KYBERNETIKA — VOLUME 17 (1981), NUMBER 3

ON THE CAPACITY OF ASYMPTOTICALLY MEAN STATIONARY CHANNELS

ŠTEFAN ŠUJAN

The properties of the information quantile capacity, the Shannon capacity, and the operational channel block coding capacity are investigated for discrete asymptotically mean stationary channels with finite alphabets. For channels decomposable into a finite number of ergodic asymptotically mean stationary components with finite input memories the positive block coding theorem and its weak converse are established.

1. INTRODUCTION

A principal goal of the Shannon theory is to relate the operational capacities defined as certain optima over actual deterministic codes to information theoretic extremum problems. This is done by proving appropriate coding theorems. The most common quantity resulting from this approach is the Shannon capacity. However, Nedoma [6] demonstrated an example of a stationary channel where the Shannon capacity strictly exceeded the operational one and this led Winkelbauer [14–17] to the concept of information quantile capacity that equals the operational one for more general channels. In this paper we investigate these concepts of capacity for asymptotically mean stationary (a.m.s.) discrete channels with finite alphabets. Similar problems were studied in [10] for channels with additive a.m.s. noise.

Let A be a finite set with $|A| > 1$ elements. We denote by A^∞ the space of all doubly-infinite sequences $x = (\ldots, x_{-1}, x_0, x_1, \ldots)$, $x_i \in A$ and let $\mathscr{F}_A$ denote the σ-field generated by the set of all cylinders of the form $C_m^n(F) = \{x : x_m^{m+n-1} \in F\}$, $F \subset A^n$, where $x_m^{m+n-1} = (x_m, \ldots, x_{m+n-1})$. Any probability measure μ on $(A^\infty, \mathscr{F}_A)$ is called a source; we shall indicate the alphabet by writing $[A, \mu]$ when convenient. The symbol T_A stands for the shift, viz. $(T_A x)_n = x_{n+1}$. As in [10] we use the symbols $\mathbf{M}_A$, $\mathbf{M}_A^*$, $\mathbf{S}_A$, $\mathbf{S}_A^*$, $\mathbf{M}_A^n$, and $\mathbf{M}_A^{bs}$ to designate the sets of all stationary (i.e. T_A-invariant) sources, of all stationary and ergodic sources, of all a.m.s. sources, of all a.m.s. and ergodic sources, of all n-stationary (i.e. T_A^n-invariant) sources, and of all block stationary sources, respectively. If $\mu \in \mathbf{S}_A$ then $\bar{\mu} \in \mathbf{M}_A$ will denote its stationary mean [2], i.e.,

$$\bar{\mu}(F) = \lim_{K\to\infty} K^{-1} \sum_{k=0}^{K-1} \mu(T_A^{-k}F), \quad F \in \mathscr{F}_A .$$

If $\mathscr{I}_A = \{F \in \mathscr{F}_A : F = T_A^{-1}F\}$ then $\mu/\mathscr{I}_A = \bar{\mu}/\mathscr{I}_A$ and $\mu \ll \bar{\mu}'$ (cf. [2], the latter fact enables us to employ Jacobs' results [4]). Hence, the entropy

$$H(\mu) = \lim_{n\to\infty} n^{-1} H_n(\mu^n)$$

where $\mu^n(F) = \mu(C_0^n(F))$, $F \subset A^n$ and $H_n(\mu^n) = -\sum_{x_0^n \in A^n} \mu^n(x_0^n) \log \mu^n(x_0^n)$ exists for any $\mu \in \mathbf{S}_A$ and the Shannon-McMillan theorem is valid as well. Throughout the paper $\log = \log_2$, $\exp = \exp_2$.

A channel with input alphabet B and output alphabet A (both finite and non-empty) is a list of probability measures $\{\nu_y;\, y \in B^\infty\}$ on $(A^\infty, \mathscr{F}_A)$ such that for each event $G \in \mathscr{F}_A$, $y \mapsto \nu_y(G)$ is measurable. A channel is denoted by $[B, \nu, A]$ or simply by ν. Given an input source $[B, \mu]$ and a channel $[B, \nu, A]$ the corresponding double source $\mu\nu$ is defined on $(B^\infty \times A^\infty, \mathscr{F}_{B\times A})$ by

$$\mu\nu(F \times G) = \int_F \nu_y(G)\, \mu(\mathrm{d}y)$$

and the output source $\widetilde{\mu\nu}$ on $(A^\infty, \mathscr{F}_A)$ by

$$\widetilde{\mu\nu}(G) = \mu\nu(B^\infty \times G), \quad G \in \mathscr{F}_A .$$

A channel $[B, \nu, A]$ is said to be stationary if

$$\nu(T_A G \mid T_B y) = \nu(G \mid y) \quad (\nu(\cdot \mid y) = \nu_y(\cdot))$$

for $G \in \mathscr{F}_A$ and $y \in B^\infty$. A stationary channel is said to be ergodic if $\mu \in \mathbf{M}_B^*$ entails $\mu\nu \in \mathbf{M}_{B\times A}^*$. If $\mu \in \mathbf{S}_B$ entails $\mu\nu \in \mathbf{S}_{B\times A}$, ν is said to be a.m.s., and an a.m.s. channel is said to be ergodic if $\mu\nu \in \mathbf{S}_{B\times A}^*$ for any input source $\mu \in \mathbf{S}_B^*$ [1].

A simple result on conditional probabilities gives the following. Given μ and ν such that $\mu\nu$ is a.m.s. there is a stationary channel $\bar{\nu}$ satisfying $\overline{\mu\nu} = \bar{\mu}\bar{\nu}$ and called the induced stationary channel of μ and ν [1]. An input independent concept is that of a stationary mean. This is a stationary channel $[B, \bar{\nu}, A]$ such that, for any $\mu \in \mathbf{M}_B$, $\overline{\mu\nu} = \mu\bar{\nu}$, $\nu_y \ll \bar{\nu}_y$ μ-a.e., and

$$\lim_{K\to\infty} K^{-1} \sum_{k=0}^{K-1} \nu(T_A^{-k}G \mid T_B^{-k}y) = \bar{\nu}_y(G) \quad \mu\text{-a.e.}, \quad G \in \mathscr{F}_A .$$

The stationary mean exists for every a.m.s. channel (cf. [1, Theorem 3]). If $\bar{\nu}$ is also ergodic then so is ν [1, Lemma 4]). The converse is true as well:

Lemma 1. The stationary mean of an a.m.s. and ergodic channel is an ergodic channel.

Proof. Let $\mu \in \mathbf{M}_B^*$. As $\mathbf{M}_B^* \subset \mathbf{S}_B^*$, $\mu\nu \in \mathbf{S}_{B\times A}^*$ so that $\mu\nu/\mathscr{I}_{B\times A} = \overline{\mu\nu}/\mathscr{I}_{B\times A} = \mu\bar{\nu}/\mathscr{I}_{B\times A} \in \{0, 1\}$. Hence $\mu\bar{\nu} \in \mathbf{M}_{B\times A}^*$ wheever $\mu \in \mathbf{M}_B^*$. □

2. THE CAPACITIES

The Shannon capacity results from the attempts to express the capacity as extremum of a functional of probability measures over an appropriately constrained space. By far the most common approach deals with maximizing the mutual information rates attainable over the channel. Put for $y_0^n \in B^n$ and $x_0^n \in A^n$

$$I_n(y_0^n, x_0^n) = n^{-1} \log \frac{\mu\nu^n(y_0^n, x_0^n)}{\mu^n(y_0^n)\,\widetilde{\mu\nu}^n(x_0^n)}\,.$$

If μ and ν are a.m.s. we can use the Shannon-McMillan theorem of Jacobs [4] for any of the a.m.s. sources μ, $\mu\nu$, and $\widetilde{\mu\nu}$ (note that $\widetilde{\mu\nu}$ is the A-marginal, and μ is the B-marginal of, the double source $\mu\nu$). As $\mu\nu \ll \mu \times \widetilde{\mu\nu}$ we conclude that the limit

$$I(y, x) = \lim_{n\to\infty} I_n(y_0^n, x_0^n)$$

exists in $L_1(\mu\nu)$ sense. If $I(\mu\nu) = H(\mu) + H(\widetilde{\mu\nu}) - H(\mu\nu)$ is the mutual information rate then

$$I(\mu\nu) = E_{\mu\nu} I = \int I(y, x)\,\mathrm{d}\mu\nu(y, x)\,.$$

The function $I(y, x)$ is shift-invariant and measurable so that $\{(y, x) : I(y, x) = I(\mu\nu)\} \in \mathscr{I}_{B\times A}$. If $\mu \in \mathbf{S}_B^*$ and ν is a.m.s. and ergodic then $\mu\nu \in \mathbf{S}_{B\times A}^*$ so that the above set is of measure zero or one. Consequently, we get the following result.

Lemma 2. Given an a.m.s. ergodic source $[B, \mu]$ and an a.m.s. ergodic channel $[B, \nu, A]$. Then

$$\mu\nu\{(y, x) : I(y, x) = I(\mu\nu)\} = 1\,.$$

Let R_A denote the set of all sequences $x \in A^\infty$ which are typical for some $\mu \in \mathbf{M}_A^*$. Then $R_A \in \mathscr{I}_A$ and $\mu(R_A) = 1$ for all $\mu \in \mathbf{M}_A$ (see [11–13]). Consequently, $\mu(R_A) = 1$ for all $\mu \in \mathbf{S}_A$ as well. If $x \in R_A$ we let μ_x denote that $\mu \in \mathbf{M}_A^*$ which has x as a typical sequence (note that a typical sequence determines the source uniquely).

Lemma 3. If $\mu \in \mathbf{S}_B$ and $[B, \nu, A]$ is a.m.s. then

$$\mu\nu\{(y, x) \in R_{B\times A} : I(y, x) = I(\mu_{(y,x)})\} = 1\,.$$

Proof. Use [13, Lemma 5] or [15] to get the conclusion for stationary μ and ν. In the a.m.s. case use the fact that the set in the conclusion belongs to $\mathscr{I}_{B\times A}$. □

Put

$$[I \leqq r] = \{(y, x) \in R_{B \times A} : I(y, x) \leqq r\} .$$

Let $[B, \nu, A]$ be a.m.s. and $\vartheta \in (0, 1)$. The ϑ-information quantile is defined by

$$C^*(\nu; \vartheta) = \sup_{\mu \in \mathbf{S}_B^*} \sup \{r : \mu\nu[I \leqq r] < \vartheta\} .$$

Theorem 1. The set $\mathbf{S}_B^*$ in the formula for the ϑ-information quantile can be replaced by any of the sets $\mathbf{S}_B$, $\mathbf{M}_B$, $\mathbf{M}_B^*$ and $\mathbf{M}_B^{bs}$, $\vartheta \in (0, 1)$.

Proof.

(I) Replacement by $\mathbf{S}_B$. Since $\mathbf{S}_B^* \subset \mathbf{S}_B$ we get the inequality

$$\sup_{\mu \in \mathbf{S}_B} \sup \{r : \mu\nu[I \leqq r] < \vartheta\} \geqq C^*(\nu; \vartheta) , \quad \vartheta \in (0, 1) .$$

Choose and fix an arbitrary $\vartheta \in (0, 1)$ and assume the strict inequality. This means we can find a source $\mu_0 \in \mathbf{S}_B \setminus \mathbf{S}_B^*$ such that

$$\sup \{r : \mu_0\nu[I \leqq r] < \vartheta\} > C^*(\nu; \vartheta) .$$

Since $[I \leqq r] \in \mathscr{I}_{B \times A}$ we can replace in the latter relation the a.m.s. measures by their respective stationary means and so get

$$\sup \{r : \overline{\mu_0\nu}[I \leqq r] < \vartheta\} > \sup_{\mu \in \mathbf{S}_B^*} \sup \{r : \overline{\mu\nu}[I \leqq r] < \vartheta\} .$$

We claim that the right hand side upperbounds

$$\sup_{y \in R_B} \sup \{r : \mu_y\bar{\nu}[I \leqq r] < \vartheta\} .$$

To prove the claim choose $y \in R_B$. Then $\mu_y \in \mathbf{M}_B^* \subset \mathbf{M}_B$ and the definition of stationary mean for channels gives $\overline{\mu_y\nu} = \mu_y\bar{\nu}$. At the same time

$$\{\mu_y : y \in R_B\} = \mathbf{M}_B^* \subset \mathbf{S}_B^*$$

and this proves the claim. Hence we have

$$(*) \qquad \sup \{r : \overline{\mu_0\nu}[I \leqq r] < \vartheta\} > \sup_{y \in R_B} \sup \{r : \mu_y\bar{\nu}[I \leqq r] < \vartheta\} .$$

Next we show

$$(**) \qquad \overline{\mu_0\nu}[I \leqq r] = \int_{R_B} \mu_y\bar{\nu}[I \leqq r] \, \bar{\mu}_0(\mathrm{d}y) .$$

Let $y \in R_B$. Then

$$\mu_y\bar{\nu}[I \leqq r] = \int \bar{\nu}([I \leqq r]_s | s) \, \mu_y(\mathrm{d}s) =$$

$$= \int \lim_{K \to \infty} K^{-1} \sum_{k=0}^{K-1} \nu(T_A^{-k}[I \leqq r]_s \mid T_B^{-k}s) \, \mu_y(\mathrm{d}s) .$$

Since I is invariant and the shifts are bijective maps we can write

$$T_A^{-k}[I \leqq r]_s = T_A^{-k}\{t : I(s, t) \leqq r\} =$$
$$= \{t : I(s, T_A^k t) \leqq r\} = \{t : I(T_B^{-k}s, t) \leqq r\} = [I \leqq r]_{T_B^{-k}s} .$$

Consequently,

$$\mu_y\bar{\nu}[I \leqq r] = \int \lim_{K\to\infty} K^{-1} \sum_{k=0}^{K-1} \nu([I \leqq r]_{T_B^{-k}s} \mid T_B^{-k}s)\, \mu_y(\mathrm{d}s) =$$
$$= \lim_{K\to\infty} K^{-1} \sum_{k=0}^{K-1} \int \nu([I \leqq r]_{T_B^{-k}s} \mid T_B^{-k}s)\, \mu_y(\mathrm{d}s) .$$

But $\mu_y = \mu_y T_B^{-k}$ for all k so that all integrals on the right hand side of the latter relation equal the same number, namely $\mu_y\nu[I \leqq r]$. Since $\mu_0\nu \in \mathbf{S}_{B\times A}$ there is the induced stationary channel $\bar{\nu}$ such that $\bar{\mu}_0\bar{\nu} = \overline{\mu_0\nu}$. However, $\bar{\mu}_0 \in \mathbf{M}_B$ and hence

$$\overline{\bar{\mu}_0\nu} = \bar{\mu}_0\bar{\nu} = \bar{\mu}_0\bar{\nu} \quad \text{so that} \quad \overline{\bar{\mu}_0\nu} = \overline{\overline{\mu_0\nu}} .$$

On the other hand, $\mu_0\nu = \bar{\mu}_0\nu$, when restricted to invariant events so that we get $(**)$. Substituting $(**)$ into $(*)$ we get

$$\sup \left\{r : \int_{R_B} \mu_y\bar{\nu}[I \leqq r]\, \bar{\mu}_0(\mathrm{d}y) \leqq \vartheta\right\} > \sup_{y\in R_B} \sup \{r : \mu_y\bar{\nu}[I \leqq r] < \vartheta\} .$$

But this is a contradiction. Indeed, this inequality entails the existence of a real number r_0 such that

$$\int_{R_B} \mu_y\bar{\nu}[I \leqq r_0]\, \bar{\mu}_0(\mathrm{d}y) < \vartheta$$

and

$$r_0 > \sup \{r : \mu_y\bar{\nu}[I \leqq r] < \vartheta\}$$

for all $y \in R_B$. The second property implies that, for each $y \in R_B$, $\mu_y\bar{\nu}[I \leqq r_0] \geqq \vartheta$, and this contradicts the first property.

(II) Replacement by $\mathbf{M}_B$. Because $\mathbf{M}_B \subset \mathbf{S}_B$, part (I) gives $\sup_{\mu\in\mathbf{M}_B} \sup \{r : \mu\nu[I \leqq r] < \vartheta\} \leqq \sup_{\mu\in\mathbf{S}_B} \sup \{r : \mu\nu[I \leqq r] < \vartheta\}$. Again assume the strict inequality. Hence, there is a $\mu_0 \in \mathbf{S}_B \setminus \mathbf{M}_B$ with the property

$$\sup_{\mu\in\mathbf{M}_B} \sup \{r : \mu\nu[I \leqq r] < \vartheta\} < \sup \{r : \mu_0\nu[I \leqq r] < \vartheta\} .$$

Choose $\bar{\nu}$ as the induced stationary channel of ν and μ_0 so that

$$\mu_0\nu[I \leqq r] = \overline{\mu_0\nu}[I \leqq r] = \bar{\mu}_0\bar{\nu}[I \leqq r] .$$

On the other hand,

$$\mu\nu[I \leqq r] = \overline{\mu\nu}[I \leqq r] = \mu\bar{\nu}[I \leqq r]$$

where $\mu \in \mathbf{M}_B$ and $\bar{\nu}$ is the stationary mean of the channel ν. The difference between these two channels is however inessential (in fact, given an input source, the double sources determined by the induced stationary channel and the stationary mean must give rise to isomorphic probability spaces), and we get

$$\sup\{r : \bar{\mu}_0\bar{v}[I \leqq r] < \vartheta\} \leqq \sup_{\mu \in \mathbf{M}_B} \sup\{r : \mu\bar{v}[I \leqq r] < \vartheta\} <$$
$$< \sup\{r : \bar{\mu}_0\bar{v}[I \leqq r] < \vartheta\},$$

a contradiction.

(III) Replacement by $\mathbf{M}_B^*$. Repeat the proof for step (II) starting with the original definition of $C^*(v; \vartheta)$.

(IV) Replacement by $\mathbf{M}_B^{bs}$. If $\mu \in \mathbf{M}_B^n$ then μ is a.m.s. and

$$\bar{\mu} = n^{-1} \sum_{j=0}^{n-1} \mu T_B^{-j}.$$

Hence $\mathbf{M}_B \subset \mathbf{M}_B^{bs} \subset \mathbf{S}_B$ and the conclusion follows by parts (I) and (II). □

The limit

$$C^*(v) = \lim_{\vartheta \to 0} C^*(v; \vartheta)$$

is said to be the information quantile capacity of the channel $[B, v, A]$. Note that the limit exists as the $C^*(v; \vartheta)$ are bounded and nonincreasing as $\vartheta \to 0$.

Now let us turn to the concept of Shannon capacity. In the a.m.s. case there are a larger number of possibilities to constraint the set of input sources:

$$C_e^s(v) = \sup_{\mu \in \mathbf{M}_B^*} I(\mu v), \quad C_e^{ams}(v) = \sup_{\mu \in \mathbf{S}_B^*} I(\mu v),$$
$$C_s^s(v) = \sup_{\mu \in \mathbf{M}_B} I(\mu v), \quad C_s^{ams}(v) = \sup_{\mu \in \mathbf{S}_B} I(\mu v),$$
$$C(v) = \sup_n \sup_{\mu \in \mathbf{M}_B^n} I(\mu v).$$

Theorem 2. Given an a.m.s. channel $[B, v, A]$,

$$C_e^s(v) = C_e^{ams}(v) = C_s^s(v) = C_s^{ams}(v) = C(v).$$

In general, $C(v) \geqq C^*(v)$. If however $[B, v, A]$ is also ergodic then $C(v) = C^*(v) = C^*(v; \vartheta)$, $\vartheta \in (0, 1)$.

Proof. Let $\mu \in \mathbf{M}_B$. Then $I(\mu v) = I(\overline{\mu v}) = I(\mu\bar{v})$ so that we can repeat the reasoning of Parthasarathy [7] showing that $C_e^s(v) = C_s^s(v)$. A direct consequence of the fact that $\mu v / \mathscr{I}_{B \times A} = \overline{\mu v} / \mathscr{I}_{B \times A}$ is the relation

$$\mu v\{(y, x) : I(y, x) \leqq c\} = \overline{\mu v}\{(y, x) : I(y, x) \leqq c\}$$

for all real c, whence the relation $C_e^{ams}(v) = C_s^{ams}(v)$ follows by the same arguments as in the stationary case. Denote $C^s(v) = C_e^s(v) = C_s^s(v)$ and similarly $C^{ams}(v)$. The inequality $C^s(v) \leqq C^{ams}(v)$ follows from the inclusion $\mathbf{M}_B \subset \mathbf{S}_B$. Assuming the strict inequality we can proceed as in parts (I) and (II) of the proof of Theorem 1 in order to get $C^s(\nu) = C^{ams}(\nu)$. As $\mathbf{M}_B \subset \mathbf{M}_B^{bs} \subset \mathbf{S}_B$, we have $C^s(\nu) = C^{ams}(\nu) = C(\nu)$, thus proving the first assertion. The proofs of the remaining ones follow from [3, Lemma 1, Corollary 1]. □

Let $\bar{\nu}$ denote the stationary mean of an a.m.s. channel ν. According to part (II) in the proof of Theorem we can work with stationary input sources when dealing with the information quantile capacity and, for any $\mu \in \mathbf{M}_B$, $I(\mu\nu) = I(\overline{\mu\nu}) = I(\mu\bar{\nu})$ so that $C^*(\nu; \vartheta) = C^*(\bar{\nu}; \vartheta)$, $\vartheta \in (0, 1)$. By similar reasoning $C(\nu) = C(\bar{\nu})$. Hence, all results are proper generalizations of the results known in the stationary case.

3. THE OPERATIONAL CAPACITY

In this paper we shall restrict ourselves to the operational channel block coding capacity [3]. However, our proof of the corresponding coding theorem makes it an easy task to get also the results for the operational source/channel block coding capacity (see [3] for an excellent survey of the work concerning the capacities and the corresponding coding theorems).

A block-length n channel block code $\mathscr{Y}$ for a discrete channel $[B, \nu, A]$ is a collection of $M = |\mathscr{Y}|$ distinct codewords $\mathbf{y}_i \in B^n$ and M disjoint decoding sets $G_i \subset A^n$, $i = 1, \ldots, M$. The rate R of a code is $R = n^{-1} \log M$. A code $\mathscr{Y} = \{(\mathbf{y}_i, G_i) : i = 1, \ldots, M\}$ has error probability ε if

$$\max_{1 \leqq i \leqq M} \sup_{y \in C_0^n(y_i)} \nu_y(C_0^n(A^n \setminus G_i)) \leqq \varepsilon$$

We call such a code also an (M, n, ε) code [18]. We say R is a permissible rate if there exist $([\exp(nR)], n, \varepsilon_n)$ *) codes such that $\varepsilon_n \to 0$ as $n \to \infty$ through a subsequence of positive integers. Define the block coding capacity $C_0(\nu)$ as the supremum of all permissible rates.

Note that it is possible to define the concept of an ε-permissible rate for any fixed $\varepsilon \in (0, 1)$. This concept is closely related with the Theorem on ε-Capacities as established by Winkelbauer [14–16].

In general, we establish that a particular number C_0 is the capacity of a given channel by proving the positive coding theorem (that is, for any $R < C_0$, there exist $([\exp(nR)], n, \varepsilon_n)$ codes with $\varepsilon_n \to 0$ as $n \to \infty$, and hence that $C_0 \leqq C_0(\nu)$) and a weak converse (i.e., given any sequence of $([\exp(nR)], n, \varepsilon_n)$ codes, $R > C_0$, there is an $\varepsilon_0 > 0$ such that $\varepsilon_n \geqq \varepsilon_0$ for n large enough; hence $C_0 \geqq C_0(\nu)$). The positive part is usually proved using Feinstein's lemma [5]. The problem is that Feinstein's lemma gives "good" codes for an artificial measure on the output n-sequences induced by the "capacity yielding input source". Under some additional

*) [·] denotes integer part of ·.

assumptions on the channel, however, we can reach at an arbitrarily good approximation of $\nu_y(\cdot)$ on output n-tuples by that artificial measure. By far the mot common assumption is that the channel has finite input memory. That is, there exists a number m such that, for any integer i, any $n \geqq 1$, and $\mathbf{x} \in A^n$, if $y_{i-m}^{i+n-1} = (y')_{i-m}^{i+n-1}$ then

$$\nu(C_i^n(\mathbf{x}) \mid y) = \nu(C_i^n(\mathbf{x}) \mid y') .$$

Then the measure ν_{in} on the n-tuples from A^n defined by the properties

$$\nu_{in}[\mathbf{x} \mid y_{i-m}^{i+n-1}] = \nu(C_i^n(\mathbf{x}) \mid y) , \quad \mathbf{x} \in A^n ;$$

$$\nu_{in}[E \mid \mathbf{y}] = \sum_{\mathbf{x} \in E} \nu_{in}[\mathbf{x} \mid \mathbf{y}] , \quad E \subset A^n , \quad \mathbf{y} \in B^{n+m}$$

is close to the artificial measure on A^n in the variational sense (see [3] for another idea of approximation based on Ornstein's $\bar{d}$-distance). Call the least number $m = m(\nu)$ such that the above condition is satisfied the duration of the input memory.

Lemma 4. Given an a.m.s. channel $[B, \nu, A]$ with finite input memory, the stationary mean $[B, \bar{\nu}, A]$ has finite input memory of the same duration.

Proof. Let $m = m(\nu)$, i, n, $\mathbf{x} \in A^n$ and $k \geqq 0$ be given. Choose y and y' coinciding in coordinates from $i - m$ up to $i + n - 1$ and use the finite input memory condition for $i + k$ and $T_B^k y$ and $T_B^k y'$, respectively. Then

$$\nu(T_A^{-k} C_i^n(\mathbf{x}) \mid T_B^{-k} y) = \nu(T_A^{-k} C_i^n(\mathbf{x}) \mid T_B^{-k} y')$$

thus

$$\bar{\nu}(C_i^n(\mathbf{x}) \mid y) = \bar{\nu}(C_i^n(\mathbf{x}) \mid y') . \qquad \square$$

The above notion of the block-length n channel block code can be reformulated as follows. Let $[B, \nu, A]$ be an a.m.s. channel with finite input memory of duration $m = m(\nu)$. Let $\psi : A^n \to B^{n+m}$. Then the set

$$\{(\mathbf{y}, \psi^{-1}\{\mathbf{y}\}) : \mathbf{y} \in \mathscr{Y}_{in}(\psi) ;$$

$$\mathscr{Y}_{in}(\psi) = \{\mathbf{y} \in B^{n+m} : \nu_{in}(\psi^{-1}\{\mathbf{y}\} \mid \mathbf{y}] > 1 - \varepsilon\}$$

is said to be a $(|\mathscr{Y}_{in}(\psi)|, n, \varepsilon)$ code at origin i. We denote $|\mathscr{Y}_{in}(\psi)| = S_{in}(\psi; \varepsilon, \nu)$ and put

$$S_{in}(\varepsilon, \nu) = \max \{S_{in}(\psi; \varepsilon, \nu) : \psi \in (B^{n+m})^{A^n}\} .$$

Theorem 3. Let $[B, \nu, A]$ be a finite composition

$$\nu_y(G) = \sum_{\lambda \in \Lambda} \nu_y^\lambda(G) \gamma_\lambda$$

($\Lambda = \{1, \ldots, k\}$, $\gamma_\lambda > 0$, $\sum \gamma_\lambda = 1$) of a.m.s. and ergodic channels $[B, \nu^\lambda, A]$ with finite input memories. Let

$$C_*(\nu) = \sup_{\mu \in M_B^*} \min_{1 \leqq \lambda \leqq k} I(\mu\nu^\lambda) .$$

Then

(I) $$\lim_{\varepsilon \to 0} \limsup_{n \to \infty} n^{-1} \log S_{0n}(\varepsilon, \nu) \leqq C_*(\nu)$$

and

(II) $$\lim_{\varepsilon \to 0} \liminf_{n \to \infty} n^{-1} \log S_{0n}(\varepsilon, \nu) \geqq C_*(\nu) .$$

Comment. One can easily check that the composed channel is a.m.s. and has finite input memory. If $k > 1$, it cannot be ergodic. By an easy modification of the proof given in [11] one can verify that the left hand sides in (I) and (II) above coincide, thus these quantities define the operational block coding capacity $C_0(\nu)$.

Proof. The proof is an adaptation of the proof for the stationary case [17]. As both (I) and (II) will be proved indirectly, Feinstein's lemma naturally appears in the proof of (II) while Fano's inequality works in (I). Since we did not establish until now what happens for codes at origins $i \neq 0$, we use the symbol $\tilde{C}(\nu)$ to designate the left hand sides of (I) and (II) and not $C_0(\nu)$.

(I) Assume the contrary, i.e., let $\tilde{C}(\nu) > C_*(\nu)$. Then there is a real number r with $\tilde{C}(\nu) > r > C_*(\nu)$. We find a stationary memoryless input source μ with $H(\mu) = r$; hence $C_*(\nu) < r = H(\mu) < \tilde{C}(\nu)$. As in [17] we choose ε and n_0 such that, for $n \geqq n_0$, $L_n(\varepsilon, \mu) < S_{0n}(\varepsilon, \nu)$, where

$$L_n(\varepsilon, \mu) = \min \{|E| : E \subset B^n,\ \mu^n(E) > 1 - \varepsilon\}$$

(note that, as $\mu \in \boldsymbol{M}_B^*$, $n^{-1} \log L_n(\varepsilon, \mu)$ converges to $H(\mu)$ for $n \to \infty$, $\varepsilon \in (0, 1)$, cf. [12]. The same is true even for $\mu \in S_B^*$ [9]). Hence there is a map $\psi : A^n \to B^{n+m}$ $(m = m(\nu))$ such that $L_n(\varepsilon, \mu) < S_{0n}(\psi; \varepsilon, \nu)$ and this makes it possible to define a block-length n source/channel block code as a pair of encoder $\varkappa : B^n \to B^{n+m}$ and decoder $\delta : A^n \to B^n$ such that the error probability

$$e'_{0n}(\mu, \nu, \varkappa, \delta) = 1 - \sum_{z \in B^n} \nu_{0n}[\delta^{-1}\{z\} \mid \varkappa z]\, \mu^n\{z\}$$

is less than 2ε. (Note that this construction is a rather particular case of what is usually called the source/channel block code. In fact, we just require that the input source producing n-tuples be encoded into the input of the channel which acts not on n-tuples but on $(n + m)$-tuples due to the memory effect.) The encoder $\varkappa$ implements a mapping on infinite strings:

$$\{(\tau z)_{-m+k(m+n)+i}\}_{0 \leqq i < m+n} = \varkappa(z_{kn}^{(k+1)n-1}), \quad z \in B^\infty .$$

We denote by $\mu\tau^{-1}$ the encoded source (i.e., the channel input process) and by $\omega = (\mu\tau^{-1})\, \nu$ the corresponding double source. As in [17] we deduce that the source/channel error probability

$$e_{0n}(\omega) = 1 - \sum_{\mathbf{x} \in A^n} \max_{\mathbf{z} \in B^n} \int_{C_0^n(\mathbf{z})} \nu(C_0^n(\mathbf{x}) \mid \tau z)\, \mu(\mathrm{d}z)$$

satisfies

$$e_{0n}(\omega) \leqq e'_{0n}(\mu, \nu, \varkappa, \delta) .$$

As one can easily check, $\mu\tau^{-1}$ is a. m.s. (in fact, $\mu\tau^{-1} \in \mathbf{M}_B^{n+m}$) and, moreover,

$$\overline{\mu\tau^{-1}} = (m + n)^{-1} \sum_{i=0}^{n+m-1} (\mu\tau^{-1}) T_B^i$$

is a stationary ergodic source with the same entropy as $\mu\tau^{-1}$. Since ω is a.m.s. as the result of joining an a.m.s. input source to an a.m.s. channel, Jacobs' arguments [4] work well to show that

$$I(\omega) = H(\mu\tau^{-1}) - H(\omega; B \mid A) ,$$

where

$$H(\omega; B \mid A) = \lim_{n\to\infty} (m + n)^{-1} H_n(\omega; B \mid A) .$$

Here, $H_n(\omega; B \mid A)$ is the $(n + m, n)$-dimensional relative entropy (see [17], p. 910) for which one has the following form of Fano's inequality

$$H_n(\omega; B \mid A) \leqq (m + n)\, e_{0n}(\omega) \log |B| + 1$$

(see [11]; in fact, the proof is a purely finite-dimensional construction so that carries over to the a.m.s. case as well). Using $\overline{\mu\tau^{-1}}$ as the input source and checking the relation

$$I(\overline{\mu\tau^{-1}}\nu) = \sum_{\lambda\in\Lambda} \gamma_\lambda I(\overline{\mu\tau^{-1}}\nu^\lambda) = I(\omega)$$

(for this use the similar relation valid for the corresponding stationary means and established in [17] and the fact that $I(\mu\nu) = E_{\mu\nu} I$, where I is a function equally ditributed under $\mu\nu$ and under $\overline{\mu\nu}$) we can derive the desired contradiction in the same way as done in [17].

(II) Assume $\tilde{C}(\nu) < C_*(\nu)$ so that $C_*(\nu) > r + \alpha > r > \tilde{C}(\nu)$ for some appropriate r and $\alpha > 0$. Thus, there is a $\mu \in \mathbf{M}_B^*$ such that

$$\min_{1\leqq\lambda\leqq k} I(\mu\nu^\lambda) > r + \alpha .$$

On the other hand, using $r > \tilde{C}(\nu)$ we deduce the existence of an $\varepsilon \in (0, 1)$ such that

$$c_0(\varepsilon, \nu) = \liminf_{n\to\infty} n^{-1} \log S_{0n}(\varepsilon, \nu) < r .$$

At this stage it is almost clear that the latter two inequalities contradict. Formally, we use the procedure employed by Khinchin [5] in his proof of Feinstein's lemma to get a map ψ_n for which

$$(\varepsilon/2) \exp(nr) < S_{0n}(\psi_n; \varepsilon, \nu) \leqq S_{0n}(\varepsilon, \nu)$$

for any n sufficiently large. Hence $c_0(\varepsilon, \nu) \geqq r$, a contradiction proving the theorem. □

4. CONCLUDING REMARKS

Let us denote by $\tilde{C}_i(\nu)$ the quantities standing on the left hand sides of assertions (I) and (II) of Theorem 3 with S_{0n} replaced by S_{in}, $i \neq 0$. We claim that $\tilde{C}_i(\nu) = C(\nu)$ This can be seen indirectly by showing that $\tilde{C}_i(\nu) = C_*(\nu)$, too. In fact, the proof of Theorem 3 makes use either of finite-dimensional constructions (like block encoders and decoders, Feinstein's lemma etc.) that can be performed for any origin i or, of the asymptotic results which, in light of [4], do not depend on the particular choice of i. More precisely, the limiting quantities corresponding to a.m.s. sources and channels (like entropy, information rate etc.) coincide with those which correspond to their stationary means, and the latter ones do not depend on i.

As a simple consequence of Theorem 3 we get the following result.

Corollary. Given an a.m.s. and erodic channel $[B, \nu, A]$ with finite input memory. Then

$$\tilde{C}_i(\nu) = \tilde{C}(\nu) = C(\nu) = C^*(\nu) .$$

However, we already know that $C^*(\nu) = C^*(\nu; \vartheta)$ for all $\vartheta \in (0, 1)$ (see Theorem 2). Thus, two natural problems appear:

- in the ergodic case, is it necessary to take the limit for $\varepsilon \to 0$ when computing $\tilde{C}_i(\nu)$?
- in the general case, does $C^*(\nu; \vartheta)$ determine the behavior of the maximum length of the n-dimensional ϑ-codes for separate values of $\vartheta \in (0, 1)$? (see [8] and [10] for the answer in affirmative in the particular case of channels with additive, stationary or a.m.s., noise).

These questions will be investigated in a subsequent paper devoted to the Theorem on ε-Capacities for a.m.s. channels.

(Received October 10, 1980.)

REFERENCES

[1] R. J. Fontana, R. M. Gray and J. C. Kieffer: Asymptotically mean stationary channels (preprint).

[2] R. M. Gray and J. C. Kieffer: Asymptotically mean stationary measures (submitted to Annals of Prob.).

[3] R. M. Gray, D. S. Ornstein: Block coding for discrete stationary $\bar{d}$-continuous noisy channels. IEEE Trans. on Inform. Theory *IT-25* (1979), 292–306.

[4] K. Jacobs: Die Übertragung diskreter Informationen durch periodische und fastperiodische Kanäle. Math. Annalen *137* (1959), 125–135.

[5] A. I. Khinchine: Mathematical Foundations of Information Theory. Dover, New York 1957.

[6] J. Nedoma: On non-ergodic channels. Trans. Second Prague Conf. Inform. Theory etc., NČSAV, Prague 1960, 143–181.

[7] K. R. Parthasarathy: On the integral representation of the rate of transmission of a stationary channel. Ill. J. Math. *2* (1961), 299–305.

[8] K. R. Parthasarathy: Effective entropy rate and transmission of information through channels with additive random noise. Sankhya *A 25* (1963), 75—84.
[9] Š. Šujan: Existence of asymptotic rate for asymptotically mean stationary sources with countable alphabets. Trans. Third Czechoslovak-Soviet-Hungarian Seminar on Inform. Theory, 1980, 201—206.
[10] Š. Šujan: Channels with additive asymptotically mean stationary noise. Kybernetika *17* (1981), 1, 1—15.
[11] K. Winkelbauer: Channels with finite past history. Trans. Second Prague Conf. Infor. Theory etc., NČSAV, Prague 1960, 685—831.
[12] K. Winkelbauer: On discrete information sources. Trans. Third Prague Conf. Inform. Theory etc., NČSAV, Prague 1964, 765—830.
[13] K. Winkelbauer: On the asymptotic rate of non-ergodic information sources. Kybernetika *6* (1970), 127—148.
[14] K. Winkelbauer: On the coding theorem for decomposable discrete information channels I. Kybernetika *7* (1971), 109—123.
[15] K. Winkelbauer: On the coding theorem for decomposable discrete information channels II. Kybernetika *7* (1971), 230—255.
[16] K. Winkelbauer: On the regularity condition for decomposable communication channels. Kybernetika *7* (1971), 314—327.
[17] K. Winkelbauer: On the capacity of decomposable channels. Trans. Sixth Prague Conf. Inform. Theory etc., Academia, Prague 1973, 903—914.
[18] J. Wolfowitz: Coding Theorems of Information Theory. Sec. ed., Springer-Verlag, Berlin 1964.

RNDr. Štefan Šujan, CSc., Ústav merania a meracej techniky SAV (Institute of Measurement and Measuring Technique — Slovak Academy of Sciences), 885 27 Bratislava. Czechoslovakia.