Kybernetika

Karel Winkelbauer

On the coding theorem for decomposable discrete information channels. I

Kybernetika, Vol. 7 (1971), No. 2, (109)--124

Persistent URL: http://dml.cz/dmlcz/124249

KYBERNETIKA — VOLUME 7 (1971), NUMBER 2

On the Coding Theorem for Decomposable Discrete Information Channels I

KAREL WINKELBAUER

The entire paper is devoted to analyzing the problem of validity of the Coding Theorem and its Converse for a class of stationary discrete information channels which may be decomposed into ergodic components.

The stationary channels we shall deal with, are to possess the property that for them the concept of finite-dimensional code as defined by Wolfowitz in [11] makes sense; to be realistic, we must require for such channels to be without anticipation in time. The class of stationary channels satisfying both the requirements, is exactly consisting of those channels that were studied by the author in [6] and called there stationary channels with finite past history.

Our result will be stated as Theorem on the existence of ε-capacity since it gives a reply to the question on the asymptotic behaviour of the maximum length of n-dimensional ε-codes. The question was first touched by Parthasarathy in [4]; his answer to the problem was given for the class of channels which have additive noise. Here a necessary and sufficient condition is stated under which the class of stationary channels with ergodic components behaves in the desired manner. Decomposable channels with finitely many components were first investigated by Nedoma in his work [3]; in this paper we do not restrict ourselves to the class of channels with finite number of components, and we shall study the general case. Let us mention that the result may be transferred to the case of denumerable alphabets if the notion of entropy and similar concepts are properly generalized (as to that cf. [5]).

CONTENTS

PART I: 1. Statement of the Theorem

2. Proof of the Theorem

References to Part I

PART II*: 3. Proof of the First Inequality
4. Formula for Capacity
5. Some Properties of Decomposable Channels

References to Part II

* Part II is to be published in the next issue of this journal.

1. STATEMENT OF THE THEOREM

In this section we shall develop the concepts together with the terminology and notations that are needed in the statement of the theorem on ε-capacity.

Our basic notations will be as follows. If V and W are sets, then the Cartesian power W^V represents the class of all transformations which map V into W; in other words, W^V is the Cartesian product

$$(1.1)\qquad W^V = \prod_{u \in V} W_u\,, \quad \text{where} \quad W_u = W \quad \text{for all} \quad u \in V\,.$$

If $w \in W^V$, then the symbol $\{w_u\}_{u \in U}$ for $U \subset V$ will designate the partial mapping of U into W determined by the mapping w; especially, $\{w_u\}_{u \in V}$ may be used as an equivalent notation for the transformation w and will be sometimes called a parameter family of elements in W (with parameters in V).

Throughout this paper the symbol $\pi(S)$ means for a finite set S the number of elements in S, and for an infinite set we define $\pi(S) = +\infty$. As usual, the set of all integers will be denoted here by I.

For the sake of brevity, given an arbitrary countable (i.e. finite or denumerably infinite) non-empty set M, we shall accept the following notations: The class of all finite-dimensional cylinders in the space M^I (cf. (1.1)) will be denoted by $\boldsymbol{K}_M$; a *finite-dimensional cylinder* in M^I is, by definition, any set of the form

$$(1.2)\qquad \{z : z \in M^I,\ \{z_i\}_{i \in J} \in E\}\,, \quad E \subset M^J\,, \quad 0 < \pi(J) < +\infty\,,$$

where J is any finite non-empty subset of the index set I. The *coordinate-shift* transformation of the space M^I (onto itself) will be designated by T_M; it is defined by the property that

$$(1.3)\qquad (T_M z)_i = z_{i+1} \quad \text{for} \quad z \in M^I\,, \quad i \in I\,.$$

Finally, the symbol $\boldsymbol{F}_M$ will be used to denote the σ-algebra of sets in M^I generated by the class $\boldsymbol{K}_M$ of finite-dimensional cylinders.

In the ergodic theory of discrete dynamical systems we investigate the class of all probability measures in M^I which are invariant with respect to the transformation T_M; a probability measure μ defined on the class $\boldsymbol{F}_M$ is said to be *invariant* if $\mu T_M^{-1} = \mu$. In the sequel we designate the class of all invariant measures in M^I by $\mathcal{M}_M$.

An invariant measure μ is said to be decomposable if there are $\mu^1, \mu^2 \in \mathcal{M}_M$ such that $\mu = (1 - \alpha)\,\mu^1 + \alpha\mu^2$ for some α, $0 < \alpha < 1$ $(\mu^1 \neq \mu^2)$. An invariant measure which is not decomposable, is said to be *indecomposable*; the class of all indecomposable (invariant) measures in M^I will be denoted by $\mathscr{M}_M^*$. It is a well-known fact that the concept of indecomposability coincides with that of ergodicity.

In what follows we shall make use of some other notations which are chosen in accordance with [8] and [9]: we shall set

$$(1.4)\qquad [z] = \{\zeta : \zeta \in M^I,\ \{\zeta_j\}_{0 \leq j < n} = z\} \quad \text{for} \quad z \in M^n\,;$$
$$M^n = M^{\{1,2,\ldots,n\}} \quad (n = 1, 2, \ldots)\,;$$

$$(1.5)\qquad \mu_n(E) = \sum_{z \in E} \mu[z] \quad \text{for} \quad E \subset M^n\,;$$

$$(1.6)\qquad L_n(\varepsilon, \mu) = \min \{\pi(E) : E \subset M^n,\ \mu_n(E) > 1 - \varepsilon\}$$
$$\text{for} \quad 0 < \varepsilon < 1\,, \quad \mu \in \mathscr{M}_M$$

(cf. (1.2)); a finite-dimensional cylinder of the form $T_M^{-i}[z]$, where $i \in I$, $z \in M^n$ $(n = 1, 2, \ldots)$, is usually said to be an *elementary* one (with base z). The *entropy rate* of an invariant measure is defined as the limit

$$(1.7)\qquad \mathscr{H}(\mu) = \lim_n\,(1/n)\,H_n(\mu)\,, \quad \mu \in \mathscr{M}_M\,,$$

where we have set

$$(1.8)\qquad H_n(\mu) = -\sum_{z \in M^n} \mu[z] \log_2 \mu[z]\,.$$

The concept of indecomposable measure enables us to classify the points of the space M^I: a point $z \in M^I$ is, by definition, *regular* if and only if there is an indecomposable measure μ_z (i.e. $\mu_z \in \mathscr{M}_M^*$) such that

$$(1.9)\qquad \mu_z(K) = \lim_n\,(1/n) \sum_{j=0}^{n-1} \chi_K(T_M^j z) \quad \text{for all} \quad K \in \boldsymbol{K}_M\,,$$

i.e. the equality holds for any finite-dimensional cylinder; here χ_K means the characteristic function of the set K. Clearly, the measure μ_z is uniquely determined by the regular point z; let us mention that it suffices to require the equality to be valid only on the (countable) class of all elementary cylinders (cf. [9], Sec. 2). The set of all regular points in the space M^I is denoted in what follows by R_M. The important property of the set R_M that constitutes a basis of ergodic theory, states that $\mu(R_M) = 1$ for any $\mu \in \mathscr{M}_M$.

Throughout the entire paper we shall assume that we are given two (not necessarily distinct) finite non-empty sets A and B; they will be interpreted as alphabets of communications channels to be studied, namely A as the *output alphabet* and B as the *input alphabet*. We shall apply the notation introduced above to the Cartesian product $M = A \times B = AB$ (i.e. AB stands for $A \times B$ as the subscript) and employ for invariant measures lying in $\mathscr{M}_{AB}$ the generic symbol ω.

For the sake of simplicity and conformity in expressions given in the sequel, we shall denote, for $x \in A^I$, $y \in B^I$, by xy the element in the space $(A \times B)^I$ which satisfies the relations

$$(1.10) \qquad (xy)_i = (x_i, y_i) \quad \text{for} \quad i \in I ;$$

as evident, the mapping $(x, y) \to xy$ establishes a one-to-one correspondence between the measurable spaces $(A^I \times B^I, \mathbf{F}_A \times \mathbf{F}_B)$ and $((A \times B)^I, \mathbf{F}_{AB})$. The same notation xy will be employed also in case that $x \in A^n$, $y \in B^n$; here $(xy)_i = (x_i, y_i)$, $xy \in (A \times B)^n$, $i = 1, 2, \ldots, n$. If $\omega \in \mathscr{M}_{AB}$, we define the *marginal* measures $\omega^A \in \mathscr{M}_A$ and $\omega^B \in \mathscr{M}_B$ by

$$(1.11) \qquad \omega^A(E) = \omega\{xy : x \in E\} \quad \text{for} \quad E \in \mathbf{F}_A ;$$
$$\omega^B(F) = \omega\{xy : y \in F\} \quad \text{for} \quad F \in \mathbf{F}_B .$$

The *information rate* $I(\omega)$ is associated with any (stationary double source) $\omega \in \mathscr{M}_{AB}$ according to the definition (cf. (1.4))

$$(1.12) \qquad I(\omega) = \lim_n (1/n) \sum_{x \in A^n, y \in B^n} \omega[xy] \log_2 \frac{\omega[xy]}{\omega^A[x]\, \omega^B[y]} ;$$

the existence of the limit is a well-known result of information theory.

By a (discrete communication) *channel* we shall mean a parameter family

$$(1.13) \qquad \nu = \{\nu_y\}_{y \in B^I}$$

of probability measures on the measurable space $(A^I, \mathbf{F}_A)$ which satisfies the following measurability condition: $\nu_y(E)$ as a function of parameter y is measurable on the space $(B^I, \mathbf{F}_B)$ for every measurable set E in A^I, i.e. $E \in \mathbf{F}_A$.

If necessary, we shall refer to a (measurable) family (1.13) as a channel with output alphabet A and input alphabet B; if convenient, we shall make use of the symbol $\nu(. \mid y)$ to denote the probability measure ν_y for $y \in B^I$.

Our interest will be devoted to the class of stationary channels: a (measurable) family (1.13) is, by definition, a *stationary* channel if

$$(1.14) \qquad \nu(T_A E \mid T_B y) = \nu(E \mid y) \quad \text{for} \quad E \in \mathbf{F}_A , \quad y \in B^I$$

(cf. (1.3)). To have a concise notation, we shall designate the class of all stationary channels (with the given alphabets A, B) by $\mathscr{N}(A \mid B)$ or $\mathscr{N}$.

Elements lying in the class $\mathscr{M}_B$ are called *stationary inputs* of channels belonging to the class $\mathscr{N}(A \mid B)$. If μ is a stationary input and ν a stationary channel (i.e. $\mu \in \mathscr{M}_B$, $\nu \in \mathscr{N}(A \mid B)$), then the symbol $\nu\mu$ will mean the (stationary double source, viz.) invariant measure in $\mathscr{M}_{AB}$ which is defined by the equation

$$(1.15)\qquad \nu\mu(G) = \int \nu_y(G_y)\,\mathrm{d}\mu(y)\,, \quad \text{where} \quad G_y = \{x : xy \in G\}\,, \quad G \in \mathbf{F}_{AB}\,.$$

As pointed out in the introductory part of this paper, the theorem on capacity we are going to state in this section, will concern the class of stationary channels with finite past history. Perhaps the nomenclature chosen by the author in [6] may be regarded as a misnomer compared with the concept of past history as developed by Wolfowitz in [11]; it could be better and more precise to call such channels as defined below, channels with finite input memory. Nevertheless, we will keep on the original terminology; it is because the finite-memory concept is well-established in literature for a narrower class of discrete channels.

A stationary channel with *finite past history* is defined as a parameter family (1.13) of probability measures which satisfies the stationarity condition (1.14) and the following finite-past-history condition

$$(1.16)\qquad \begin{gathered} \nu([x] \mid y') = \nu([x] \mid y) \quad \text{for} \quad x \in A^n\,, \quad y, y' \in B^I \\ \text{if} \quad y'_i = y_i \quad \text{for} \quad -m \leqq i < n \quad (n = 1, 2, \ldots)\,, \end{gathered}$$

for some non-negative integer m. The least non-negative integer m for which the latter condition is satisfied, will be denoted by $m(\nu)$; it represent the *duration* of the past history (i.e. of the input memory). The condition (1.16) implies the validity of the measurability condition required for channels. We may denote the class of all stationary channels with finite past history, given the alphabets A, B, by $\mathscr{N}_{\text{f.past}}$ or $\mathscr{N}_{\text{f.past}}(A \mid B)$.

The following notations make sense only for channels satisfying condition (1.16): we shall set

$$(1.17)\qquad \begin{gathered} \nu_n(E \mid y) = \sum_{x \in E} \nu([x] \mid \eta) \quad \text{for} \quad E \subset A^n \\ \text{if} \quad \eta \in T^m[y]\,, \quad y \in B^{m+n}\,; \quad m \geqq m(\nu)\,; \end{gathered}$$

$$(1.18)\qquad \begin{gathered} S_n(\psi; \varepsilon, \nu) = \pi\{y : y \in B^{m(\nu)+n},\ \nu_n(\psi^{-1}\{y\} \mid y) > 1 - \varepsilon\} \\ \text{for} \quad \psi \in (B^{m(\nu)+n})^{A^n}\,; \end{gathered}$$

$$(1.19)\qquad \begin{gathered} S_n(\varepsilon, \nu) = \max\{S_n(\psi; \varepsilon, \nu) : \psi \in (B^{m(\nu)+n})^{A^n}\} \\ \text{for} \quad 0 < \varepsilon < 1 \quad (n = 1, 2, \ldots)\,, \quad \nu \in \mathscr{N}_{\text{f.past}}\,. \end{gathered}$$

The notations are chosen in accordance with [6] and [8].

Remark 1. Given $\psi : A^n \to B^{m+n}$ $(m = m(\nu))$, then the parameter family

$$\{(y, \psi^{-1}\{y\})\}_{y \in Y(\psi)}\,,$$

where $Y(\psi)$ is the set of those parameters y in B^{m+n} that satisfy the inequality $\nu_n(\psi^{-1}\{y\} \mid y) > 1 - \varepsilon$, is nothing else than an n-dimensional ε-code of length $N = S_n(\psi; \varepsilon, \nu)$ (in [11] a code (n, N, λ) with $\lambda = 1 - \varepsilon$ is such a code for which the latter inequality is taken in the non-strict sense, i.e. with $\geqq$); the number $S_n(\varepsilon, \nu)$ represents the maximum length of n-dimensional ε-codes.

Remark 2. The Coding Theorem is concerning the asymptotic behaviour of the sequence

$$\frac{1}{n} \log_2 S_n(\varepsilon, \nu)\,, \quad n = 1, 2, \ldots;$$

taking in the above definitions any integer $m \geqq m(\nu)$ instead of $m(\nu)$ itself, and denoting the corresponding maximum length of extended n-dimensional ε-codes by $S'_n(\varepsilon, \nu)$, then the asymptotic behaviour of the sequence

$$\frac{1}{n} \log_2 S'_n(\varepsilon, \nu)\,, \quad n = 1, 2, \ldots,$$

must be the same since

$$S_n(\varepsilon, \nu) \leqq S'_n(\varepsilon, \nu) \leqq S_n(\varepsilon, \nu)\,[\pi(B)]^{m-m(\nu)}\,,$$

as immediately follows from the definitions.

In this paper a stationary channel ν with finite past history is said to be *strongly stable* if there is a real number C such that

$$\text{(1.20)} \qquad \frac{1}{n} \log_2 S_n(\varepsilon, \nu) \to C\,, \quad n \to \infty\,; \quad 0 < \varepsilon < 1\,;$$

i. e. the limit exists and does not depend on ε.

By definition, an *ergodic* channel is a stationary channel with finite past history such that, for any ergodic input μ, i.e. $\mu \in \mathscr{M}^*_B$, the (double source) $\nu\mu$ is an ergodic (i.e. indecomposable) measure, i.e. $\nu\mu \in \mathscr{M}^*_{AB}$. Denoting by $\mathscr{N}_{\text{erg}}$ (or $\mathscr{N}_{\text{erg}}(A \mid B)$) the class of all ergodic channels (with the alphabets given), we shall designate the class of those ergodic channels which are at the same time strongly stable, by $\mathscr{N}^*_{\text{erg}}$ (or $\mathscr{N}^*_{\text{erg}}(A \mid B)$); the latter class will play a fundamental role in our investigations. Summarizing in symbols:

$$\text{(1.21)} \qquad \begin{aligned} \mathscr{N}^*_{\text{erg}} &= \{\nu : \nu \in \mathscr{N}_{\text{erg}},\ \nu \text{ satisfies } (1.20)\}\,; \\ \mathscr{N}_{\text{erg}} &= \{\nu : \nu \in \mathscr{N}_{\text{f.past}},\ \nu\mu \in \mathscr{M}^*_{AB} \text{ for any } \mu \in \mathscr{M}^*_B\}\,; \\ \mathscr{N}_{\text{f.past}} &= \{\nu : \nu \in \mathscr{N},\ \nu \text{ satisfies } (1.16) \text{ for some } m\}\,. \end{aligned}$$

The notations employed are partly chosen in accordance with [8] (cf., in particular, Sec. 1.6).

Remark 3. Stationary channels with finite past history are, in general, non-ergodic. Examples were first given by Nedoma in [2]. Especially, any non-degenerate decomposable channel as defined in the sequel, represents an example of non-ergodic channel. On the other hand, the well-known finite-memory channels constitute a class of ergodic channels which are strongly stable: the condition of strong stability as given in (1.20) is nothing else than a restatement of the Coding Theorem together with its Strong Converse in the sense of Wolfowitz. A class of channels which are, in general, not of finite memory but which may be at the same time ergodic, is formed by channels with additive noise as studied by Parthasarathy in [4].

As well-known, for any stationary channel ν the *ergodic* and *stationary* (information-rate, or transmission-rate) *capacities* are defined by the formulas

$$\begin{aligned}(1.22)\qquad \mathscr{C}_{\mathrm{erg}}(\nu) &= \sup\{I(\nu\mu) : \mu \in \mathscr{M}_{\mathrm{erg}}\}\,, \quad \mathscr{M}_{\mathrm{erg}} = \mathscr{M}_B^*\,,\\ \mathscr{C}_{\mathrm{st}}(\nu) &= \sup\{I(\nu\mu) : \mu \in \mathscr{M}_{\mathrm{st}}\}\,, \quad \mathscr{M}_{\mathrm{st}} = \mathscr{M}_B\,,\end{aligned}$$

as the corresponding suprema of information rates $I(\nu\mu)$; both the capacities coincide with one another (as shown independently by Jacobs, and by Parthasarathy), and we shall denote their common value by $\mathscr{C}(\nu)$:

$$(1.22)\qquad \mathscr{C}(\nu) = \mathscr{C}_{\mathrm{erg}}(\nu) = \mathscr{C}_{\mathrm{st}}(\nu) \quad \text{for} \quad \nu \in \mathscr{N}\,.$$

An immediate consequence of Theorem 3, Sec. 21 in [6] is the fact that (cf. (1.20))

$$(1.23)\qquad \frac{1}{n}\log_2 S_n(\varepsilon, \nu) \to \mathscr{C}(\nu)\,, \quad n \to \infty\,; \quad 0 < \varepsilon < 1\,; \quad \nu \in \mathscr{N}^*_{\mathrm{erg}}\,;$$

i.e. (1.23) is valid for any strongly stable ergodic channel ν.

A parameter family $\{\nu^\alpha\}_{\alpha\in\mathscr{A}}$ of channels with parameters in a measurable space $(\mathscr{A}, \boldsymbol{A})$ is said to be *measurable* if the function $\nu^\alpha(E \mid y)$ as a function of parameter α is measurable on $(\mathscr{A}, \boldsymbol{A})$ for every $E \in \boldsymbol{F}_A$ and $y \in B^I$. In what follows any probability measure ξ defined on $\boldsymbol{A}$ will be referred to as a *channel distribution* of the (measurable) family $\{\nu^\alpha\}_{\alpha\in\mathscr{A}}$.

Let $\{\nu^\alpha\}_{\alpha\in\mathscr{A}}$ be a measurable family of strongly stable ergodic channels having a channel distribution ξ. It is an easy consequence of definition (1.18) that $S_n(\psi; \varepsilon, \nu^\alpha)$ as a function of parameter α is measurable so that from (1.23) it follows that capacity $\mathscr{C}(\nu^\alpha)$ as a function of parameter α is measurable as well. We shall set

$$(1.24)\qquad \tilde{c}(\theta, \xi) = \inf\{r : \xi\,\{\alpha : \mathscr{C}(\nu^\alpha) \leqq r\} \geqq \theta\}\,, \quad 0 < \theta \leqq 1\,.$$

As immediately follows from definition (1.24), $\tilde{c}(\theta, \xi)$ as a function of parameter θ is monotonically increasing in the open interval (0, 1) of real numbers (the quantity $\tilde{c}(\theta, \xi)$ represents the quantile of order θ of the probability distribution of the random variable $\mathscr{C}(\nu^\alpha)$). The quantity defined by (1.24) will play a fundamental role in the statement of the theorem on ε-capacity given below.

Regularity condition. Let ξ be a channel distribution of a measurable family $\{\nu^\alpha\}_{\alpha\in\mathscr{A}}$ of strongly stable ergodic channels, let F be the probability distribution function of the random variable $\mathscr{C}(\nu^\alpha)$, i.e.

$$F(t) = \xi\{\alpha : \mathscr{C}(\nu^\alpha) \leqq t\}\,, \quad t \text{ real}\,,$$

and let D be the (countable, maybe empty) set of all discontinuity points of the distribution function F, i.e.

$$D = \{t : F(t + 0) - F(t - 0) > 0\} .$$

Then the channel distribution ξ is said to be *regular* if and only if both (1) the condition that,

(1.25) for any $t < r$ and for any $\theta > F(r - 0)$, there is some $\mu \in \mathscr{M}_B^*$ (i.e. an ergodic input) such that

$$\xi\{\alpha : I(\nu^\alpha \mu) < t\} \leqq \theta$$

(cf. (1.12), (1.15)) is satisfied for all those $r \in D$ having the property that either $F(r - 0) = 0$, or $F(r - 0) = \gamma > 0$ and $\tilde{c}(\gamma, \xi) < r$, and (2) the condition that,

(1.26) for any real t, either there is some $\mu \in \mathscr{M}_B^*$ such that

$$\xi\{\alpha : I(\nu^\alpha \mu) \leqq t\} < F(r - 0) ,$$

or

$$\xi\{\alpha : I(\nu^\alpha \mu) \leqq t, \mathscr{C}(\nu^\alpha) < r\} \geqq F(r - 0) \quad \text{for all} \quad \mu \in \mathscr{M}_B^*$$

is fulfilled for every r lying either in D and such that $F(r - 0) = \gamma > 0$, $\tilde{c}(\gamma, \xi) = r$, or in some countable set Q of nonnegative real numbers such that $\tilde{c}(F(r), \xi) = r \notin D$ for every $r \in Q$, and such that Q is dense in the numerical set

$$\{r : r \notin D, r = \tilde{c}(\theta, \xi) \quad \text{for some} \quad \theta(0 < \theta \leqq 1) .$$

The sense of the regularity condition just stated will be fully cleared up in the subsequent analysis (in the subsequent sections). Let us point out that examples may be found in which the regularity condition is not satisfied.

In the sequel a stationary channel ν is said to be *decomposable* if there is a measurable family $\{\nu^\alpha\}_{\alpha \in \mathscr{A}}$ of channels $\nu^\alpha \in \mathscr{N}_{\mathrm{f.past}}$ with parameters in $(\mathscr{A}, \boldsymbol{A})$ such that $m(\nu^\alpha) \leqq m(\alpha \in \mathscr{A})$ for some integer m, and

$$(1.27) \qquad \nu(E \mid y) = \int_{\mathscr{A}} \nu^\alpha(E \mid y) \, d\xi(\alpha) , \quad E \in \boldsymbol{F}_A , \quad y \in B^I , \quad \text{symbolically}$$

$$\nu = \int_{\mathscr{A}} \nu^\alpha \, d\xi(\alpha)$$

for some channel distribution ξ of the family $\{\nu^\alpha\}_{\alpha \in \mathscr{A}}$; for our purposes it will be convenient to have an alternative notation and to put

$$(1.28) \qquad \tilde{c}(\theta, \nu) = \tilde{c}(\theta, \xi) , \quad 0 < \theta \leqq 1 ,$$

replacing the symbol of channel distribution by that of decomposable channel itself.

It is clear from the definition that a decomposable channel ν is of finite past history, i.e. $\nu \in \mathscr{N}_{\mathrm{f.past}}$ (cf. (1.21)). Our main interest will be directed to channels decomposable into strongly stable ergodic components, i.e. for which the components ν^α are lying in $\mathscr{N}^*_{\mathrm{erg}}$; for the latter channels we shall now state the main theorem of this paper, viz. the theorem on the existence of ε-capacity.

Theorem on ε-Capacity. *Let $\{\nu^\alpha\}_{\alpha\in\mathscr{A}}$ be a measurable family of strongly stable ergodic channels, with a channel distribution ξ and such that the duration of past history $m(\nu^\alpha)$ of channels ν^α is bounded a.s.*, i.e.

$$\xi\{\alpha : m(\nu^\alpha) \leqq m\} = 1 \quad \text{for some nonnegative integer } m .$$

Then a necessary and sufficient condition that, for any real number r less than the essential supremum of $\mathscr{C}(\nu^\alpha)$, i.e.

$$r < \text{ess. sup}\,\{\mathscr{C}(\nu^\alpha) : \alpha \in \mathscr{A}[\xi]\} ,$$

there be a countable subset E_r of the open interval $(0,1)$ of real numbers such that the limit

$$\lim_n \frac{1}{n} \log_2 S_n(\varepsilon, \nu^{[r]})$$

exist and equal $\bar{c}(\varepsilon, \nu^{[r]})$ for all $\varepsilon \notin E_r$, $0 < \varepsilon < 1$, where $\nu^{[r]}$ is the decomposable channel defined by

$$(1.29) \qquad \nu^{[r]} = \frac{1}{\xi(\Lambda_r)} \int_{\Lambda_r} \nu^\alpha d\xi(\alpha) ,$$
$$\Lambda_r = \{\alpha : \mathscr{C}(\nu^\alpha) \geqq r\} ,$$

is that the channel distribution ξ be regular.

Especially, if ν is a decomposable channel with strongly stable ergodic components $\nu^\alpha(\alpha \in \mathscr{A})$ the channel distribution ξ of which is regular, then the limit

$$(1.30) \qquad C_\varepsilon(\nu) = \lim_n \frac{1}{n} \log_2 S_n(\varepsilon, \nu)$$

exists and equals

$$(1.31) \qquad \inf\{r : \xi\{\alpha : \mathscr{C}(\nu^\alpha) \leqq r\} \geqq \varepsilon\}$$

for all ε, $0 < \varepsilon < 1$, except a countable set of ε's.

A much more general class of strongly stable ergodic channels than the class of finite-memory channels is that of channels with additive noise whose noise distributions are ergodic. A channel with *additive noise*, as defined by Parthasarathy in [4], the noise distribution of which is an invariant measure $\mu \in \mathscr{M}_A$, is a channel ν satisfying the condition

$$(1.32) \qquad \nu(E \mid y) = \mu(E - y) , \quad E \in \mathbf{F}_A , \quad y \in B^I ,$$

provided that A is an additive Abelian group, and that both the alphabets A and B are identical with each other; the additive group operation in A^I is defined by coordinates, and $E - y$ is the set of $x \in A^I$ such that $x + y \in E$. If the noise distribution μ is ergodic, the channel ν is said to be with *additive ergodic noise.*

A channel with additive noise as defined with the aid of an invariant measure $\mu \in \mathscr{M}_A$ by (1.32), is always stationary and with zero past history. It may easily be shown that a channel with additive ergodic noise is ergodic; it is, at the same time, strongly stable as follows from Theorem 3.1 stated in [4].

Corollary. *If ν is a channel decomposable into components with additive ergodic noise each, then the limit $C_\varepsilon(\nu)$ given by (1.30) exists and equals the quantity (1.31) for $0 < \varepsilon < 1$, except a countable set of ε's.*

We shall see that the statement of the Corollary to the main theorem follows from the fact that any channel of described type behaves regularly in the sense of conditions (1.25), (1.26).

The next section is devoted to the proof of the theorem on the existence of ε-capacity. The proof will be based upon a group of theorems which may be of some interest taken only for themselves. The group of theorems just quoted will be proved in Part II of this paper.

2. PROOF OF THE THEOREM

Before proceeding to the proof of the main theorem, we must develop some auxiliary notations. First of all, we shall make use of the following concise notations:

$$(2.1) \qquad c(\varepsilon, \nu) = \liminf_n \frac{1}{n} \log S_n(\varepsilon, \nu)\,; \quad \log = \log_2\,;$$

$$\bar{c}(\varepsilon, \nu) = \limsup_n \frac{1}{n} \log S_n(\varepsilon, \nu)\,, \quad \nu \in \mathscr{N}_{\mathrm{f.past}}\,.$$

We shall set

$$(2.2) \qquad c^*(\theta, \nu) = \sup\{c^*(\theta, \nu\mu) : \mu \in \mathscr{M}_B^*\}\,,$$

$$c^*(\theta, \nu\mu) = \sup\{r : \nu\mu\{I_{xy} \geqq r\} \geqq 1 - \theta\}\,, \quad 0 < \theta < 1\,, \quad \nu \in \mathscr{N}_{\mathrm{f.past}}\,,$$

where

$$\nu\mu\{I_{xy} \geqq r\} = \nu\mu\{xy : xy \in R_{AB},\, I_{xy} \geqq r\}, \quad I_z = I(\omega_z)\,, \quad z \in R_{AB}\,.$$

The following theorem will be proved in Section 3.

Theorem 1. *For any stationary channel ν with finite past history*, i.e. $\nu \in \mathscr{N}_{\mathrm{f.past}}$, *the following inequality holds:*

$$c^*(\theta, \nu) \leqq c(\varepsilon, \nu) \quad \text{for} \quad 0 < \theta < \varepsilon < 1\,.$$

The capacity $C(\nu)$ of a channel $\nu \in \mathcal{N}_{\mathrm{f.past}}$ as introduced in [6], may be defined by the formula

$$C(\nu) = \lim_{\varepsilon \to 0} c(\varepsilon, \nu)\,.$$

For the capacity we obtain the relation (proved in Section 4):

Theorem 2. *If ν is a stationary channel with finite past history,* i.e. $\nu \in \mathcal{N}_{\mathrm{f.past}}$, *then its capacity $C(\nu)$ has the following property:*

$$C(\nu) = \lim_{\theta \to 0} c^*(\theta, \nu) = c^*(0+, \nu)\,.$$

As an immediate consequence of the well-known facts of ergodic theory we obtain that

$$\nu\mu_z\{I_{xy} \geqq r\} = \nu_y\{I_{xy} \geqq r\} \quad \text{a.s.} \quad [\mu_z]\,, \quad z \in R_B \quad (\nu \in \mathcal{N})\,; \tag{2.3}$$

cf. (1.9). It is because I_{xy} is an invariant function (with respect to the automorphism T_{AB}; cf. (1.3)) and so is $\nu_y\{I_{xy} \geqq r\}$; on the other hand, μ_z for $z \in R_B$ is an indecomposable measure so that $\nu_y\{I_{xy} \geqq r\}$ as a function of y must be constant a.s.

In the remainder of this section the letter ν will denote a decomposable channel with strongly stable ergodic components $\{\nu^\alpha\}_{\alpha \in \mathcal{A}}$ the channel distribution of which will be designated by the letter ξ; cf. (1.27). This convention enables us to simplify our notations in suppressing the letter ν, e.g. $c^*(\theta)$, $\tilde{c}(\theta)$ will stand for $c^*(\theta, \nu)$, $\tilde{c}(\theta, \nu)$; cf. (1.28), (1.24).

Lemma 2.1. *There is a countable subset E of the interval $(0,1)$ such that*

$$c^*(\theta) \leqq \tilde{c}(\theta) \quad \text{for all} \quad \theta \notin E\,, \quad 0 < \theta < 1\,.$$

Proof. Under the assumptions made for the decomposable channel together with the ergodicity of its components we have

$$\nu\mu\{I_{xy} \geqq r\} = \xi\{I(\nu^\alpha\mu) \geqq r\}\,, \quad \mu \in \mathcal{M}_B^*\,,$$

since

$$\nu\mu(G) = \int \nu^\alpha\mu(G)\, d\xi(\alpha) \quad \text{for any} \quad G \in \mathbf{F}_{AB}\,,$$

and $\nu^\alpha\mu\{I_{xy} \geqq r\} = 0$ or 1 according to $I(\nu^\alpha\mu) < r$ or $\geqq r$. Using the latter equality, we easily find on the basis of the definitions of c^* and $\tilde{c}$ that

$$c^*(\theta_1) \leqq \tilde{c}(\theta_2) \quad \text{for} \quad \theta_1 < \theta_2\,, \tag{2.4}$$

since the contrary would lead to the contradictory inequalities

$$\tilde{c}(\theta_2) < r_1 < r_2 < c^*(\theta_1) \quad \text{for some} \quad r_1, r_2\,.$$

From (2.4) it immediately follows that $c^*(\theta) \leqq \tilde{c}(\theta)$ must be valid at least at all continuity points θ of the function $\tilde{c}$.

We shall now state two theorems the proofs of which are postponed into Section 5.

Theorem 3. *If ν is a decomposable channel with strongly stable ergodic components, then the following inequality is satisfied:*

$$\bar{c}(\varepsilon, \nu) \leqq \tilde{c}(\theta, \nu) \quad \text{for} \quad 0 < \varepsilon < \theta < 1\,.$$

Theorem 4. *If ν is a decomposable channel with strongly stable ergodic components, then*

$$\lim_{\theta \to 1} c^*(\theta, \nu) = \text{ess. sup}\,\{\mathscr{C}(\nu^\alpha) : \alpha \in \mathscr{A}[\xi]\}\,.$$

Let us define another auxiliary quantity by

$$\tilde{c}'(\theta) = \tilde{c}'(\theta, \nu) = \sup\{r : \xi\{\alpha : \mathscr{C}^\alpha \geqq r\} \geqq 1 - \theta\}\,, \quad 0 < \theta < 1\,, \tag{2.5}$$

where we have set

$$\mathscr{C}^\alpha = \mathscr{C}(\nu^\alpha)\,, \quad \alpha \in \mathscr{A}\,. \tag{2.6}$$

It is an immediate consequence of the definition that

$$\tilde{c}'(\theta_1) \leqq \tilde{c}(\theta_2) \quad \text{for} \quad 0 < \theta_1 < \theta_2 < 1\,; \tag{2.7}$$

it is because $\tilde{c}'$ is monotonically increasing similarly as $\tilde{c}$.

Now we shall prove a basic lemma which shows the sense of the regularity condition stated in Section 1.

Lemma 2.2. *If the channel distribution ξ is regular, then there is a countable subset E of the interval $(0,1)$ such that*

$$c^*(\theta) \geqq \tilde{c}(\theta) \quad \text{for all} \quad \theta \notin E\,, \quad 0 < \theta < 1\,.$$

Proof. I. Let E be a countable subset of the interval $(0, 1)$ having the property that, for all $\theta \notin E$, $0 < \theta < 1$,

$$c^*, \tilde{c} \text{ are continuous at } \theta, \text{ and } \tilde{c}'(\theta) = \tilde{c}(\theta)\,; \tag{1}$$

the existence of E follows from the elementary properties of the functions considered, and especially from (2.7).

Let $\gamma (0 < \gamma < 1)$ be such that $\gamma \notin E$ and $\tilde{c}(\gamma) = r \notin D$, where D is the set of all discontinuities of the distribution function F of the random variable $\mathscr{C}^\alpha$. Owing to (1), we easily obtain that

$$\xi\{\alpha : \mathscr{C}^\alpha < r\} = \gamma\,.$$

Let us define an auxiliary decomposable channel ν' by

(2) $$\nu' = \frac{1}{\gamma} \int_A \nu^\alpha \, d\xi(\alpha) ,$$
$$A = \{\alpha : \mathscr{C}^\alpha < r\} .$$

Making use of Theorem 4, we get the equalities

$$\lim_{\theta \to 1} c^*(\theta, \nu') = \text{ess. sup} \{\mathscr{C}^\alpha : \alpha \in A[\xi]\} = r$$

because of (1). On the other hand, it again follows from (1) and $\gamma \notin E$ that

$$\lim_{\theta \to 1} c^*(\theta, \nu') = \lim_{\theta \to 1} c^*(\theta \, . \, \gamma, \nu) = c^*(\gamma -) = c^*(\gamma) ,$$

i.e. $c^*(\gamma) = r = \tilde{c}(\gamma)$; cf. Lemma I in [10].

II. Let θ $(0 < \theta < 1)$ be such that $\tilde{c}(\theta) = r \in D$, and that the point $\gamma = F(r - 0)$ is a point of continuity for $\tilde{c}$, i.e. $r = \tilde{c}(\gamma +) = \tilde{c}(\gamma)$. (Let us point out that $\tilde{c}$ is continuous from the left.) Since $\xi\{\alpha : \mathscr{C}^\alpha < r\} = \gamma$, we may define channel ν' by (2) as above and find by the same reasoning that $c^*(\gamma) \geqq \tilde{c}(\gamma) = r$; hence

$$c^*(\theta) \geqq c^*(\gamma) \geqq r = c^*(\theta) \quad \text{for} \quad \gamma \leqq \theta \leqq \gamma' = F(r + 0) .$$

Because of the fact that $\tilde{c}(\theta) \neq r$ outside the interval $[\gamma, \gamma']$ we have proved the implication

(3) $$\text{if} \quad \tilde{c}(\theta) = r \quad \text{then} \quad c^*(\theta) \geqq \tilde{c}(\theta)$$

for the case considered.

III. Let θ $(0 < \theta < 1)$ be such $\tilde{c}(\theta) = r \in D$, and that $\gamma = F(r - 0)$ is a discontinuity point of $\tilde{c}$, i.e. $\tilde{c}(\gamma) < \tilde{c}(\gamma +) = r$ for $\gamma > 0$, or $\gamma = 0$. Applying the regularity condition (1.25), we immediately obtain according to (2.2) that

$$c^*(\theta) \geqq r = \tilde{c}(\theta) \quad \text{for} \quad \gamma < \theta \leqq \gamma' = F(r + 0) .$$

Hence it follows the implication

(4) $$\text{if} \quad \tilde{c}(\theta) = r \quad \text{then} \quad c^*(\theta) \geqq \tilde{c}(\theta)$$

for the discontinuity point r of F having the above properties.

Summarizing the facts obtained during the proof (cf. especially (3) and (4)), we have shown that

$$c^*(\theta) \geqq \tilde{c}(\theta) \quad \text{for} \quad \theta \notin E , \quad 0 < \theta < 1 ,$$

which is the desired result.

Proof of the Theorem on ε-capacity. I. Under the assumptions made in the theorem, let us assume first that the channel distribution ξ is regular. Owing to

 Lemma 2.1 and Lemma 2.2, there is a countable set E of numbers lying inside the interval $(0,1)$ such that

$$\tilde{c} \text{ is continuous at } \theta\,, \quad c^*(\theta) = \tilde{c}(\theta) \quad \text{for} \quad \theta \notin E\,, \quad 0 < \theta < 1\,.$$

Applying Theorem 1 and Theorem 3 to channel $\nu^{[r]}$ as defined by (1.29), and making use of the equalities (cf. Lemma II in [10])

$$(1) \qquad c^*(\theta, \nu^{[r]}) = c^*(\theta_r + \theta\,.\,(1 - \theta_r))\,,$$
$$\tilde{c}(\theta, \nu^{[r]}) = \tilde{c}(\theta_r + \theta\,.\,(1 - \theta_r))\,,$$

where $\theta_r = 1 - \xi(\Lambda_r)$, we obtain the relations

$$\tilde{c}(\theta, \nu^{[r]}) = c^*(\theta, \nu^{[r]}) \leqq c(\varepsilon, \nu^{[r]}) \leqq \tilde{c}(\varepsilon, \nu^{[r]}) \leqq \tilde{c}(\theta', \nu^{[r]})$$
$$\text{for} \quad \theta < \varepsilon < \theta'\,, \quad \theta'' \notin E\,, \quad \text{where} \quad \theta(1 - \theta_r) + \theta_r = \theta''\,;$$

hence,

$$\lim_{\theta \to \varepsilon} \tilde{c}(\theta, \nu^{[r]}) = c(\varepsilon, \nu^{[r]}) = \tilde{c}(\varepsilon, \nu^{[r]}) \quad \text{for} \quad \varepsilon(1 - \theta_r) + \theta_r = \varepsilon' \notin E$$

because of the continuity of $\tilde{c}$ at ε', which shows the sufficiency of the regularity condition.

II. Let us make the assumption that the channel distribution ξ is *not* regular. Using the notations of the statement of the regularity condition given in Section 1, we may assert that either (a) there is an $r \in D$ such that

$$(2) \qquad c^*(\theta) < \tilde{c}(\theta) = r \quad \text{for} \quad \gamma < \theta \leqq \theta_0\,, \quad \gamma = F(r - 0)$$

for some $\theta_0 \leqq F(r + 0)$, or (b) there is an $r \notin D$ such that $\tilde{c}(F(r)) = \tilde{c}'(F(r)) = r$, and that

$$(3) \qquad c^*(\gamma-) = c^*(\gamma) = c^*(\gamma+) < \tilde{c}(\gamma) = r \quad \text{for} \quad \gamma = F(r)\,.$$

Taking the channel $\nu^{[r]}$ defined by (1.29), applying Theorem 2 to channel $\nu^{[r]}$, and using the above equalities (1), we obtain

$$c^*(\gamma+) = \lim_{\theta \to 0} c^*(\theta, \nu^{[r]}) = \lim_{\varepsilon \to 0} c(\varepsilon, \nu^{[r]})\,.$$

The latter equalities together with (1) and (2), or (3) show that the assertion of the theorem cannot be valid for r just considered, which shows the necessity of the regularity condition, Q.E.D.

Rephrasing the terminology of Section 1, we may say that a decomposable channel is *regular* if the channel distribution defining the channel is regular.

The assertion of the Corollary to Theorem on ε-capacity for channels with components having additive ergodic noise immediately follows from the subsequent theorem:

Theorem 5. *Any channel decomposable into components with additive ergodic noise is regular.*

The proof of Theorem 5 will be given in Section 5 (cf. Part II of this paper to be published in the next issue).

(Received November 13, 1970.)

REFERENCES

[1] A. I. Khinchin: Mathematical foundations of information theory. Dover Publications, New York 1957.

[2] J. Nedoma: Capacity of a discrete channel. Transact. First Prague Conf. on Inform. Theory etc. Prague 1957, 143—181.

[3] J. Nedoma: On non-ergodic channels. Transact. Second Prague Conf. on Inform. Theory etc. Prague 1960, 363—395.

[4] K. R. Parthasarathy: Effective entropy rate and transmission of information through channels with additive random noise. Sankhya *A 25* (1963), 75—84.

[5] V. A. Rohlin: New progress in the theory of transformation with invariant measure (in Russian). Usp. Mat. Nauk *15* (1960), 3—26.

[6] K. Winkelbauer: Channels with finite past history. Transact. Second Prague Conf. on Inform. Theory etc. Prague 1960, 685—831.

[7] K. Winkelbauer: On discrete information sources. Transact. Third Prague Conf. on Inform. Theory etc. Prague 1964, 765—830.

[8] K. Winkelbauer: Axiomatic definition of channel capacity and entropy rate. Transact. Fourth Prague Conf. on Inform. Theory etc. Prague 1967, 661—705.

[9] K. Winkelbauer: On the asymptotic rate of non-ergodic information sources. Kybernetika *6* (1970), 2, 127—148.

[10] K. Winkelbauer: On the regularity condition for decomposable communication channels. Kybernetika *7* (1971), 4 (in press).

[11] J. Wolfowitz: Coding theorems of information theory. Second ed. Berlin 1964.

K větě o kódování pro rozložitelné diskrétní sdělovací kanály I

KAREL WINKELBAUER

Příspěvek je věnován analýze problému platnosti věty o kódování pro třídu stacionárních diskrétních kanálů, jež lze rozložit na ergodické složky. Výsledek je formulován jako teorém o existenci ε-kapacity, ježto podává odpověď na otázku o asymptotickém chování maximální délky n-rozměrných ε-kódů pro každé ε zvlášť.

Doc. RNDr. Mg. Mat. Karel Winkelbauer DrSc., Ústav teorie informace a automatizace ČSAV (Institute of Information Theory and Automation — Czechoslovak Academy of Sciences), Vyšehradská 49, Praha 2.