Kybernetika

Igor Vajda

Alphabet-oriented and user-oriented source coding theories

Kybernetika, Vol. 23 (1987), No. 5, 388--406

Persistent URL: http://dml.cz/dmlcz/125873

KYBERNETIKA — VOLUME 23 (1987), NUMBER 5

ALPHABET-ORIENTED AND USER-ORIENTED SOURCE CODING THEORIES

IGOR VAJDA

Dedicated to the Memory of Dr. Bedřich Váňa.

The aim of this, and of the follow-up paper [16], is to present the mathematics used in the technical field of speech coding. This paper remains mostly within a „classical'' source coding theory based on alphabetic distortion measures. We prove a new general joint source/channel coding theorem for these measures. By an example of two users we demonstrate that the "classical" theory, which is in the above stated sense alphabet-oriented, need not always be user-oriented. The first of these users is a statistician and the second one is a telephone owner. At the end we prove a statement which opens the possibility to build up a spectrum-oriented source coding theory, which is user-oriented for a special category of users, e.g. for the telephone owner. Such a theory is presented in [16].

1. INTRODUCTION

In information theory, source coding means coding of signals (messages) subject to a fidelity – or distortion – criterion. This coding leads to codebooks attaining, at a given information rate $R > 0$, the smallest possible average distortion $\delta_n(R)$ of original source messages of length $n = 1, 2, \ldots$ Since the distortion-rate functions $\delta_n(R)$, or the limits

$$\delta(R) = \liminf_{n\to\infty} \delta_n(R)\,,$$

are the basic operational characteristics of *source coding theory*, this theory is also known as distortion-rate theory or *rate-distortion theory*. The source coding theory is the important "half" of information theory which is complementary to the more traditional channel theory. The source coding theory had its origins already in Shannon [13], and was further developed mainly by Shanon [14], Berger [2], Gray and Davisson [8], and by some more recent papers which can be found in the review paper of Šujan [17].

Let us note that the mathematical importance of source coding theory exceeds the framework of information theory. A source coding method was used by Ornstein

[12] in his famous solution of the classical problem of H. Poincaré, namely to find out a necessary and sufficient condition for isomorphy of two Bernoulli sequences (for details about this problem we refer to Billingsley [3]). In return, the Ornstein's solution led to more concise and strong formulation of source coding theorems in information theory (in this respect we refer again to Šujan [17]).

We shall describe briefly the basic concepts of the source coding theory developed in the above cited papers, Let us consider an abstract *alphabet* $(A, \mathscr{A})$ where A is a closed nonempty subset of the real line $\mathbb{R}$ and $\mathscr{A}$ is the σ-algebra of all Borel subsets of A. Put

$$A^n = \times A(n \text{ times}) \quad \text{for} \quad n = 1, 2, \ldots, \infty$$

and

$$\mathscr{A}^n = \otimes \mathscr{A}(n \text{ times}) \quad \text{for} \quad n = 1, 2, \ldots,$$

and let $\mathscr{A}^\infty$ be the σ-algebra of subsets of A^∞ generated by the class of subsets

$$\{E \times A^\infty | E \in \mathscr{A}^n, n = 1, 2, \ldots\}.$$

By an *information source* we mean a probability space $(A^\infty, \mathscr{A}^\infty, \nu)$. The probability spaces $(A^n, \mathscr{A}^n, \nu_n)$, $n = 1, 2, \ldots$, defined by

$$\nu_n(E) = \nu(E \times A^\infty), \quad E \in \mathscr{A}^n,$$

are called *n-sources* and elements $\boldsymbol{x} \in A^n$ *n*-messages. We distinguish two particular cases, the first with $A = \{1, 2, \ldots, r\}$, which is called the *discrete source*, and the second with $A = \mathbb{R}$ and with a measure ν such that, for any finite n, ν_n is absolutely continuous with respect to the n-dimensional Lebesgue measure. In the second case we speak about the *continuous source*.

A fundamental concept of any source coding theory is a *basic distortion measure*. In the theory developed in the above cited papers this measure is simply an *alphabet distortion measure* d which is an $\mathscr{A} \otimes \mathscr{A}$-measurable function defined on $A \times A$ with values in $[0, \infty)$ such that $d(x, x) = 0$. The alphabet distortion measure is used to define an $\mathscr{A}^n \otimes \mathscr{A}^n$-measurable *distortion function* $d_n: A^n \times A^n \to [0, \infty)$, $n = 1, 2, \ldots$, by

$$d_n(\boldsymbol{x}, \boldsymbol{y}) = 1/n \sum_{k=1}^{n} d(x_k, y_k) \tag{1}$$

where $\boldsymbol{x} = (x_1, \ldots, x_n)$, $\boldsymbol{y} = (y_1, \ldots, y_n)$. For example, in a discrete source with $A = \{1, \ldots, r\}$, the Hamming alphabet distortion measure

$$d(x, y) = \delta_{x,y} = \begin{cases} 0 & \text{if} \quad x = y \\ 1 & \text{if} \quad x \neq y \end{cases}$$

defines the *Nedoma's distortion function* (see (6.1) in [11])

$$d_n(\boldsymbol{x}, \boldsymbol{y}) = 1/n \sum_{k=1}^{n} \delta_{x_k, y_k} \tag{2}$$

Analogically in a continuous source, the squared error alphabet distortion

$$d(x, y) = (x - y)^2$$

defines the *average squared error distortion function*

$$d_n(x, y) = 1/n \sum_{k=1}^{n} (x_i - y_i)^2 . \tag{3}$$

It is apparent from Šujan [17] that the restriction to distortion functions of the form of average per-letter distortion as considered in (1), which is characteristic for the source coding theory developed in the above cited papers, opens the possibility to employ in this theory powerful methods of ergodic theory and to obtain very general and very strong source coding theorems. We afford ourselves to speak about *alphabet-oriented source coding theory*, in order to distinguish this theory from alternative source coding theories motivated by examples that follow later in this paper (one such theory is presented in [16]). But the concepts as well as the main result stated below make sense not only in the alphabet-oriented source coding theory, to which we restrict ourselves in the present paper, but also in any theory dealing with $\mathscr{A}^n \otimes \mathscr{A}^n$-measurable distortion functions d_n: $A^n \times A^n \to [0, \infty)$. An example of such a theory is the classical discrete information theory which is based on the Hamming-distance distortion functions

$$d_n(x, y) = \delta_{x,y} = \begin{cases} 0 & \text{if} \quad x = y \\ 1 & \text{if} \quad x \neq y . \end{cases}$$

It is easy to prove by contradiction that this function is not satisfying (1) for any alphabet distortion measure d. On the other hand, some of the concepts that follow give up their sense e.g. when passing to the theory developed in [16].

For any finite n we consider a *codebook* C_n which is a collection of n-messages y_s, $s = 1, \ldots, \|C_n\|$, drawn from the finite set

$$\tilde{A}^n = \times \tilde{A}(n \text{ times}) \subset A^n$$

where $\tilde{A}$ is a finite subset of A containing $\|\tilde{A}\| \geqq 1$ alphabet letters and called an *available alphabet*. It is assumed that every source n-message $x \in A^n$ is encoded into a codeword from C_n, symbolically denoted by $y(x)$, which minimizes the function $d_n(x, \cdot)$, x fixed, on the codebook. Hence the quantity

$$d_n(x, C_n) = d_n(x, y(x))$$

describes the minimum distortion of the n-message x by the codebook. The function $d_n(x, C_n)$ of variable $x \in A^n$, as a minimum of finitely many $\mathscr{A}^n$-measurable functions $d_n(x, y)$, $y \in C_n$ (cf. the measurability assumption concerning d and (1)), is $\mathscr{A}^n$-measurable. The *average distortion* of the n-source $(A^n, \mathscr{A}^n, \nu_n)$ by the codebook C_n is defined by

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

On the other hand, *information rate* of the codebook C_n is given by

$$R(C_n) = (1/n) \log_2 \|C_n\| .$$

Clearly, the information rate takes on values between 0 and $\log_2 \|\tilde{A}\|$.

If

$$\mathbb{C}_n(R) = \{C_n \subset \tilde{A}^n \mid R(C_n) \leqq R\} \quad 0 \leqq R \leqq \log_2 \|\tilde{A}\|,$$

then it is clear that $\mathbb{C}_n(R)$ is nonempty for every $0 \leqq R \leqq \log_2 \|\tilde{A}\|$. For every fixed $0 \leqq R \leqq \log_2 \|\tilde{A}\|$ we are interested in the smallest average distortion of the n-source $(A^n, \mathscr{A}^n, \nu_n)$ attainable by codebooks from $\mathbb{C}_n(R)$, i.e. in the quantity

$$\delta_{n,\nu}(R) = \inf d_{n,\nu}(C_n),$$

where the infimum extends over all $C_n \in \mathbb{C}_n(R)$, and in the limit of this quantity

$$\delta_\nu(R) = \liminf_{n \to \infty} \delta_{n,\nu}(R).$$

We shall refer to this limit, considered as a function of $0 \leqq R \leqq \log_2 \|\tilde{A}\|$, as to a *distortion-rate function.* The distortion-rate function $\delta_\nu(R)$ is said ***regular*** if for every fixed $0 \leqq R \leqq \log_2 \|\tilde{A}\|$ there exist codebooks $C_n \in \mathbb{C}_n(R)$, $n = 1, 2, \ldots$, and a constant $\gamma \geqq 0$ such that

$$\delta_\nu(R) = \lim_{n \to \infty} d_{n,\nu}(C_n) \tag{4}$$

and

$$\int_{A^n} d_n(x, \Phi(x))\, \mathrm{d}\nu_n(x) \leqq \gamma \tag{5}$$

for every $\mathscr{A}^n$-measurable mapping $\Phi: A^n \to C_n$.

From the point of view of a source $(A^\infty, \mathscr{A}^\infty, \nu)$, the regularity of a distortion-rate function $\delta_\nu(R)$ is a property depending solely on the available alphabet $\tilde{A} \subset A$ and on the alphabet distortion measure d, which are both assumed to be constants of our source coding model and, as such, they are not explicitly denoted.

We shall formulate sufficient conditions for regularity of the distortion-rate function $\delta_\nu(R)$. First we define some properties of the source $(A^\infty, \mathscr{A}^\infty, \nu)$. Let for every $E \in \mathscr{A}^\infty$

$$E^* = \{(z_2, z_3, \ldots) \in A^\infty \mid (z_1, z_2, \ldots) \in E\}. \tag{6}$$

The set E^* belongs to $\mathscr{A}^\infty$ too – to this end it suffices to consider first the sets E of the form $E_n \times A^\infty$ where $E_n \in \mathscr{A}^n$ and then to take into account the definition of $\mathscr{A}^\infty$. The source $(A^\infty, \mathscr{A}^\infty, \nu)$ is said *stationary* if for every $E \in \mathscr{A}^\infty$ it holds $\nu(E^*) = \nu(E)$. If we define on the probability space $(A^\infty, \mathscr{A}^\infty, \nu)$ a discrete-time random processes $(X_n \mid n = 1, 2, \ldots)$ as the $\mathscr{A}^\infty$-measurable mappings

$$X_n(z_1, z_2, \ldots) = z_n, \quad n = 1, 2, \ldots \tag{7}$$

(coordinate-projections; note that $\mathscr{A}$ was supposed to contain all singletons $\{x\} \subset A$!) then it is not very difficult to prove that $(A^\infty, \mathscr{A}^\infty, \nu)$ is stationary iff the process $(X_n \mid n = 1, 2, \ldots)$ is stationary in the sense of Doob [6].

Proposition 1. If $(A^\infty, \mathscr{A}^\infty, \nu)$ is a stationary source such that, for every $y \in \tilde{A}$,

$$\int_A d(x, y)\, \mathrm{d}\nu_1(x) < \infty \tag{8}$$

then the distortion-rate function $\delta_\nu(R)$, $0 \leqq R \leqq \log_2 \|\tilde{A}\|$, is regular.

Proof. (I) Let

$$\gamma = \|\bar{A}\| \max_{y \in \bar{A}} \int_A d(x, y)\, \mathrm{d}\nu_1(x)\,.$$

By assumptions, it holds $0 \leqq \gamma < \infty$. It suffices to prove that for every natural n and every $\mathscr{A}^n$-measurable function $\Phi: A^n \mapsto \bar{A}^n$ it holds (5). But, by (1) and by the nonnegativity of d,

$$\int_{A^n} d_n(x, \Phi(x))\, \mathrm{d}\nu_n(x) = \int_{A^n} (1/n) \left[\sum_{i=1}^{n} d(x_i, (\Phi(x))_i)\right] \mathrm{d}\nu_n(x) \leqq$$

$$\leqq \int_{A^n} (1/n) \left[\sum_{i=1}^{n} \sum_{y \in \bar{A}} d(x_i, y)\right] \mathrm{d}\nu_n(x) = \sum_{y \in \bar{A}} \int_A d(x, y)\, \mathrm{d}\nu_1(x) \leqq \gamma\,.$$

(II) Now we shall prove (4). Let us fix $0 \leqq R \leqq \log_2 \|\bar{A}\|$. By definition, for every natural n there exists a codebook $C_n \in \mathbb{C}_n(R)$ such that

$$\delta_{n,\nu}(R) \leqq d_{n,\nu}(C_n) < \delta_{n,\nu}(R) + 1/n\,. \tag{9}$$

If $n \geqq 2$ and there exists $1 \leqq r < n$ such that

$$d_{n,\nu}(C_r \times C_{n-r}) = \min_{s=1,\dots,n-1} d_{n,\nu}(C_s \times C_{n-s}) < d_{n,\nu}(C_n)$$

then replace C_n by $C_r \times C_{n-r} \in \mathbb{C}_n(R)$. Applying this procedure for $n = 3, 4, \dots$ we obtain at the end a sequence of codebooks $C_n \in \mathbb{C}_n(R)$ satisfying in addition to (9) the condition

$$d_{n+m,\nu}(C_{n+m}) \leqq d_{n+m,\nu}(C_n \times C_m) \quad \text{for all} \quad m, n = 1, 2, \dots \tag{10}$$

Now, in order to prove (4), take first into account that, by (9) and by the definition of $\delta_\nu(R)$, it holds

$$\delta_\nu(R) = \liminf_{n \to \infty} d_{n,\nu}(C_n)\,. \tag{11}$$

Further, by definition, it holds for every $y \in \bar{A}^n$

$$d_n(x, C_n) \leqq d_n(x, y)$$

so that, by definition of $d_{n,\nu}(C_n)$ and (5),

$$d_{n,\nu}(C_n) \leqq \gamma\,.$$

Finally it follows from the definition of $d_{n,\nu}(C_n)$ and from the stationarity of ν that it holds for every $m, n = 1, 2, \dots$

$$(n + m)\, d_{n+m,\nu}(C_n \times C_m) \leqq n\, d_{n,\nu}(C_n) + m\, d_{m,\nu}(C_m)$$

so that, by Lemma on p. 113 of Gallager [7], the sequence $(d_{n,\nu}(C_n) \mid n = 1, 2, \dots)$ is convergent. It follows from here and from (11) that (4) holds. □

2. THE MAIN RESULTS

We are now going to introduce the concepts figuring in the main result of this paper. Some of the examples are not necessary for the present paper but they are helpful here and effectively used in the follow up paper [16].

Let $(A_1^\infty, \mathscr{A}_1^\infty)$, $(A_2^\infty, \mathscr{A}_2^\infty)$ be defined in the same way as $(A^\infty, \mathscr{A}^\infty)$ – it is assumed neither that $(A_1^\infty, \mathscr{A}_1^\infty)$ and $(A_2^\infty, \mathscr{A}_2^\infty)$ are mutually equal nor that either of them equals $(A^\infty, \mathscr{A}^\infty)$. By a *communication channel* we mean a triple

$$((A_1^\infty, \mathscr{A}_1^\infty), (P_z \mid z \in A_1^\infty), (A_2^\infty, \mathscr{A}_2^\infty)) \tag{12}$$

where P_z are probability measures on $(A_2^\infty, \mathscr{A}_2^\infty)$. Hereafter we assume that this channel is *nonanticipating*, i.e. that, for every natural n and every $x \in A_1^n$, and $E \in \mathscr{A}_2^n$, the probability $P_z(E \times A_2^\infty)$ is constant for all extensions $z \in A_1^\infty$ of the vector x. For brevity, this probability is denoted by $P_{n,x}(E)$.

Let $A_{1,n} \subset A_1^n$ for $n = 1, 2, \ldots$ be nonempty sets representing *input constraints* of the channel (12): input n-messages (codewords) of the channel must be from $A_{1,n}$. If $A_{1,n} = A_1^n$, for $n = 1, 2, \ldots,$ then there are in fact no input constraints. The other extreme is when $A_{1,n}$ are singletons for $n = 1, 2, \ldots$ The *channel capacity* $C \geqq 0$ is then the supremum of all $R \geqq 0$ such that for every $\varepsilon > 0$ there exists n_0 with the property that if $n > n_0$ then there exists a *channel input codebook* $B_n \subset A_{1,n}$ of $2^{[nR]}$ elements and an $\mathscr{A}_2^n$-measurable mapping ψ: $A_2^n \to B_n$, called *B_n-decoder*, for which

$$P_{n,x}(\psi^{-1}(x)) > 1 - \varepsilon \quad \text{for every} \quad x \in B_n\,.$$

We see that our concept of channel capacity is close to the capacity $C(0+)$ considered in Section 7.7 of Wolfowitz [19].

Now we consider three examples – the last one presenting a practically very important channel with nontrivial input constraints.

Example 1. Let us consider an information source $(\mathbb{R}^\infty, \mathscr{A}^\infty, \nu)$ with alphabet $A = \mathbb{R}$. Denote by $(z_1, z_2, \ldots)$ an element of $\mathbb{R}^\infty$ and suppose

$$\int_{\mathbb{R}^\infty} z_k^2 \,\mathrm{d}\nu(z_1, z_2, \ldots) < \infty\,, \quad k = 1, 2, \ldots.$$

Then the expressions

$$c(i, j) = \int_{\mathbb{R}^\infty} z_i z_j \,\mathrm{d}\nu(z_1, z_2, \ldots)$$

exist and are finite for every $i, j = 1, 2, \ldots$. If we consider the covariance function of the process (7), in the sense of Doob [6], then its values at i, j coincide with $c(i, j)$. Thus the following terminology corresponds to that of Doob. The source $(\mathbb{R}^\infty, \mathscr{A}^\infty, \nu)$ is called *stationary in wide sense* if

$$\int_{\mathbb{R}^\infty} z_k \, d\nu(z_1, z_2, \ldots) = 0 \quad \text{for every} \quad k = 1, 2, \ldots$$

and if there exists a function r: $\{0, \pm 1, \ldots\} \to \mathbb{R}$ such that $r_{-k} = r_k$ and, for every $i, j = 1, 2, \ldots,$

$$c(i, j) = r_{i-j}\,.$$

The function r is then said the *covariance function* of the source. The covariance function of any source which is stationary in wide sense satisfies the condition

$$\sum_{i,j=1}^{n} r_{i-j} \lambda_i \bar{\lambda}_j \geqq 0\,, \quad n = 1, 2, \ldots \tag{13}$$

for any choice of complex numbers $\lambda_1, \ldots, \lambda_n$ (cf. § 3 in Chap. X of Doob [6]). It follows from the results of § 3 in Chap. II of Doob [6] that the converse is true as well: If $r: \{0, \pm 1, \ldots\} \mapsto \mathbb{R}$ is a function satisfying (13) then there exists a source $(\mathbb{R}^\infty, \mathscr{A}^\infty, \nu)$ stationary in wide sense, with the covariance function r.
If

$$\sum_{k=-\infty}^{\infty} |r_k| < \infty$$

then we define a *spectral density* of the source by

$$\varphi(\omega) = \sum_{k=-\infty}^{\infty} r_k \, e^{-ik\omega}, \quad -\pi \leqq \omega \leqq \pi .$$

It follows from IV, 3 in Anděl [1] that $\varphi(\omega)$ takes on a.s. values from the interval $[0, \infty)$. Since for every $n \geqq 1$ and every complex numbers $\lambda_1, \ldots, \lambda_n$

$$\sum_{j,k=1}^{n} r_{k-j} \lambda_k \bar{\lambda}_j = \int_{-\pi}^{\pi} \Big| \sum_{k=1}^{n} \lambda_k \, e^{ik\omega} \Big|^2 \varphi(\omega) \, d\omega ,$$

it is clear that if $\varphi(\omega) > 0$ for every $-\pi \leqq \omega \leqq \pi$ then all $n \times n$ matrices $[r_{k-j}]$, $j, k = 1, \ldots, n$, are positively definite.

Example 2. The source $(\mathbb{R}^\infty, \mathscr{A}^\infty, \nu)$ is called *stationary Gaussian* if it is stationary in wide sense, with a covariance function r, and if for every natural n the n-source $(\mathbb{R}^n, \mathscr{A}^n, \nu_n)$ has an n-dimensional normal distribution $N(\boldsymbol{O}_n, \boldsymbol{V}_n)$ where

$$\boldsymbol{O}_n = (0, \ldots, 0) \in \mathbb{R}^n ,$$

and $\boldsymbol{V}_n$ is a Toeplitz-symmetric $n \times n$ matrix with r_k on the kth diagonal, i.e.

$$\boldsymbol{V}_n = \begin{bmatrix} r_0 & r_1 & \cdots & r_{n-2} & r_{n-1} \\ r_1 & r_0 & \ddots & \ddots & r_{n-2} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ r_{n-2} & \ddots & \ddots & r_0 & r_1 \\ r_{n-1} & r_{n-2} & \cdots & r_1 & r_0 \end{bmatrix} .$$

By (13), $\boldsymbol{V}_n$ is *positively semidefinite.* If all matrices $\boldsymbol{V}_n$, $n = 1, 2, \ldots$, are regular then the stationary Gaussian process is said *regular.* A positively semidefinite matrix $\boldsymbol{V}_n$ is regular iff it is *positively definite.*

Example 3. We shall say that (12) is a *stationary Gaussian channel* if $A_2 = \mathbb{R}$, in which case – by definition – $\mathscr{A}_2$ is the σ-algebra of all Borel sets on $\mathbb{R}$, and if for every $z \in A_1^\infty$ there is a covariance function r satisfying (13) such that $P_{n,(z_1,\ldots,z_n)} = N((z_1, \ldots, z_n), \boldsymbol{V}_n)$ where $\boldsymbol{V}_n$ is as in Example 2. A stationary Gaussian channel is called *memoryless* if for every $z \in A_1^\infty$

$$r_k = 0 \quad \text{for all} \quad k = 1, 2, \ldots$$

It is easy to see that the stationary Gaussian channel is nonanticipating. Let us denote by $\mathscr{A}$ the σ-algebra of all Borel subsets of $\mathbb{R}$ and let us consider a stationary Gaussian memoryless channel

$$((\mathbb{R}^\infty, \mathscr{A}^\infty), (P_z \mid z \in \mathbb{R}^\infty), (\mathbb{R}^\infty, \mathscr{A}^\infty))$$

with $r_0 > 0$ and with the input constraints

$$A_{1,n} = \{x \in \mathbb{R}^n \mid (1/n) \sum_{k=1}^{n} x_i^2 \leqq \varepsilon\}, \quad \varepsilon > 0, \quad n = 1, 2, \ldots.$$

For this channel the capacity, first deduced by Shannon [14], is

$$C = \tfrac{1}{2} \log_2 \left(1 + \frac{\varepsilon}{r_0}\right),$$

as rigorously proved in Chap. 9 of Wolfowitz [19]. There the above considered memoryless stationary Gaussian channel is studied with some other input constraints $A_{1,n}$ as well, e.g. with

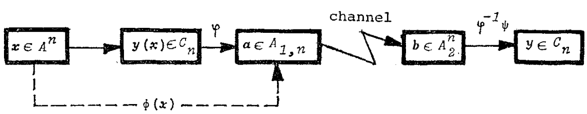
$$A_{1,n} = \left\{\frac{1}{2m}, \frac{3}{2m}, \ldots, \frac{2m-1}{2m}\right\}^n,$$

where m is a natural number (such "available channel input alphabets" are considered under the assumption $r_0 = 1$).

We shall now describe a coding and decoding scheme for a source $(A^\infty, \mathscr{A}^\infty, \nu)$ and the channel (12). Let n be an arbitrary natural number and let C_n be a source codebook from $\mathbb{C}_n(R)$ containing not more elements than the channel input set $A_{1,n}$. Obviously, there exists at least one $0 \leqq R \leqq \log_2 \|\tilde{A}\|$, namely $R = 0$, such that $\mathbb{C}_n(R)$ contains at least one such codebook. Let φ be a C_n-*coder*, i.e. a one-to-one mapping from C_n into $A_{1,n}$ and define

$$\Phi(x) = \varphi(y(x)), \quad x \in A^n,$$

where $x \to y(x)$ is the mapping figuring in the definition of $d_n(x, C_n)$ above. The source message $x \in A^n$, according to the block diagram



is transmitted as an admissible input codeword $\Phi(x) \in A_{1,n}$ through the channel, received as b at the channel output and, finally, delivered to the user as $y = \varphi^{-1} \psi(b)$ from the source codebook C_n where ψ is a $\Phi(A^n)$-decoder defined above (since $\Phi(A^n) = \varphi(C_n)$, we shall refer to ψ as to a $\varphi(C_n)$-decoder). Now, if $b \in A_2^n$ is the channel output n-message, then

$$d_n(x, \varphi^{-1} \psi(b))$$

is the distortion of the source message x and

$$\tilde{\delta}_n(x) = \int_{A_2{}^n} d_n(x, \varphi^{-1} \psi(b)) \, \mathrm{d}P_{n,\Phi(x)}(b) = \sum_{y \in C_n} d_n(x, y) \, P_{n,\Phi(x)}(\psi^{-1}(\varphi(y)))$$

is the *expected distortion* of this message. Notice that on the right side there is a sum of $\mathscr{A}^n$-measurable functions. Indeed, the $\mathscr{A}^n$-measurability of $\Phi(x)$ follows from the same argument as the $\mathscr{A}^n$-measurability of $d_n(x, C_n)$ above and the rest is clear. The average expected distortion

$$(14) \qquad \tilde{\delta}_n = \int_{A^n} \tilde{\delta}_n(x) \, \mathrm{d}\nu_n(x) = \sum_{v,y \in C_n} (P_{n,\varphi(v)} (\psi^{-1} \varphi(y)) \int_{\Psi(v)} d_n(x, y) \, \mathrm{d}\nu_n(x)) ,$$

where

$$\Psi(v) = \{x \in A^n : y(x) = v\} ,$$

thus describes a source distortion at the channel output, briefly an *output distortion*, attained by the codebook C_n, the C_n-coder φ and the $\varphi(C_n)$-decoder ψ. When convenient, we write $\tilde{\delta}_n(C_n, \varphi, \psi)$ instead of $\tilde{\delta}_n$.

Let us note that the output distortion (14) with an alphabet distortion measure was probably first studied by Nedoma [11] who considered a discrete model where the source alphabets $A = \tilde{A}$ as well as the channel input and output alphabets A_1 and A_2 are finite and the distortion functions $d_n(x, y)$ are defined by (2) (the Hamming distances $d_n(x, y) = \delta_{x,y}$ were also considered). The output distortion has been termed *risk* by him and the interest was concentrated on sources and channels for which $\tilde{\delta}_n(C_n, \varphi, \psi)$ can be made arbitrarily close to 0 for sufficiently large n (cf. Theorem 1, 2 in Sec. III there).

Next follows the main result of this paper which is a general information transmission theorem (joint source/channel coding theorem). It generalizes many former theorems of this kind, e.g. Theorem 9.2.1. of Gallager [7] or Theorem 2.4 of Csiszár and Körner [5]. It is formulated for arbitrary nonanticipating channels (neither the channel stationarity nor even the $\mathscr{A}_1^\infty$-measurability of channel probabilities $P_z(E)$, $E \in \mathscr{A}_2^\infty$, are assumed). (Of course, in practical applications, where source n-messages are transmitted through the channel subsequently, block after the block, the block-stationarity and block-independence of the channel simplify interpretation of the theorem.) Analogically the source need not in principle be stationary – but the sources for which we guarantee in this paper the assumed distortion-rate function regularity are stationary – see Proposition 1. Nevertheless there are trivial examples of nonstationary sources for which the regularity assumption holds too.

Theorem. Let us consider an information source $(A^\infty, \mathscr{A}^\infty, \nu)$ with an available reproducing alphabet $\tilde{A} \subset A$ of size $\|\tilde{A}\| > 1$ and with a regular distortion-rate function $\delta_\nu(R)$, $0 \leqq R \leqq \log_2 \|\tilde{A}\|$ and a nonanticipating communication channel (12) with a capacity $0 \leqq C \leqq \log_2 \|\tilde{A}\|$. Then for every $\varepsilon > 0$ there exists a natural number n_0 such that for all natural numbers $n > n_0$ there exist a source codebook $C_n \subset \tilde{A}^n$, a C_n-coder φ and a $\varphi(C_n)$-decoder ψ such that the output distortion $\tilde{\delta}_n(C_n, \varphi, \psi)$ defined by (14) satisfies the inequality

$$\tilde{\delta}_n(C_n, \varphi, \psi) \leqq \delta_\nu(C-) + \varepsilon \quad \text{where} \quad \delta_\nu(C-) = \lim_{R \uparrow C} \delta_\nu(R) .$$

Proof. Let us fix $\varepsilon > 0$.

(I) It is clear from the definition of $\delta_\nu(R)$ that the function δ_ν is nonincreasing on the domain $[0, \log_2 \|\tilde{A}\|]$. Therefore, for given C and ε, there exists $0 < R_0 < C$ such that for all $R_0 < R < C$

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

Let R be an arbitrary fixed point from the interval (R_0, C).

(II) It follows from the regularity of δ_ν, in particular from the assumption (4), that there exists natural n_1 such that, for all $n > n_1$, thete exist $C_n \in \mathbb{C}_n(R)$ with the property

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

or, by (I), with the property

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

(III) Let $\gamma \geqq 0$ be a constant satisfying (5). It follows from the definition of channel capacity C and from the inequality $0 \leqq R < C$ that there exists $n_0 > n_1$ such that, for every $n > n_0$, there exist a C_n-coder φ and $\varphi(C_n)$-decoder ψ for which

$$P_{n,\varphi(y)}(\psi^{-1}(\varphi(y)) > 1 - \frac{\varepsilon}{3(\gamma+1)} \quad \text{for every} \quad y \in C_n$$

It follows from here

$$\sum_{\substack{y \in C_n \\ y \neq v}} P_{n,\varphi(v)}(\psi^{-1}(\varphi(y)) < \frac{\varepsilon}{3(\gamma+1)} \quad \text{for every} \quad v \in C_n\,.$$

(IV) Let $n > n_0$ be fixed. It follows from (14) that

$$\tilde{\delta}_n(C_n, \varphi, \psi) \leqq \mathscr{E}(1) + \mathscr{E}(2)$$

where

$$\mathscr{E}(1) = \sum_{y \in C_n} \int_{\Psi(y)} d_n(x, y)\, \mathrm{d}\nu_n(x) = \int_{A^n} d_n(x, C_n)\, \mathrm{d}\nu_n(x) = d_{n,\nu}(C_n) \leqq \delta_\nu(C-) + \tfrac{2}{3}\varepsilon \qquad \text{(cf. (II))}$$

and

$$\mathscr{E}(2) = \sum_{\substack{v, y \in C_n \\ v \neq y}} [P_{n,\varphi(v)}(\psi^{-1}(\varphi(y))) \int_{\Psi(v)} d_n(x, y)\, \mathrm{d}\nu_n(x)] \leqq$$

$$\leqq \sum_{v \in C_n} [\sum_{\substack{y \in C_n \\ y \neq v}} P_{n,\varphi(v)}(\psi^{-1}(\varphi(y)) \int_{\Psi(v)} d_n(x, y_v)\, \mathrm{d}\nu_n(x)]$$

where $y_v \in C_n$ maximizes the integral

$$\int_{\Psi(v)} d_n(x, y)\, \mathrm{d}\nu_n(x)$$

on C_n. Let us define an $\mathscr{A}^n$-measurable mapping Φ: $A^n \to C_n$ by

$$\Phi(x) = y_v \quad \text{for} \quad x \in \Psi^{-1}(v)\,, \quad v \in C_n\,.$$

Then it follows from the last inequality and from (III)

$$\mathscr{E}(2) \leqq \sum_{v \in C_n} \left(\frac{\varepsilon}{3(\gamma+1)}\right) \int_{\Psi^{-1}(v)} d_n(x, y_v)\, \mathrm{d}\nu_n(x)) =$$

$$= \frac{\varepsilon}{3(\gamma + 1)} \int_{A^n} d_n(x, \Phi(x)) \, dv_n(x) \leqq \frac{\varepsilon\gamma}{3(\gamma + 1)} < \tfrac{1}{3}\varepsilon . \qquad \text{(cf. (5)).}$$

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

$$\tilde{\delta}_n(C_n, \varphi, \psi) \leqq \delta_v(C-) + \varepsilon . \qquad \square$$

Remark. Theorem says that at least as small distortion as can be attained in the rate domain $0 \leqq R < C \leqq \log_2 \|\tilde{A}\|$ by a codebook in available reproducing alphabet $\tilde{A}$ can be attained at the output of any channel with capacity C by means of a "real time" block coding.

The applicability of the distortion-rate function $\tilde{\delta}_v(R)$, which is clear from Theorem, would be of no practical importance if we were not able to evaluate this function explicitly, or at least to compute its values for preselected rates R. Fortunately, this is not the case. In the next part of our paper we present one proposition about evaluation of $\delta_v(R)$ by means of convex optimization methods and some examples relevant in coding speech and image signals.

Let us consider a source $(A^\infty, \mathscr{A}^\infty, v)$ and, for every natural n, n-sources $(\tilde{A}^n, \tilde{\mathscr{A}}^n, P_{n,x})$, $x \in A^n$, where $\tilde{A} \subset A$ is a finite available alphabet and $\tilde{\mathscr{A}}$ is the σ-algebra of all subsets of $\tilde{A}$ and where, for every $\tilde{E} \in \tilde{\mathscr{A}}^n$, $P_{n,x}(\tilde{E})$ as a function of $x \in A^n$ is $\mathscr{A}^n$-measurable. Then we can define for the input n-source $(A^n, \mathscr{A}^n, v_n)$ a double source $(A^n \times A^n, \mathscr{A}^n \otimes \mathscr{A}^n, v_n \otimes P_n)$ by

$$(v_n \otimes P_n)(E \times \tilde{E}) = \int_E P_{n,x}(\tilde{E}) \, dv_n(x)$$

and by the standard extension argument. We can also define a marginal output source $(\tilde{A}^n, \tilde{\mathscr{A}}^n, Q_n)$ by

$$Q_n(\tilde{E}) = (v_n \otimes P_n)(A^n \times \tilde{E}) \quad \text{for} \quad \tilde{E} \in \tilde{\mathscr{A}}^n$$

and we can consider the I_1-divergence $I_1(v_n \otimes P_n \parallel v_n \times Q_n)$ of the joint measure $v_n \otimes P_n$ and of the product $v_n \times Q_n$ of its marginals on the double space (see Liese and Vajda [9]). Let us denote by $\mathscr{P}_n(R)$, $R \geqq 0$, the class of all families $(P_{n,x} \mid x \in A^n)$ of the above considered type with the property

$$I_1(v_n \otimes P_n \parallel v_n \times Q_n) \leqq R$$

and let for some d_n under consideration

$$D_{n,v}(R) = \begin{cases} \inf\limits_{P_n \in \mathscr{P}_n(R)} \int_{A^n \times \tilde{A}^n} d_n(x, y) \, d(v_n \otimes P_n) & \text{if} \quad \mathscr{P}_n(R) \neq \emptyset \\ \infty \quad \text{otherwise} \end{cases}$$

and

$$D_v(R) = \liminf_{n \to \infty} D_{n,v}(R) , \quad R \geqq 0 . \tag{15}$$

Let us remind that the source $(A^\infty, \mathscr{A}^\infty, v)$ is called *ergodic* if $v(E^*) = 0$ or 1 for all sets $E \in \mathscr{A}^\infty$ identical with their "star sets" E^* defined by (6).

Proposition 2. (Berger [2]). If the conditions of Proposition 1 hold then (15) hold with "lim inf" replaced by "lim". Inf, moreover, the source $(A^\infty, \mathscr{A}^\infty, \nu)$ is ergodic then

$$D_\nu(R) = \delta_\nu(R), \quad 0 \leqq R \leqq \log_2 \|\tilde{A}\|.$$

In the literature the *distortion-rate function* means the function (15) rather than (4). Analogically as (4), the distortion-rate function (15) is nonincreasing. The former was defined for $0 \leqq R \leqq \log_2 \|\tilde{A}\|$, the latter is defined for all $R \geqq 0$, but it is a constant equal to $D_\nu(\log_2 \|\tilde{A}\|)$ for $R > \log_2 \|\tilde{A}\|$ because the information $I_1(\nu_n \otimes P_n \| \nu_n \times Q_n)$ is bounded above by $\log_2 \|\tilde{A}\|$. The inverse function $R_\nu(D) = D_\nu^{-1}(R)$ defined by

$$R_\nu(D) = \inf\{R: D_\nu(R) \leqq D\}, \quad D \geqq 0,$$

si called the *rate-distortion function.*

Next follow two examples, both relevant in speech coding. In both these examples we consider available alphabets

$$\tilde{A} = \left\{0, \pm\frac{1}{2^N}, \pm\frac{2}{2^N}, \ldots, \pm\frac{N}{2^N}\right\},$$

and we use the existence of the limit

$$R_\nu^*(D) = \lim_{N\to\infty} R_\nu(D)$$

which follows from Lemma (8·30) in Vajda [18]. It follows from the same lemma, that $R_\nu^*(D)$ is the rate-distortion function under no restriction on the available alphabet $\tilde{A}$, i.e. under $\tilde{A} = A$ which is considered in the theorems of Berger [2], in the two examples. Thus, based on the cited results of Berger and Vajda, the two examples present an approximation $R_\nu^*(D)$ of the function $R_\nu(D)$ which is tight for large sizes of alphabet $\tilde{A}$.

Example 4 (4.5.3 in Berger [2]). Let us consider the square error distortion function (3) and the stationary Gaussian source of Example 2 with covariance function r and zero mean. Assume that the spectral density

$$\varphi(\omega) = \sum_{k=-\infty}^{\infty} r_k \mathrm{e}^{-ik\omega}, \quad \omega \in [-\pi, \pi],$$

exists and has the essential infimum positive and the essential supremum finite. Then for every θ between 0 and the essential supremum

$$R_\nu^*(D) = \frac{1}{4\pi} \int_{-\pi}^{\pi} \max\left\{0, \log_2 \frac{\varphi(\omega)}{\theta}\right\} \mathrm{d}\omega$$

if $D \geqq 0$ satisfies the equation

$$D = \frac{1}{2\pi} \int_{-\pi}^{\pi} \min\{\theta, \varphi(\omega)\} \,\mathrm{d}\omega.$$

If the source is memoryless, i.e. $r_k = 0$ for $k = 1, 2, \ldots$ then $\varphi(\omega) = r_0$ and we obtain from the above formulas

$$R_v(D) = \max\left\{0, \tfrac{1}{2}\log_2 \frac{r_0}{D}\right\},$$

i.e.

$$D_v(R) = r_0\, 2^{-2R}\,.$$

Autoregressive stationary sources are extraordinary important models of segment of speech signals (cf. e.g. Buzo et al. [4] or the follow-up paper [16]). We first present a definition and basic facts about autoregressive processes. Let us consider a probability space $(\Omega, \mathcal{S}, P)$ and, on it, a discrete-time real valued independent process $\mathscr{Z} = (Z_n \mid n = 0, \pm 1, \ldots)$ such that, for all n,

$$\mathsf{E}Z_n = \int_\Omega Z_n \,\mathrm{d}P = 0 \tag{16}$$

$$\sigma^2 = \mathsf{E}Z_n^2 = \int_\Omega Z_n^2 \,\mathrm{d}P > 0\,.$$

Each $\mathscr{Z}$ is uniquely determined by a probability distribution function F on $\mathbb{R}$ – we shall use to write $\mathscr{Z}(F)$ instead of $\mathscr{Z}$. Thus $\mathscr{Z}(F)$ is considered for every p.d.f. F on $\mathbb{R}$ such that

$$\int_{\mathbb{R}} x \,\mathrm{d}F(x) = 0\,, \quad \sigma^2 = \int_{\mathbb{R}} x^2 \,\mathrm{d}F(x) > 0\,. \tag{17}$$

For every natural m we define an $\mathbb{A}$-*subset* $\mathbb{A}_m$ of $\mathbb{R}^m$ by the condition that $\boldsymbol{a} = (a_1, \ldots, a_m) \in \mathbb{R}^m$ belongs to $\mathbb{A}_m$ iff the complex polynomial

$$a(\lambda) = 1 + a_1\lambda^{-1} + \ldots + a_m\lambda^{-m} \tag{18}$$

has all roots inside the unit circle. According to V, 6 in Anděl [1], for every $\mathscr{Z}(F)$ and $\boldsymbol{a} \in \mathbb{A}_m$ there exists a unique wide-sense stationary process $\mathscr{X}(F, \boldsymbol{a}) = (X_n \mid n = 0, \pm 1, \ldots)$ on $(\Omega, \mathcal{S}, P)$ such that X_k and Z_n are independent for $k < n$ and, for all natural k,

$$X_k = -\sum_{j=1}^{m} a_j X_{k-j} + Z_k\,. \tag{19}$$

The process $\mathscr{X}(F, \boldsymbol{a})$ is called an *autoregressive* (AR)-process of order m with parameters F, $\boldsymbol{a}$. If

$$r_k = \int_\Omega X_0 X_k \,\mathrm{d}P \quad k = 0, \pm 1, \ldots,$$

is the covariance function of an AR-process $\mathscr{X}(F, \boldsymbol{a})$, $\boldsymbol{a} \in \mathbb{A}_m$, then it is shown ibid that the *spectral density*

$$\varphi(\omega) = \sum_{k=-\infty}^{\infty} r_k \,\mathrm{e}^{-\mathrm{i}k\omega}\,, \quad -\pi \leqq \omega \leqq \pi\,,$$

of $\mathscr{X}(F, \boldsymbol{a})$ exists and satisfies the relation

$$\varphi(\omega) = \frac{\sigma^2}{|a(\mathrm{e}^{\mathrm{i}\omega})|^2}\,, \quad -\pi \leqq \omega \leqq \pi\,. \tag{20}$$

It follows from here in particular that there is $\Delta > 0$ for which

(21) $$0 < \varphi(\omega) < \Delta\,, \quad -\pi \leqq \omega \leqq \pi$$

and that the covariance function of the AR-process $\mathscr{X}(F, \boldsymbol{a})$ satisfies the relation

(22) $$r_k = \frac{\sigma^2}{2\pi} \int_{-\pi}^{\pi} \frac{e^{ik\omega}}{|\boldsymbol{a}(e^{i\omega})|^2}\, d\omega\,, \quad k = 0, \pm 1, \ldots.$$

Thus the second order properties are the same for all AR-processes $\mathscr{X}(F, \boldsymbol{a})$ with F satisfying (17) for some fixed $\sigma^2 > 0$. In view of this we prefer to speak about classes $\mathscr{X}(\sigma^2, \boldsymbol{a})$ of AR-processes with parameters $(\sigma^2, \boldsymbol{a}) \in (0, \infty) \times \mathbb{A}_m$. Speaking about one AR-process $\mathscr{X}(\sigma^2, \boldsymbol{a})$ we shall mean one arbitrary element of the class $\mathscr{X}(\sigma^2, \boldsymbol{a})$. It follows from (22) that the parameters $(\sigma^2, \boldsymbol{a}) \in (0, \infty) \times \mathbb{A}_m$ uniquely define all covariances r_k of all processes from $\mathscr{X}(\sigma^2, \boldsymbol{a})$. We shall prove that, conversely, the covariances $r_0, \ldots, r_m$ of any AR-process $\mathscr{X}$ of order $m \geqq 1$ uniquely determine the parameters $(\sigma^2, \boldsymbol{a}) \in (0, \infty) \times \mathbb{A}_m$ with the property

$$\mathscr{X} \in \mathscr{X}(\sigma^2, \boldsymbol{a})\,.$$

Indeed, by 5 in V, 13 of Anděl [1], if $\mathscr{X} \in \mathscr{X}(\sigma^2, \boldsymbol{a})$ then

(23) $$\sigma^2 = r_0 + \sum_{j=1}^{m} r_j a_j\,.$$

Further, by the definition of covariance function and by (19), $\mathscr{X} \in \mathscr{X}(\sigma^2, \boldsymbol{a})$ implies

$$r_k = \int_\Omega \left[\left(Z_0 - \sum_{j=1}^{m} a_j X_{k-j}\right) X_0\right] dP = - \sum_{j=1}^{m} a_j \int_\Omega X_{k-j} X_0\, dP$$

i.e.

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

Since, by the definition of spectral density, it holds for all complex $\lambda_1, \ldots, \lambda_m$

$$\sum_{j,k=1}^{m} r_{k-j} \lambda_k \bar{\lambda}_j = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left|\sum_{k=1}^{m} \lambda_k e^{-ik\omega}\right|^2 \varphi(\omega)\, d\omega\,,$$

it follows from (21) that the $m \times m$ matrix $[r_{k-j}]$ occurring in (24) is positively definite. The positive definiteness of $[r_{k-j}]$ implies that there is a unique solution $\boldsymbol{a} = (a_1, \ldots, a_m)$ of the equation (23) which, combined with (23), uniquely determines σ^2 as well. Thus there is a one-to-one relation between covariance vectors $\boldsymbol{r} = (r_0, \ldots, r_m)$ of AR-processes of order m and parameters $(\sigma^2, \boldsymbol{a}) \in (0, \infty) \times \mathbb{A}_m$ of these processes.

Let us consider an arbitrary AR-process $\mathscr{X}(F, \boldsymbol{a}) = (X_n \mid n = 0, \pm 1, \ldots)$, $\boldsymbol{a} \in \mathbb{A}^m$. The subprocess $(X_n \mid n = 1, 2, \ldots)$ is uniquelly determined by a subprocess $(Z_n \mid n = 1, 2, \ldots)$ of independent identically – by F – distributed random variables, by the random m-vector $(X_{-m+1}, \ldots, X_0)$ independent of $(Z_n \mid n = 1, 2, \ldots)$ and by the parameter $\boldsymbol{a}$ (i.e. by the relation (19)). The same is true for classes of AR-processes when F is replaced by $\sigma^2 > 0$. The vector $(X_{-m+1}, \ldots, X_0)$ is a random "initial state" of the subprocess. This observation serves as a basis for the definition that follows.

Let $(X_n \mid n = 1, 2, \ldots)$ be a random process on $(\Omega, \mathscr{S}, P)$. This process is said *AR-process of order* $m \geqq 1$ if there exist F satisfying (17) and $\boldsymbol{a} \in \mathbb{A}_m$ and a random vector $(X_{-m+1}, \ldots, X_0)$ such that (19) holds for $k = 1, 2, \ldots$, where $(Z_n \mid n = 1, 2, \ldots)$ is a sequence of mutually and on $(X_{-m+1}, \ldots, X_0)$ independent random variables distributed by F. Analogically as above, we consider AR-processes with parameters $(\sigma^2, \boldsymbol{a}) \in (0, \infty) \times \mathbb{A}_m$. The "one-sided" AR-processes are wide-sense stationary and the above considered properties of "doubly-sided" AR-processes apply to these process with obvious modifications.

Let us consider an information source $(\mathbb{R}^\infty, \mathscr{A}^\infty, \nu)$. If the process $(X_n \mid n = 1, 2, \ldots)$ defined by (7) is an AR-process of order $m \geqq 1$ then the source $(\mathbb{R}^\infty, \mathscr{A}^\infty, \nu)$ is said an *AR-source* of order m. It is clear from Example 2 that the Gaussian AR-sources of arbitrary order are stationary in the ordinary sense (the same holds for Gaussian AR-processes). Now we are done to present the next example.

Example 5. Let us consider a Gaussian AR-source $(\mathbb{R}^\infty, \mathscr{A}^\infty, \nu)$ of order $m \geqq 1$ with parameters $(\sigma^2, \boldsymbol{a}) \in (0, \infty) \times \mathbb{A}_m$ and with the square error distortion function (3) and let

$$\psi(\omega) = \frac{1}{\sigma^2} \left|1 + \sum_{k=1}^{m} a_k \,\mathrm{e}^{-\mathrm{i}k\omega}\right|^2, \quad -\pi \leqq \omega \leqq \pi .$$

Let θ be between 0 and $1/\Delta$ where $\Delta > 0$ is the esssup of $\psi(\omega)$ on $[-\pi, \pi]$. Then it follows from Example 4 and from (20) that it holds

$$R_\nu^*(D) = \frac{1}{4\pi} \int_{-\pi}^{\pi} \log_2 \frac{1}{D\, \psi(\omega)} \,\mathrm{d}\omega = \tfrac{1}{2} \log_2 \frac{\sigma^2}{D}$$

if

$$0 \leqq D \leqq \Delta .$$

For example, let us consider a first-order AR source with parameters $(\sigma^2, a_1) \in \in (0, \infty) \times \mathbb{A}_1$. Here $\mathbb{A}_1 = (-1, 1)$ for $-a_1$ is the root of the polynomial $\boldsymbol{a}(\lambda) = = 1 + a_1 \lambda^{-1}$. We have

$$\psi(\omega) = \frac{1 + 2a_1 \cos \omega + a_1^2}{\sigma^2}$$

so that $\Delta = (1 + |a_1|)^2/\sigma^2$. If this source is Gaussian then it follows from the general result above

$$R(D) = \tfrac{1}{2} \log_2 \frac{\sigma^2}{D} \quad \text{for every} \quad 0 \leqq D \leqq \frac{\sigma^2}{(1 + |a_1|)^2} .$$

The alphabet-oriented source coding theory presented in this paper has applications, e.g. in the compression of speech signals. For example, codebooks with 8-bit binary available reproducing alphabets and with information rates of order 10^4 bits per second are of great practical importance in digital transmission of speech – see e.g. Markel and Gray [10] or Šedivý [15].

3. TWO NONSTANDARD USERS

Distortion functions of the form of average per-letter distortions defined by (1), which are characteristic for the alphabet-oriented source coding theory, lead to codings (and in the case of joint source/channel model also to decodings) with the philosophy to approximate or reconstruct the message letter by letter as exactly as possible. This is why we speak about letter-riented coding. Letter-oriented means in many cases the same as user-oriented since an "average user" is probably satisfied by the uniform letter-by-letter proximity of the information he obtains. In order to motivate one source coding theory which is presented in the follow-up paper [16], we shall present examples of two users which are surely not "average" in the above considered sense.

Example 6. Let us consider a stationary *memoryless source* $(\mathbb{R}^\infty, \mathscr{A}^\infty, \nu)$ which is defined by a probability measure ν_0 on $(\mathbb{R}, \mathscr{A})$ and by the condition

$$\nu_n = \nu_0^n \quad \text{for every} \quad n = 1, 2, \ldots.$$

Let us further assume that ν_0 describes the uniform distribution on $(a, b) \subset \mathbb{R}$ (so that the source is in our terminology continuous!). Let finally the user be a statistician who expects to be paid for telling the numbers a and b to an institution and who decided to pay half of the expected income to the Universal Source Coding Company for loading 72-bit data $x_1, \ldots, x_{10000}$ he borrowed from the institution to his 8-bit personal computer. This fictive statistician will be deeply dissappointed when he learns that the first and the last order statistics $x_{(1)}$ and $x_{(10000)}$ have been loaded into his computer with approximately the same precision as the remaining 9998 numbers which are nothing but a mere balast in the borrowed data packet. Next time he will prefer services of the Uniform Source Coding Company which is operating with the distortion measure

$$d_n(x, y) = |x_{(1)} - y_{(1)}| + |x_{(n)} - y_{(n)}|$$

provided he learns that such a company exists.

Example 7. Let us consider a source $(\mathbb{R}^\infty, \mathscr{A}^\infty, \nu)$ – the measure ν is unimportant in this case, the emphasis will be on distortion measures $d_n(x, y)$ for n-messages $x, y \in \mathbb{R}^n$. Let us consider that $x = (x_1, \ldots, x_n)$ is obtained by sampling a short segment (about 20 msec) of a speech signal. An experimentally verified fact is that the human ear hears spectra of segments of signals $x = (x_1, \ldots, x_n)$ rather than the coordinates $x_1, \ldots, x_n$ – in this respect we refer to Markel and Gray Jr. [10]. The mere fact that, from the point of view of hearing, the distortion measures (1) are unsatisfactory can be demonstrated by two periodic signals $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$ where

$$x_i = (-1)^i, \quad y_i = (-1)^{i+1}.$$

The ear will hear in both cases the same tone – it is not able to distinguish whether the signal starts with -1 or $+1$ – while the distortion (1) in this case equals

$$d_n(x, y) = d(1, -1)$$

which is typically far from zero.

Now we explain what was meant by the "spectrum of the signal x" above. Let us consider a stationary ergodic information source $(\mathbb{R}^\infty, \mathscr{A}^\infty, \nu)$ with a covariance function r satisfying the condition

$$\sum_{k=-\infty}^{\infty} |r_k| < \infty \tag{25}$$

and with a spectral density φ positive on $[-\pi, \pi]$ (cf. Example 1). As shown in Example 1, for every $s \geqq 0$, the matrices $[r_{k-j}]$, $j, k = 0, \ldots, s$, are positively definite. Let us consider arbitrary $n > m \geqq 1$ and $x \in \mathbb{R}^n$ and let r_x: $\{\ldots, -1, 0, 1, \ldots\} \mapsto \mathbb{R}$ be defined by

$$r_{x,k} = \begin{cases} 0 & \text{if} \quad k > m \\ 1/(n-k)\sum_{j=1}^{n-k} x_j x_{j+k} & \text{if} \quad 0 \leqq k \leqq m \\ r_{x,-k} & \text{if} \quad k < 0\,. \end{cases} \tag{26}$$

The event $\mathbb{B}_n \subset \mathbb{R}^n$ defined by the condition that all matrices $[r_{x,k-j}]$, $j, k = 0, \ldots, s$, $s = 0, 1, \ldots$ are positively definite is a measurable set from $\mathscr{A}^n$, for the condition is equivalent to the positivity of countably many determinants whose elements are $\mathscr{A}^n$-measurable functions (26). Similar argument, combined with that, by the ergodicity, $r_{x,k}$ tend in ν-probability to r_k as $n \to \infty$ and with the fact that all functions (26) except of those for $k = 0, \pm 1, \ldots, \pm m$ are identically zero (m is fixed!) implies

$$\lim_{n\to\infty} \nu_n(\mathbb{B}_n) = 1\,.$$

Thus the complement $\mathbb{B}_n^c \in \mathscr{A}^n$ is an asymptotically negligible set of messages. Our neglection of this set is apparent from that, ignoring what is considered by spectra of signals $x \in \mathbb{B}_n^c$ in the acoustics (cf. the above cited book of Markel and Gray Jr., or Buzo et al. [4]) we define these spectra by the spectral densities of Example 1 for $r = r_x$ defined by

$$r_{x,k} = \delta_0(k)\,(1/n) \sum_{j=1}^{n} x_i^2 \tag{27}$$

instead of (26). For signals x from the "practically important" set $\mathbb{B}_n$ we also define spectra by the spectral densities of Example 1, but for $r = r_x$ defined by (26). We hope that on $\mathbb{B}_n$ we are in agreement with what is considered in the accoustics. In both cases, all matrices $[r_{x,k-j}]$, $j, k = 0, \ldots, s$, $s = 0, 1, \ldots,$ are positively definite (for $x \in \mathbb{B}_n^s$ these matrices are the unit matrices I_s multiplied by the right-hand factor on the right-hand side of (27)). Thus the spectral densities $\varphi_x(\omega)$, $-\pi \leqq \omega \leqq \pi$, under consideration are defined by

$$\varphi_x(\omega) = \begin{cases} \sum_{k=-m}^{m} r_{x,k}\, e^{-ik\omega} & \text{for} \quad x \in \mathbb{B}_n \\ r_{x,0} & \text{for} \quad x \in \mathbb{B}_n^c\,. \end{cases} \tag{28}$$

We can summarize the results leading to our spectral representation of messages $x \in \mathbb{R}^n$, $n \geqq 1$, from ergodic information sources as follows.

Proposition 3. Let $(\mathbb{R}^\infty, \mathscr{A}^\infty, \nu)$ be a stationary ergodic source satisfying (25). Then for every $m \geqq 1$ and $\varepsilon > 0$ there exists $n_0 > m$ such that, for every $n > n_0$, there exists a set $\mathbb{B}_n \in \mathscr{A}^n$ such that

$$\nu_n(\mathbb{B}_n) > 1 - \varepsilon$$

and that for every $x \in \mathbb{R}^n$ there exists a wide-sense stationary information source $(\mathbb{R}^\infty, \mathscr{A}, \nu_x)$ with the covariance function defined by (26) or (27), depending on whether $x \in \mathbb{B}_n$ or $x \in \mathbb{B}_n^c$ respectively, and with the spectral density (28).

Proof. The only point which was not made clear above is the existence of the source $(\mathbb{R}^\infty, \mathscr{A}, \nu_x)$ which is however clear from what has been said in Example 1. □

By Proposition 3, it is meaningful to consider for ergodic sources $(\mathbb{R}^\infty, \mathscr{A}^\infty, \nu)$ distortion functions $d_n(x, y)$ defined as arbitrary distances of the corresponding spectral densities φ_x, φ_y, e.g.

$$d_n(x, y) = \|\varphi_x - \varphi_y\|_\alpha\,, \quad x, y \in \mathbb{R}^n\,, \tag{29}$$

where $\|\cdot\|_\alpha$ is the norm of the Banach L_α-space on $[-\pi, \pi]$ with the Lebesgue measure, $1 \leqq \alpha \leqq \infty$. The more, one can consider codebooks C_n consisting of $\|C_n\|$ arbitrary wide-sense stationary information sources $(\mathbb{R}^\infty, \mathscr{A}^\infty, \mu)$ instead of $\|C_n\|$ messages $y \in \mathbb{R}^n$, which is typical of the classical alphabet-oriented source coding theory, with (1) replaced e.g. by

$$d_n(x, \mu) = \|\varphi_x - \varphi_\mu\|_\alpha\,, \quad x \in \mathbb{R}^n\,, \quad \mu \in \mathfrak{M}\,, \tag{30}$$

where $\mathfrak{M}$ denotes the set of all wide-sense stationary measures μ on $(\mathbb{R}^\infty, \mathscr{A}^\infty)$ with spectral densities $\varphi_\mu \in L_\alpha$. This opens new possibilities in the source coding theory, partly exploited in the forthcoming paper [16].

(Received November 7, 1986.)

REFERENCES

[1] J. Anděl: Statistical Analysis of Time Series (in Czech). SNTL, Prague 1976. German edition: Akademie Verlag, Berlin 1984.

[2] T. Berger: Rate Distortion Theory. Prentice-Hall, Englewood Cliffs 1971.

[3] P. Billingsley: Information and Ergodic Theory. J. Wiley, New York 1965.

[4] A. Buzo, A. H. Gray, Jr., R. M. Gray and J. D. Markel: Speech coding based upon vector quantization. IEEE Trans. on Acoustics Speech and Signal Proc. *28* (1980), 562–574.

[5] I. Csiszár and J. Körner: Information Theory. Academic Press, New York 1981.

[6] J. L. Doob: Stochastic Processes. J. Wiley, New York 1953.

[7] R. G. Gallager: Information Theory and Reliable Communication. J. Wiley, New York 1968.

[8] R. M. Gray and L. D. Davisson: Source coding without the ergodic assumption. IEEE Trans. Inform. Theory *20* (1974), 502–516.

[9] F. Liese and I. Vajda: Convex Statistical Distances. Teubner, Leipzig 1987.

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

[11] J. Nedoma: The capacity of a discrete channel. In: Trans. 1st Prague Conf. on Inform. Theory, Statist. Dec. Functions, Random Processes, Publ. House Czechosl. Acad. Sci., Prague 1957.
[12] D. S. Ornstein: Bernoulli shifts with the same entropy are isomorphic. Adv. in Math. *4* (1970), 338–352.
[13] C. E. Shannon: A mathematical theory of communication. Bell. System Tech. J. *27* (1948), 623–656.
[14] C. E. Shannon: Coding theorems for a discrete source with a fidelity criterion. IRE Nat. Convention Record, Part 4, 142–163.
[15] J. Šedivý: A DPCM vector quantizer for speech transmission. In: Circuit Theory and Design 85 (V. Zíma, J. Kvasil, eds.), Academia, Prague/North Holland, Amsterdam 1985, pp. 588–591.
[16] O. Šefl and I. Vajda: Spectrum-oriented source coding theory. Kybernetika *23* (1987), 6 (to appear).
[17] Š. Šujan: Ergodic theory, entropy and coding problems of information theory. Kybernetika *19* (1983), supplement, pp. 1–67.
[18] I. Vajda: Basic Theory of Information (to be published).
[19] J. Wolfowitz: Coding Theorems of Information Theory. Second edition. Springer-Verlag, Berlin 1964.

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.