which can in principle be used to encode the source n-messages $\boldsymbol{x} \in \mathbb{R}^n$ for all $n > m$. The following identities are used in the sequel

$$(14) \qquad \mu = (\sigma^2, \boldsymbol{a}) , \quad \mathfrak{A} = (0, \infty) \times \mathbb{A}_m .$$

The source coding theory based on these concepts and assumptions is called *spectrum oriented source coding theory*.

Let us consider an arbitrary fixed spectral distance D_f and, for all natural, n, a *codebook* $C_n \subset \mathfrak{A}$ containing $1 \leqq \|C_n\| < \infty$ AR-models $\mu = (\sigma^2, \boldsymbol{a})$. Let $F_{1,n}$: $\mathbb{R}^n \mapsto C_n$ be a mapping defined by

$$D_f(\varphi_{\boldsymbol{x}}, \varphi_{F_{1,n}(\boldsymbol{x})}) = \min_{\mu \in C_n} D_f(\varphi_{\boldsymbol{x}}, \varphi_\mu) , \quad \boldsymbol{x} \in \mathbb{R}^n .$$

By definition, $F_{1,n}$ is an optimum coding of source n-messages $\boldsymbol{x} \in \mathbb{R}^n$ into the codebook C_n under the spectral distance D_f. In accordance with (4), we denote

$$(15) \qquad d_n(x, C_n) = D_f(\varphi_x, \varphi_{F_{1,n}(x)}) , \quad x \in \mathbb{R}^n .$$

Proposition 3. For every natural n, the functions $F_{1,n}(\boldsymbol{x})$ and $d_n(\boldsymbol{x}, C_n)$ are $\mathscr{A}^n$-measurable on $\mathbb{R}^n$.

Proof. It suffices to prove that, for every natural n and every $\mu \in \mathfrak{A}$, the function $D_f(\varphi_x, \varphi_\mu)$ is $\mathscr{A}^n$-measurable on $\mathbb{R}^n$. It follows from (26)–(28) in [5] that $\varphi_x(\omega)$ is $\mathscr{A}^n \otimes \mathscr{B}$-measurable on $\mathbb{R}^n \times [-\pi, \pi]$, where $\mathscr{B}$ denotes the σ-algebra of Borel subsets of $[-\pi, \pi]$. Further, by (3), $\varphi_\mu(\omega)$ is $\mathscr{B}$-measurable on $[-\pi, \pi]$. Consequently, in view of the piecewise monotonicity of f assumed in Example 2,

$$f\left(\frac{\varphi_x(\omega)}{\varphi_\mu(\omega)}\right)$$

is $\mathscr{A}^n \otimes \mathscr{B}$ measurable on $\mathbb{R}^n \times [-\pi, \pi]$. The desired assertion now follows from 26C in Halmos [1] and from (9). □

Let us return to the function (15) describing the least achievable distortion of source n-messages by a codebook C_n. The *average distortion* of the n-source $(\mathbb{R}^n, \mathscr{A}^n, \nu_n)$ by the codebook C_n is defined as follows

$$d_{n,\nu}(C_n) = \int_{\mathbb{R}^n} d_n(x, C_n) \, \mathrm{d}\nu_n(x) .$$

Let, as in [5],

$$R(C_n) = \frac{1}{n} \log_2 \|C_n\| \geqq 0$$

be the *information rate* of the codebook C_n and let

$$\begin{aligned} \mathbb{C}_n(R) &= \{C_n \subset \mathfrak{A} \mid R(C_n) \leqq R\}\,, & R \geqq 0\,, \\ \delta_{n,\nu}(R) &= \inf\{d_{n,\nu}(C_n) \mid C_n \in \mathbb{C}_n(R)\}\,, & R \geqq 0\,, \\ \delta_\nu(R) &= \liminf_{n\to\infty} \delta_{n,\nu}(R)\,, & R \geqq 0\,. \end{aligned}$$

The function $\delta_\nu(R)$, $R \geqq 0$, is called *distortion-rate function* of the source $(\mathbb{R}^\infty, \mathscr{A}^\infty, \nu)$. A distortion-rate function $\delta_\nu(R)$ is said *regular* if for every $R \geqq 0$ there exist codebooks $C_n \in \mathbb{C}_n(R)$, $n = 1, 2, \ldots$ and a constant $\gamma \geqq 0$ such that

$$\delta_\nu(R) = \lim_{n\to\infty} d_{n,\nu}(C_n)\,, \tag{16}$$

$$\max_{\mu_1,\mu_2 \in C_n} D_f(\varphi_{\mu_1}, \varphi_{\mu_2}) \leqq \gamma\,. \tag{17}$$

If the function f figuring in the spectral distance D_f is decreasing on $[0, 1]$ and increasing on $[1, \infty)$ and if $m = 1$ and $\mathfrak{A}_1 \subset \mathfrak{A}$ is the set of all first-order AR-sources (σ^2, a_1) such that

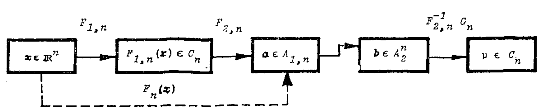
$$\varepsilon_1 \leqq \sigma^2 \leqq 1/\varepsilon_1\,, \quad |a_1| \leqq 1 - \varepsilon_2\,,$$

where $0 < \varepsilon_1, \varepsilon_2 < 1$ are arbitrary fixed, then (17) holds for all codebooks $C_n \in \mathfrak{A}_1$.

Let us now consider a nonanticipating *communication channel*

$$((A_1^\infty, \mathscr{A}_1^\infty), (P_z \mid z \in A_1^\infty), (A_2^\infty, \mathscr{A}_2^\infty)) \tag{18}$$

with *input constraints* $(A_{1,n} \in \mathscr{A}_1^n \mid n = 1, 2, \ldots)$ and with a *capacity* $C \geqq 0$ (cf. (12) in [5]). We shall describe a joint source/channel coding and decoding scheme for an arbitrary source $(\mathbb{R}^\infty, \mathscr{A}^\infty, \nu)$ under consideration. Let n be an arbitrary natural number and C_n



a codebook from $\mathbb{C}_n(R)$, $R \geqq 0$, containing not more elements than $A_{1,n}$ (such a codebook exists for at least one $R \geqq 0$, namely for $R = 0$). Let $F_{2,n}$ be a C_n*-coder*, i.e. a one-to-one mapping from C_n into $A_{1,n}$ and put

$$F_n(x) = F_{2,n}(F_{1,n}(x))\,, \quad x \in \mathbb{R}^n\,,$$

where $F_{1,n}$ is the optimum coder of source n-messages into the codebook C_n defined earlier in this paper. Thus each source n-message $x \in \mathbb{R}^n$ is transmitted as an admissible codeword $F_n(x) \in A_{1,n}$ through the channel (18) and received as an output n-message

$\boldsymbol{b} \in A_2^n$ (cf. the diagram). The output message is random, distributed in accordance with the probability measure $P_{n,F_n(\boldsymbol{x})}$ on $(A_2^n, \mathscr{A}_2^n)$ (cf. [5]). Let

$$F_n(\mathbb{R}^n) = \{F_n(\boldsymbol{x}) \mid \boldsymbol{x} \in \mathbb{R}^n\}$$

and let G_n be the B_n-decoder for $B_n = F_n(\mathbb{R}^n) \subset A_{1,n}$ defined in [5]. In this case $G_n(\boldsymbol{b}) \in F_n(\mathbb{R}^n)$ is a random image of the channel input $F_n(\boldsymbol{x})$ at the output of the channel and

$$\mu = F_{2,n}^{-1}\, G_n(\boldsymbol{b}) \in C_n$$

is a "message" delivered to the user instead of $\boldsymbol{x} \in \mathbb{R}^n$. Thus

$$d_n(\boldsymbol{x}, F_{2,n}^{-1}\, G_n(\boldsymbol{b})) = D(\varphi_{\boldsymbol{x}}, \varphi_{F_{2,n}^{-1}G_n(\boldsymbol{b})})$$

is a random spectral distortion of $\boldsymbol{x}$ and

$$\bar{\delta}_n(\boldsymbol{x}) = \int_{A_2^n} d_n(\boldsymbol{x}, F_{2,n}^{-1}\, G_n(\boldsymbol{b}))\, \mathrm{d}P_{n,F_n(\boldsymbol{x})}(\boldsymbol{b})$$

is an *expected distortion* of $\boldsymbol{x}$, corresponding to the codebook, C_n, C_n-coder $F_{2,n}$ and $F_n(\mathbb{R}^n)$-decoder G_n. The average expected distortion

$$\bar{\delta}_n(C_n, F_{2,n}, G_n) = \int_{A_1^n} \bar{\delta}_n(x)\, \mathrm{d}\nu_n(x)$$

is a relevant spectral-distance-based distortion of the n-source $(\mathbb{R}^n, \mathscr{A}^n, \nu_n)$ for the user. This distortion is called simply *output distortion* of the source under consideration (at the output of the channel under consideration). It is easy to see that the following identity holds

$$(19) \qquad \bar{\delta}_n(C_n, F_{2,n}, G_n) = \sum_{\mu,\tilde{\mu} \in C_n} P_{n,F_{2,n}(\tilde{\mu})}\, (G_n^{-1}(F_{2,n}(\mu))) \int_{F_{1,n}^{-1}(\tilde{\mu})} D(\varphi_{\boldsymbol{x}}, \varphi_\mu)\, \mathrm{d}\nu_n(x)\,.$$

Theorem 1. Let us consider a stationary ergodic source $(\mathbb{R}^\infty, \mathscr{A}^\infty, \nu)$ with a regular distortion-rate function $\delta_\nu(R)$, $R \geqq 0$, defined by means of a spectral distance D_f, and a nonanticipating communication channel (17) with a capacity $C > 0$. Then for every $\varepsilon > 0$ there exists a natural number n_0 such that for all $n > n_0$ there exist a codebook $C_n \subset \mathfrak{A}$, a C_n-coder $F_{2,n}$ and an $F_n(C_n)$-decoder G_n such that the output distortion satisfies the inequality

$$\bar{\delta}_n(C_n, F_{2,n}, G_n) \leqq \delta_\nu(C-) + \varepsilon\,, \quad \text{where} \quad \delta_\nu(C-) = \lim_{R \uparrow C} \delta_\nu(R)\,.$$

Proof. (I) By definition, $\delta_\nu(R)$ is nonincreasing in the domain $R \geqq 0$. Therefore, for given C and ε, there exists $0 < R < C$ such that

$$\delta_\nu(C-) \leqq \delta_\nu(R) < \delta_\nu(C-) + \tfrac{1}{3}\varepsilon\,.$$

(II) It follows from the regularity property (16) that there exists natural n_1 such that, for all $n > n_1$, one can find a codebook C_n in $\mathbb{C}_n(R)$ with the property

$$d_{n,\nu}(C_n) < \delta_\nu(R) + \tfrac{1}{3}\varepsilon$$

i.e., in view of (I), with the property

$$d_{n,\nu}(C_n) < \delta_\nu(C-) + \tfrac{2}{3}\varepsilon\,.$$

(III) Let γ be the number figuring in the regularity property (17). It follows from the definition of capacity C and from the inequality $0 < R < C$ that there exists $n_0 > n_1$ such that, for every $n > n_0$, there exists a C_n-coder $F_{2,n}$ and an $F_{2,n}(C_n)$-decoder G_n for which

$$P_{n,F_{2,n}(\tilde{\mu})}(G_n^{-1}(F_{2,n}(\tilde{\mu}))) > 1 - \frac{\varepsilon}{3(\gamma + \delta_v(C -) + \varepsilon)}, \quad \tilde{\mu} \in C_n .$$

Therefore, for every $\tilde{\mu} \in C_n$,

$$\sum_{\substack{\mu \in C_n \\ \mu \neq \tilde{\mu}}} P_{n,F_{2,n}(\tilde{\mu})}(G_n^{-1}(F_{2,n}(\mu))) < \frac{\varepsilon}{3(\gamma + \delta_v(C -) + \varepsilon)} .$$

(IV) It follows from (19)

$$\tilde{\delta}_n(C_n, F_{2,n}, G_n) \leqq \mathscr{E}(1) + \mathscr{E}(2)$$

where

$$\mathscr{E}(1) = \sum_{\mu \in C_n} \int_{F_{1,n}^{-1}(\mu)} D(\varphi_x, \varphi_\mu) \, dv_n(x) =$$

$$= \int_{A^n} d_n(x, C_n) \, dv_n(x) = d_{n,v}(C_n) < \delta_v(C -) + \tfrac{2}{3}\varepsilon \quad \text{(cf. (II))}$$

and

$$\mathscr{E}(2) = \sum_{\substack{\mu, \tilde{\mu} \in C_n \\ \mu \neq \tilde{\mu}}} P_{n,F_{2,n}(\tilde{\mu})}(G_n^{-1}(F_{2,n}(\mu))) \int_{F_{1,n}^{-1}(\tilde{\mu})} D(\varphi_x, \varphi_\mu) \, dv_n(x) =$$

$$= \sum_{\tilde{\mu} \in C_n} \Big(\sum_{\substack{\mu \in C_n \\ \mu \neq \tilde{\mu}}} P_{n,F_{2,n}(\tilde{\mu})}(G_n^{-1}(F_{2,n}(\mu))) \int_{F_{1,n}^{-1}(\tilde{\mu})} D(\varphi_x, \varphi_\mu) \, dv_n(x) \Big) .$$

Since for every $\mu, \tilde{\mu} \in C_n$

$$\int_{F_{1,n}^{-1}(\tilde{\mu})} D(\varphi_x, \varphi_\mu) \, dv_n(x) \leqq \int_{F_{1,n}^{-1}(\tilde{\mu})} D(\varphi_x, \varphi_{\tilde{\mu}}) \, dv_n(x) + \int_{F_{1,n}^{-1}(\tilde{\mu})} D(\varphi_{\tilde{\mu}}, \varphi_\mu) \, dv_n(x) \leqq$$

$$\leqq \int_{F_{1,n}^{-1}(\tilde{\mu})} D(\varphi_x, \varphi_{\tilde{\mu}}) \, dv_n(x) + \gamma v_n(F_{1,n}^{-1}(\tilde{\mu})) ,$$

it follows from (III) and from the definition of $F_{1,n}$

$$\mathscr{E}(2) \leqq \frac{\varepsilon}{3(\gamma + \delta_v(C -) + \varepsilon)} \Big[\sum_{\tilde{\mu} \in C_n} \int_{F_{1,n}^{-1}(\tilde{\mu})} D(\varphi_x, \varphi_{\tilde{\mu}}) \, dv_n(x) + \gamma \Big] =$$

$$= \frac{\varepsilon}{3(\gamma + \delta_v(C -) + \varepsilon)} \left[d_{n,v}(C_n) + \gamma \right] .$$

Hence, by (II),

$$\mathscr{E}(2) < \tfrac{1}{3}\varepsilon .$$

This, together with the above established upper bound to $\mathscr{E}(1)$ implies

$$\mathscr{E}(1) + \mathscr{E}(2) < \delta_v(C -) + \varepsilon . \qquad \square$$

3. EVALUATION OF PARAMETERS OF AR-MODELS

In the rest of this paper we restrict ourselves to spectral distances $D: \Phi \times \Phi \mapsto \mapsto [0, \infty]$ (cf. (5), (66) with the following property: for every $\varphi_\mu \in \Phi$ there exists $\mu_* = (\sigma^2, \boldsymbol{a}) \in (0, \infty) \times \mathbb{A}_m$ such that

$$D(\varphi_\mu, \varphi_{\mu_*}) \leqq D(\varphi_\mu, \varphi_{\mu_0}), \quad \mu_0 \in (0, \infty) \times \mathbb{A}_m,$$

iff

(20) $$r_{\mu,k} = r_{\mu_*,k}, \quad k = 0, \ldots, m.$$

It is easy to see that, for every $1 \leqq \alpha < \infty$ and $m \geqq 1$,

$$D(\varphi, \psi) = \left(\sum_{k=-m}^{m} |r_{\varphi,k} - r_{\psi,k}|^\alpha \right)^{1/\alpha}$$

is a spectral distance possessing the desired property. It can be shown (cf. Vajda [6]), that $D_f(\varphi, \psi)$ defined by (9) for

$$f(x) = -\ln x + x - 1$$

(which is not satisfying (11)) is a spectral distance possessing the desired property too, uniformly for all $m \geqq 1$. A great advantage of distances with this property consists in that the minimum spectral distance AR-models of order m for signals $\boldsymbol{x} \in \mathbb{R}^n$, $n > m$, depen only on the first $m + 1$ covariances $r_{x,0}, \ldots, r_{x,m}$ defined by (26), (27) in [5]. Moreover, it follows from (20) and from what follows (24) in [5] that the parameters $\mu_* = (\sigma^2, \boldsymbol{a})$ of the minimum distance AR-models are defined uniquely by these covariances. The rest of this paper is devoted to mathematical methods of obtaining the corresponding minimum distance parameters $\mu_* = = (\sigma^2, \boldsymbol{a}) \in (0, \infty) \times \mathbb{A}_m$ defined by the condition (20) for arbitrary wide-sense stationary sources $\mu \in \mathfrak{M}$. For simplicity we write simply r instead of r_μ. We prefer terminology of stochastic processes considered before Example 5 in [5] for the methods below are close to the problems of prediction of wide-sense stationary processes.

Let us consider a probability space $(\Omega, \mathcal{S}, P)$ and let $\mathcal{H}$ be the set of all real valued random variables X defined on this space with $\mathsf{E}X = 0$, $\mathsf{E}X^2 < \infty$. The random variables $X, Y \in \mathcal{H}$ with $\mathsf{E}(X - Y)^2 = 0$ are considered identical. Let us define in the usual manner addition and multiplication by a real number in $\mathcal{H}$ and let $\langle \cdot, \cdot \rangle$ be a scalar product defined on $\mathcal{H}$ by

$$\langle X, Y \rangle = \mathsf{E}(XY).$$

Then, as well known, $\mathcal{H}$ is a Hilbert space with the norm

$$\|X\| = \langle X, X \rangle^{1/2} = (\mathsf{E}X^2)^{1/2}, \quad X \in \mathcal{H}.$$

A random process $\mathcal{X} = (X_n \mid n = 0, \pm 1, \ldots)$ in the Hilbert space $\mathcal{H}$ is wide-sense stationary iff there is a function $\boldsymbol{r} = (r_n \mid n = 0, \pm 1, \ldots)$, called covariance function of $\mathcal{X}$, such that, for all $k, j = 0, \pm 1, \ldots.$

$$\langle X_k, X_k \rangle = r_0, \quad \langle X_k, X_j \rangle = r_{k-j} = r_{j-k}$$

(cf. the definitions in Example 1 in [5]). We restrict ourselves to wide-sense stationary processes $\mathscr{X}$ from $\mathscr{H}$ with spectral densities $\varphi(\omega)$ positive everywhere on $[-\pi/2, \pi/2]$. As shown at the end of Example 1 in [5], this implies that all $(n \times n)$-matrices $[r_{j-k}]$ are strictly positively definite).

By $\mathscr{H}_{n,m}$, $n \geqq m \geqq 1$, we denote the linear span of $X_{n-m}, \ldots, X_{n-1}$ in $\mathscr{H}$ and by $\hat{X}^+_{n|m}$ the projection of X_n on $\mathscr{H}_{n,m}$. As well known from the theory of Hilbert spaces, this projection is uniquely characterized by the condition

$$\|X_n - \hat{X}^+_{n|m}\| = \min_{Y \in \mathscr{H}_{n,m}} \|X_n - Y\| . \tag{21}$$

It follows from the definition of $\mathscr{H}_{n,m}$ that $\hat{X}^+_{n|m}$ can be written in the form

$$\hat{X}^+_{n|m} = -\sum_{j=1}^{m} a_j^{(m)} X_{n-j} . \tag{22}$$

By (21), $\hat{X}^+_{n|m}$ is the best, in the sense of norm, of all estimates

$$\sum_{j=1}^{m} c_j X_{n-j}$$

of X_n. Interpreting the subscript as a time, we call $\hat{X}^+_{n|m}$ an *optimum linear one-step prediction* of the sequence $X_{n-m}, \ldots, X_{n-1}$ (in the sequel simply prediction, or prediction of order m).

By the definition of orthogonal projection, the error

$$\mathscr{E}^+_{n|m} = X_n - \hat{X}^+_{n|m} = X_n + \sum_{j=1}^{m} a_j^{(m)} X_{n-j} \tag{23}$$

must be perpendicular to $\mathscr{H}_{n|m}$, i.e. to the basis $\{X_{n-m}, \ldots, X_{n-1}\}$. This leads to the equations

$$\langle \mathscr{E}^+_{n|m}, X_{n-k} \rangle = 0 , \quad k = 1, \ldots, m .$$

equivalent to

$$r_k + \sum_{j=1}^{m} a_j^{(m)} r_{k-j} = 0 , \quad k = 1, \ldots, m . \tag{24}$$

The norm of the error is given by

$$\|\mathscr{E}^+_{n|m}\|^2 = \langle X_n + \sum_{j=1}^{m} a_j^{(m)} X_{n-j}, X_n \rangle = r_0 + \sum_{j=1}^{m} a_j^{(m)} r_j . \tag{25}$$

Analogically as above, the orthogonal projection $\hat{X}^-_{n-m-1|m}$ of X_{n-m-1} on $\mathscr{H}_{n,m}$ (a backward prediction of order m) is given by

$$\hat{X}^-_{n-m-1|m} = -\sum_{j=1}^{m} b_j^{(m)} X_{n-j} \tag{26}$$

where the coefficients $b_j^{(m)}$ may be expressed in term of a solution of (24) as follows

$$b_j^{(m)} = a_{m+1-j}^{(m)} , \quad j = 1, \ldots, m . \tag{27}$$

It follows from here for the backward prediction error

(28) $$\|\mathscr{E}^-_{n-m-1|m}\|^2 = \langle X_{n-m-1} + \sum_{j=1}^{m} a_j^{(m)} X_{n-m-1+j}, X_{n-m-1}\rangle = \|\mathscr{E}^+_{n|m}\|^2 .$$

The optimum forward as well as backward prediction of order m is thus described by a solution $a_1^{(m)}, \ldots. a_m^{(m)}$ of the equations (24). Since, by assumption, the $(m + 1) \times \times (m + 1)$-matrix $[r_{k-j}]$ of these equations is positively definite, the solution $a_1^{(m)}, \ldots, a_m^{(m)}$ is unique. We shall describe a recursive algorithm for evaluation of this solution first published by V. Levinson in 1948 and later modified by several authors (cf. Markel and Gray [3]).

Let us replace the basis $X_{n-m}, \ldots, X_{n-1}$ in $\mathscr{H}_{n,m}$ by an orthogonal basis $V_{n-m}, \ldots \ldots, V_{n-1}$ using the well-known Gram-Schmidt orthogonalization. We get

$$V_{n-1} = X_{n-1}$$

$$V_{n-k} = X_{n-k} - \sum_{j=1}^{k-1} c_j^{(k)} V_{n-j}, \quad k = 2, 3, \ldots,$$

where

$$c_j^{(k)} = - \frac{\langle X_{n-k}, V_{n-j}\rangle}{\|V_{n-j}\|^2}$$

(since $\mathscr{X}$ is stationary, $c_j^{(k)}$ is independent of n).

Proposition 4. It holds for every $k > 1$

$$V_{n-k} = \mathscr{E}^-_{n-k|k-1}$$

i.e. $\mathscr{E}^-_{n-k|k-1}$, $k = 1, \ldots, m$, with $\mathscr{E}^-_{n-1|0} = X_{n-1}$, is an orthogonal basis in $\mathscr{H}_{n,m}$.

Proof. The expression

$$\sum_{j=1}^{k-1} c_j^{(k)} V_{n-j}$$

is, by definition, an orthogonal projection of X_{n-k} on $\mathscr{H}_{n,k}$ so that the stated equality holds. The rest is clear. □

It follows from Proposition 4 that $\hat{X}^+_{n|m}$ as a projection of X_n on $\mathscr{H}_{n|m}$ can be represented as follows

(29) $$\hat{X}^+_{n|m} = \sum_{j=1}^{m} k_j V_{n-j} .$$

Proposition 5. If $\mathscr{E}^{\pm}_{i|0} = X_i$ then it holds for every $m \geqq 1$

$$\hat{X}^+_{n|m} = \hat{X}^+_{n|m-1} + k_m \mathscr{E}^-_{n-m|m-1}$$

$$\mathscr{E}^+_{n|m} = \mathscr{E}^+_{n|m-1} - k_m \mathscr{E}^-_{n-m|m-1}$$

where k_m satisfies the equation

(30) $$\langle \mathscr{E}^+_{n|m-1}, X_{n-m}\rangle - k_m \langle \mathscr{E}^-_{n-m|m-1}, X_{n-m}\rangle = 0 .$$

Proof. If $m = 1$ then assertion follows directly from the definition and assumptions. Let $m > 1$. It follows from (29)

$$\hat{X}^+_{n|m} = \sum_{j=1}^{m-1} k_j V_{n-j} + k_m V_{n-m} = \hat{X}^+_{n|m-1} + k_m V_{n-m}$$

where, by Proposition 4, $V_{n-m} = \mathscr{E}^-_{n-m|m-1}$ so that the first identity holds. The second identity follows from the first one and from (23). The equation for k_m follows from the second identity and from the last identity preceding (24). □

Analogically as $V_{n-1}, \ldots, V_{n-m}$, the sequence

$$W_1 = X_{n-m}\,, \quad W_2 = \mathscr{E}^+_{n-m+1|1}, \ldots, W_m = \mathscr{E}^+_{n-1|m-1}$$

is an orthogonal basis of $\mathscr{H}_{n,m}$ too and the backward prediction can be written as follows

$$\hat{X}^-_{n-m-1|m} = \sum_{j=1}^{m} l_j W_j\,.$$

The next statement is then an analogy to Proposition 5.

Proposition 6. If $\mathscr{E}^\pm_{i|0} = X_i$ then it holds for every $m \geqq 1$

$$\hat{X}^-_{n-m-1|m} = \hat{X}^-_{n-m-1|m-1} + \ell_m \mathscr{E}^+_{n-1|m-1}$$

$$\mathscr{E}^-_{n-m-1|m} = \mathscr{E}^-_{n-m-1|m-1} - \ell_m \mathscr{E}^+_{n-1|m-1}$$

and

(31) $$\ell_m = k_m\,.$$

Proof. The first two identities follow in a similar manner as the first two identities of Proposition 5. The same applies to the following analogy of (30)

(32) $$\langle \mathscr{E}^-_{n-m-1|m-1}, X_{n-1} \rangle - \ell \langle \mathscr{E}^+_{n-1|m-1}, X_{n-1} \rangle = 0\,.$$

Using the obvious identities (hereafter we assume that the sum is zero if the summation domain is empty)

$$\mathscr{E}^-_{n-m-1|m-1} = X_{n-m-1} + \sum_{j=1}^{m-1} a_j^{(m-1)} X_{n-m-1+j}$$

$$\mathscr{E}^+_{n-1|m-1} = X_{n-1} + \sum_{j=1}^{m-1} a_j^{(m-1)} X_{n-1-j}$$

valid for all $m \geqq 1$ one obtains

$$\langle \mathscr{E}^-_{n-m-1|m-1}, X_{n-1} \rangle = r_m + \sum_{j=1}^{m-1} a_j^{(m-1)} r_{m-j} = \langle \mathscr{E}^+_{n|m-1}, X_{n-m} \rangle$$

and

$$\langle \mathscr{E}^+_{n-1|m-1}, X_{n-1} \rangle = r_0 + \sum_{j=1}^{m-1} a_j^{(m-1)} r_j = \langle \mathscr{E}^-_{n-m|m-1}, X_{n-m} \rangle\,.$$

It follows from here and from (30), (32) that (31) holds and that, moreover, k_m satisfies the equation

(33) $$r_0 + \sum_{j=1}^{m-1} a_j^{(m-1)} r_j - k_m\Big(r_m + \sum_{j=1}^{m-1} a_j^{(m-1)} r_{m-j}\Big) = 0\,.$$ □

Theorem 2. (Levinson recursive formula). For every $m \geqq 1$ there is a unique solution k_m of the equation (33) and it satisfies the relations

$$a_m^{(m)} = -k_m \qquad \text{if} \quad m \geqq 1 .$$

$$a_j^{(m)} = a_j^{(m-1)} - k_m a_{m-j}^{(m-1)} , \quad j = 1, \ldots, m-1 , \quad \text{if} \quad m > 1 .$$

Proof. (I) The solution of (33) may not be unique only if

$$r_m + \sum_{j=1}^{m-1} a_j^{(m-1)} r_{m-j} = 0$$

in which case it must be

$$r_0 + \sum_{j=1}^{m-1} a_j^{(m-1)} r_j = 0 .$$

These two equations together with the equations (24) for $a_1^{(m-1)}, \ldots, a_{m-1}^{(m-1)}$ contradict the assumption that the $(m+1) \times (m+1)$-matrix $[r_{k-j}]$ is positively definite. (II) It holds for every $m \geqq 1$

$$\hat{X}^+_{n|m} = -\sum_{i=1}^{m} a_j^{(m)} X_{n-j}$$

$$\hat{X}^+_{n|m-1} = \begin{cases} a_1^{(1)} X_n & \text{if} \quad m = 1 \\ -\sum_{j=1}^{m-1} a_j^{(m-1)} X_{n-j} & \text{if} \quad m > 1 . \end{cases}$$

Inserting these expressions into the first identity of Proposition 5 and comparing the left and right sides one obtains the desired relations.

The Levinson recursive formula yields the following *Levinson algorithm* for evaluation of the solution $a_1^{(m)}, \ldots, a_m^{(m)}$ of equation (24) in $m \geqq 1$ steps:

Step 1: Compute $k_1 = r_1/r_0$ and put $a_1^{(1)} = -k_1$.

Step $s > 1$: Compute

$$k_s = \frac{r_0 + \sum_{j=1}^{s-1} a_j^{(s-1)} r_j}{r_s + \sum_{j=1}^{s-1} a_j^{(s-1)} r_{s-j}}$$

and put

$$a_j^{(s)} = a_j^{(s-1)} - k_s a_{s-j}^{(s-1)} , \quad j = 1, \ldots, s-1 , \quad a_s^{(s)} = -k_s .$$

End if $s = m$.

The parameters $k_1, \ldots, k_m$ defined by equations (33) and satisfying the identities

(34) $$k_s = -a_s^{(s)} , \quad s = 1, \ldots, m ,$$

are called *coefficients of reflexion*. We shall present a direct method of evaluation of these coefficients, not requiring to evaluate simultaneously the vectors $a_1^{(s)}, \ldots$ $\ldots, a_s^{(s)}$ for $s = 1, \ldots, m$. The authors of this method are Le Roux and Gueguen [2].

The method is based on the following two propositions.

Proposition 7. If $\mathscr{E}^+_{n|0} = X_n$ then it holds for every $n \geqq m \geqq 1$

$$\|\mathscr{E}^+_{n|m}\|^2 = (1 - k_m^2)\,\|\mathscr{E}^+_{n|m-1}\|^2 .$$

Proof. By the repeated use of the second identity in Proposition 5 one obtains

$$\mathscr{E}^+_{n|m} = \mathscr{E}^+_{n|m-1} - k_m\mathscr{E}^-_{n-m|m-1} =$$
$$= \mathscr{E}^+_{n|m-2} - k_{m-1}\mathscr{E}^-_{n-m+1|m-2} - k_m\mathscr{E}^-_{n-m|m-1} = \ldots = X_n - k_1\mathscr{E}^-_{n-1|0} - \ldots$$
$$\ldots - k_m\mathscr{E}^-_{n-m|m-1}$$

Since, as proved above, $\mathscr{E}^-_{n-1|0}, \ldots, \mathscr{E}^-_{n-m|m-1}$ are mutually orthogonal, it holds

$$\|\mathscr{E}^+_{n|m} - X_n\|^2 = \sum_{j=1}^{m} k_j^2\|\mathscr{E}^-_{n-j|j-1}\|^2 .$$

Using the obvious identity

$$\langle\mathscr{E}^+_{n|m}, X_n\rangle = \langle\mathscr{E}^+_{n|m}, \mathscr{E}^+_{n|m}\rangle = \|\mathscr{E}^+_{n|m}\|^2$$

one obtains

$$\|\mathscr{E}^+_{n|m} - X_n\|^2 = \|X_n\|^2 - \|\mathscr{E}^+_{n|m}\|^2 ,$$

i.e.

$$\|X_n\|^2 - \|\mathscr{E}^+_{n|m}\|^2 = \sum_{j=1}^{m} k_j^2\|\mathscr{E}^-_{n-j|j-1}\|^2 .$$

Replacing m by $m - 1$ one obtains

$$\|X_n\|^2 - \|\mathscr{E}^+_{n|m-1}\|^2 = \sum_{j=1}^{m-1} k_j^2\|\mathscr{E}^-_{n-j|j-1}\|^2 .$$

Substracting this equality from the preceding one and using (28) one obtains the desired relation. □

Proposition 8. For every $n \geqq m \geqq 1$ it holds

$$\langle\mathscr{E}^+_{n|m}, X_{n-j}\rangle = \begin{cases} 0 & \text{if } j = 1, \ldots, m \\ \langle\mathscr{E}^+_{n|m-1}, X_{n-j}\rangle - k_m\langle\mathscr{E}^+_{n|m-1}, X_{n-m+j}\rangle & \text{if } j > m . \end{cases}$$

Proof. The assertion for $j = 1, \ldots, m$ follows from the definition of $\mathscr{E}^+_{n|m}$. Inserting into the left hand from the second identity of Proposition 5 one obtains

$$\langle\mathscr{E}^+_{n|m}, X_{n-j}\rangle = \langle\mathscr{E}^+_{n|m}, X_{n-j}\rangle - k_m\langle\mathscr{E}^-_{n-m|m-1}, X_{n-j}\rangle . \tag{35}$$

The second right hand term can be rewritten as follows

$$\langle\mathscr{E}^-_{n-m|m-1}, X_{n-j}\rangle = \langle X_{n-m} + \sum_{i=1}^{m-1} a_i^{(m-1)}X_{n-m+i}, X_{n-j}\rangle =$$
$$= r_{m-j} + \sum_{i=1}^{m-1} a_i^{(m-1)}r_{m-j-i} = \langle X_n + \sum_{i=1}^{m-1} a_i^{(m-1)}X_{n-i}, X_{n-m+j}\rangle =$$
$$= \langle\mathscr{E}^+_{n|m-1}, X_{n-m+j}\rangle .$$

Substituting this into (35) we obtain the desired result. □

Let us denote for brevity

$$e_{s|j} = \langle \mathscr{E}^+_{n|s-1}, X_{n-j} \rangle\,, \quad s \geqq 1\,,$$

where $\mathscr{E}^+_{n|0} = X_n$. Clearly, $e_{s|0} = \|\mathscr{E}^+_{n|s-1}\|^2$. The following *Le Roux-Gueguen algorithm* for evaluation of reflection coefficients, $k_1, \ldots, k_m$ in $m \geqq 1$ steps follows directly from the definition of $e_{s|j}$ and from Propositions 7 and 8.

Step $s \geqq 1$: Put for every $j = s, \ldots, m$

$$e_{s|j} = \begin{cases} r_j & \text{if} \quad s = 1 \\ e_{s-1|j} - k_{s-1} e_{s-1|j+1-s} & \text{if} \quad s > 1 \end{cases}$$

$$e_{s|0} = \begin{cases} r_0 & \text{if} \quad s = 1 \\ (1 - k_{s-1})^2 \, e_{s-1|0} & \text{if} \quad s > 1 \end{cases}$$

and compute

$$k_s = \frac{e_{s|s}}{e_{s|0}}\,.$$

End if $s = m$.

The next theorem explains how are the above considered results related to the problem formulated at the beginning of the present part of this paper.

Theorem 3. Let $(X_n \mid n = 0, \pm 1, \ldots)$ be a wide-sense stationary process with a positively definite covariance function $(r_r \mid n = 0, \pm 1, \ldots)$ and let us consider the optimum predictor of order $m \geqq 1$ for this process, i.e. the solution $(a_1^{(m)}, \ldots, a_m^{(m)})$ of equations (24). Then the following assertions holds:

(i) $\boldsymbol{a}^{(m)} = (a_1^{(m)}, \ldots, a_m^{(m)})$ belongs to the set $\mathbb{A}_m$ of all vectors $\boldsymbol{a} = (a_1, \ldots, a_m) \in \mathbb{R}^m$ such that the complex polynomials

$$\boldsymbol{a}(\lambda) = 1 + a_1 \lambda^{-1} + \ldots + a_m \lambda^{-m}$$

have all roots inside the unit circle and the quantity

$$\sigma_m^2 = r_0 + \sum_{j=1}^{m} a_j^{(m)} r_j \tag{36}$$

is positive.

(ii) The AR-process with parameters $\mu = (\sigma_m^2, \boldsymbol{a}^{(m)})$ defined in (i) is the unique AR-process of order m with covariances $(r_{\mu,0}, \ldots, r_{\mu,m})$ equal to $(r_0, \ldots, r_m)$.

Proof. Theorem of Schur-Cohn (see e.g. Prouza [4]) says that for every $m \geqq 1$ and $\boldsymbol{a} = (a_1, \ldots, a_m) \in \mathbb{R}^m$ the polynomial

$$Q(\lambda) = (\lambda^m + a_1 \lambda^{m-1} + \ldots + a_m)$$

has all roots inside the unit circle iff it holds $|a_m)| < 1$ and the polynomial

$$\lambda^{-1}[Q(\lambda) - a_m \, Q(\lambda^{-1}) \, \lambda^m]$$

of degree $m - 1$ has all roots inside the unit circle. We shall prove the present Theorem in two steps.

(I) Let $1 \leqq s \leqq m$. It holds

$$\langle \mathscr{E}^-_{n-s|s-1}, X_{n-s} \rangle = \|\mathscr{E}^-_{n-s|s-1}\|^2 = \|\mathscr{E}^+_{n|s-1}\|^2$$

(cf. the stationarity assumption). Further it follows from the Schwartz inequality that the expression $|\langle \mathscr{E}^+_{n|s}, X_{n-s} \rangle|$ attains maximum equal to $\|\mathscr{E}^+_{n|s}\|^2$ for $X_{n-s} = \mathscr{E}^+_{n|s}$. Since the last equality contradicts the positive definiteness assumption, it holds

$$\langle \mathscr{E}^+_{n|s-1}, X^{n-s} \rangle < \|\mathscr{E}^+_{n|s-1}\|^2 .$$

Therefore it follows from (30), (34)

$$|a_s^{(s)}| = |k_s| < 1 , \quad s = 1, \ldots, m . \tag{37}$$

The assertion (i) of Theorem 3 holds if $m = 1$. Indeed, by the Levinson algorithm,

$$a_1^{(1)} = -r_1/r_0 , \quad \sigma_1^2 = r_0 - r_1^2/r_0$$

and the rest is clear from the positive definiteness of the matrix

$$\begin{bmatrix} r_0 & r_1 \\ r_1 & r_0 \end{bmatrix} .$$

It thus suffices to prove that (i) holds for $m > 1$ under the assumptions that it holds for $m - 1 \geqq 1$. Let us denote for every $s \geqq 1$

$$Q_s(\lambda) = (\lambda^s + a_1^{(s)}\lambda^{s-1} + \ldots + a_s^{(s)}) .$$

It holds

$$\boldsymbol{a}^{(s)}(\lambda) = \lambda^{-s} Q_s(\lambda) .$$

Thus, by assumption, $Q_{m-1}(\lambda)$ has all roots inside the unit circle and we need to prove the same for $Q_m(\lambda)$. It follows from the second relation in Proposition 5

$$\boldsymbol{a}^{(m)}(\lambda) = \boldsymbol{a}^{(m-1)}(\lambda) - k_m \lambda^{-m} \boldsymbol{a}^{(m-1)}(1/\lambda)$$

i.e.

$$Q_m(\lambda) = \lambda Q_{m-1}(\lambda) - k_m \lambda^{m-1} Q_{m-1}(1/\lambda) .$$

Inserting $1/\lambda$ at the place of λ we get

$$Q_m(1/\lambda) = \lambda^{-1} Q_{m-1}(1/\lambda) - k_m \lambda^{-m+1} Q_{m-1}(\lambda) .$$

Multiplying this equality by $k_m \lambda^m$ and adding it to the former equality we obtain

$$(1 - (a_m^{(m)})^2)\, Q_{m-1}(\lambda) = \lambda^{-1}[Q_m(\lambda) - \lambda^m a_m^{(m)} Q_m(1/\lambda)]$$

It follows from here, from the assumption above and from (37) that all assumptions of the Schurr-Cohn theorem hold. Consequently, all roots of $Q_m(\lambda)$ are inside the unit circle. The assumption $\sigma_m^2 = 0$, analogically as in the proof of Theorem 2, contradicts the assumption that r is positively definite. This completes the proof of assertion (i).

(II) Let us consider an AR-process with parameters $\mu = (\sigma_m^2, \boldsymbol{a}_m^{(m)})$. By the Levinson recursive formula, $\boldsymbol{a}^{(m)}$ uniquely determines the set of vectors $\{\boldsymbol{a}^{(m-1)}, \boldsymbol{a}^{(m-2)}, \ldots, \boldsymbol{a}_1^{(1)}\}$ as well as the set $\{k_m, k_{m-1}, \ldots, k_1\}$ of reflection coefficients. It follows from here and from the Levinson algorithm that $\boldsymbol{a}^{(m)}$ uniquely determines the ratios

$$r_1/r_0, \ldots, r_m/r_0 \tag{38}$$

of coefficients of the equations (24) (i.e., up to the multiplicative factor r_0, there is a one-to-one correspondence between coefficients $r_1, \ldots, r_m$ and solutions $\boldsymbol{a}^{(m)}$ of those equations). The factor $r_0 > 0$ is uniquely determined by the ratios (38) and $\sigma_m^2 > 0$ from the equation (36). Hence there is a one-to-one relation between parameters $\mu = (\sigma_m^2, \boldsymbol{a}^{(m)})$ of the AR-processes and coefficients $r_0, r_1, \ldots, r_m$ of equations (24) and (36) used to define these parameters. Since the covariances $r_{\mu,0}, r_{\mu,1}, \ldots, r_{\mu,m}$ of the AR-process with parameters $\mu = (\sigma_m^2, \boldsymbol{a}^{(m)})$ are satisfying formaly the same equations as (24) and (36), i.e.

$$\sigma_m^2 = r_{\mu,0} + \sum_{j=1}^{m} a_j^{(m)} r_{\mu,j} \quad \text{(cf. (23) in [5])}$$

$$r_{\mu,k} + \sum_{j=1}^{m} a_j^{(m)} r_{\mu,k-j} = 0\,, \quad k = 1, \ldots, m \quad \text{(cf. (24) in [5])}\,,$$

it holds

$$(r_{\mu,0}, \ldots, r_{\mu,m}) = (r_0, \ldots, r_m)\,.$$

The uniqueness of the AR-process with this property follows from the one-to-one relation between parameters μ and covariances $(r_\mu,0, \ldots, r_{\mu,m})$ of the AR-processes with parameters established before Example 5 in [5]. □

It follows from the two algorithms proved above and from Theorem 3 that the following assertion holds.

Corollary. For every $m \geqq 1$ there is a one-to-one relation between any two of the following three vectors:

(i) positively definite covariances $(r_0, \ldots, r_m)$,
(ii) reflection coefficients $(k_1, \ldots, k_m) \in ((-1, 0) \cup (0, 1))^m$ and $\sigma_m^2 \in (0, \infty)$,
(iii) autoregressive parameters $(\sigma_m^2, \boldsymbol{a}^{(m)}) \in (0, \infty) \times \mathbb{A}_m$.

The relation between covariances and reflection coefficients is recursively described by the Le Roux-Gueguen algorithm (forward and backward), and the simultaneous recursive evaluation of reflexion coefficients and autoregressive parameters for given covariances is described by the Levinson algorithm. Simultaneous recursive evaluation of reflexion coefficients and covariances for given autoregressive parameters $(\sigma_m^2, \boldsymbol{a}^{(m)})$ in $2m$ steps is described by the following algorithm:

Step $s > 0$: Put $k_{m-s} = -a_s^{(m-s)}$ and compute

$$a_j^{(m-s-1)} = \frac{a_j^{(m-s)} + k_{m-s} a_{m-s-j}^{(m-s)}}{1 - k_{m-s}^2}\,, \quad j = 1, \ldots, m - s - 1\,.$$

If $s = m - 2$ *then step* $t = 1$: Put $\varrho_1 = k_1$.

Step $t > 1$: Compute

$$\varrho_t = -\sum_{j=1}^{t-1} a_j^{(t-1)} \varrho_{t-j} + \frac{1 + \sum_{j=1}^{t-1} a_j^{(t-1)} \varrho_j}{k_t},$$

If $t = m$ *then step* $t = m + 1$: Compute

$$r_0 = \frac{\sigma_m^2}{1 + \sum_{j=1}^{m} a_j^{(m)} \varrho_j},$$

$$r_j = r_0 \varrho_j, \quad j = 1, \ldots, m.$$

End.

(Received January 21, 1987.)

REFERENCES

[1] P. R. Halmos: Measure Theory. Van Nostrand, Princeton 1964.

[2] J. Le Roux and G. Geuguen: Fixed point computation of partial correlation coefficients. IEEE Trans. Acoust. Speech Signal Process. *ASSP-25* (1977), 257—259.

[3] J. D. Markel and A. H. Gray Jr.: Linear Prediction of Speech. Springer-Verlag, Berlin 1976.

[4] L. Prouza: Introduction to the Theory and Applications of Linear Pulse Systems (in Czech). Academia, Praha 1967.

[5] I. Vajda: Alphabet-oriented and user-oriented source coding theories. Kybernetika *23* (1987), 5, 388—406.

[6] I. Vajda: Theory of Statistical Inference and Information. Reidel Publishing Company (in print).

[7] J. Kadlec: Moving Probabilistic Identification of Autoregressive Models of Unknown Order by Means of Lattice Structures. Res. Rept. No. 1313, ÚTIA — ČSAV, Prague 1985.

RNDr. Otakar Šefl, CSc., Ing. Igor Vajda, CSc., Ústav teorie informace a automatizace ČSAV (Institute of Information Theory and Automation — Czechoslovak Academy of Sciences), Pod vodárenskou věží 4, 182 08 Praha 8. Czechoslovakia.