KYBERNETIKA — VOLUME 17 (1981), NUMBER 6

CONTINUITY AND QUANTIZATION OF CHANNELS WITH INFINITE ALPHABETS

ŠTEFAN ŠUJAN

Block coding theorems are obtained for transmission of abstract alphabet stationary and ergodic sources over abstract alphabet stationary and weakly continuous channels in case when all data are subject to quantization errors.

1. INTRODUCTION

Winkelbauer [1] proposed a method for proving block coding theorems for stationary channels decomposable into ergodic components (cf. also [2]). A slight modification gives analogous results for a general class of stationary non-ergodic channels termed $\bar{d}$-continuous [3].

In spite of the fact that Winkelbauer follows in formulation the traditional approach which treats source and channel coding problems separately, his proofs actually give a joint source/channel coding theorem. Joint source/channel codes are designed for transmission of a source over a channel, therefore we speak also about coding theorems for transmission. Gray and Orstein [3] show that a good joint source/channel code can always be constructed from a good channel code, the latter being understood in the usual sense of Wolfowitz [4].

On the other hand, the way of constructing good joint source/channel codes is somewhat roundabout. First we construct a good channel code without any reference to the source to be transmitted over the channel. In the second the source is taken into consideration in order to construct the desired joint source/channel code. Thus, one expects that a direct construction will require weaker assumptions on the channel itself, these being compensated by the knowledge of the source. More technically, the usual notion of a block channel code seems to be unnecessarily strong from the point of view of the transmission theorems. These ideas have been made rigorous by Kieffer [5] who shows that a weaker notion of code can do the same job within the setup of block transmission.

This observation has the following important consequences. The error probability of a block channel code is calculated with the aid of the actual channel probabilities. As observed in [3], this requires continuity of the map $x \mapsto \nu_x$ carrying each input sequence into the corresponding noise component. The weaker notion of code due to Kieffer requires, however, a weaker continuity assumption. Indeed, we merely have to require that the map $\lambda \mapsto \lambda\nu$ carrying each stationary and ergodic input source into the joint input/output distribution be continuous. Kieffer [5] shows that weak continuity is appropriate and thereby obtains the most general class of stationary channels for which one would need a coding theorem.

The aim of the present paper is to investigate the problem of block transmission of abstract alphabet sources over abstract alphabet channels when all data are subject to quantization. Winkelbauer [1] suggested the way of attacking this problem, however, a simple combination of available finite alphabet results with a quantization procedure fails to work. This is in contrast with source coding where this idea works well and results in interesting results (cf. [6] and the references therein). The reason is very simple. The structure obtained by quantization of the input and output alphabets of a channel is not a channel, unless the structure of noise is extremely simple (e.g., for channels with additive noise where the components of noise are just different shifts of a single source of noise; cf. [7]).

That is why Kieffer's observations become so important. Indeed, by quantization of the joint input/output distribution we again obtain a joint distribution, namely that of the quantized pair process. Hence, the question reads whether one can prove block coding theorems for transmission without reference to the actual channel probabilities but merely to the corresponding joint input/output distributions. If the answer was in affirmative, then one could combine quantizations of alphabets with finite alphabet results and obtain the desired coding theorems. The rest of the paper is devoted to the discussion of conditions which make this idea possible.

2. PRELIMINARIES

Let $(A, \mathscr{A})$ denote a standard measurable space (we can assume that A is a complete separable metric space and $\mathscr{A}$ is the Borel σ-field, and let d_A stand for the metric on A). Let $(A^n, \mathscr{A}^n)$ and $(A^\infty, \mathscr{A}^\infty)$ denote the standard spaces of all n-tuples $z^n = (z_0, \ldots, z_{n-1})$ and of all sequences $z = (\ldots, z_{-1}, z_0, z_1, \ldots)$ with $z_i \in A$; the σ-fields $\mathscr{A}^n$ and $\mathscr{A}^\infty$ being the usual product σ-fields. For $z \in A^\infty$ let $Z_i(z) = z_i$ denote the projection of z onto its i-th coordinate, and let T_A denote the shift on A^∞, i.e., $(T_A z)_i = z_{i+1}$. Let μ be a probability measure on $(A^\infty, \mathscr{A}^\infty)$. The sequence of random variables $Z_i(z) = Z_0(T_A^i z)$ is then a discrete time random process and the whole structure is abbreviated as $[A, \mu, Z]$, $[A, \mu]$ or μ as convenient, and called a *source*. If $\mu = \mu T_A^{-n}$ then μ is called *n-stationary*, in symbols, $\mu \in \mathscr{M}_A^n(A)$. A source $\mu \in \mathscr{M}^n(A)$ is said to be *n-ergodic* if $E \in \mathscr{A}^\infty$ and $E = T_A^{-n} E$ imply that $\mu(E) \in \{0, 1\}$.

We let $\mathscr{E}^n(A)$ designate the set of all n-ergodic sources. A 1-stationary (1-ergodic) source is called simply a *stationary* (*ergodic*) source, and we shall use the notations $\mathscr{M}(A) = \mathscr{M}^1(A)$, $\mathscr{E}(A) = \mathscr{E}^1(A)$, respectively. Also, let $\mathscr{M}_b(A)$ denote the set of all block stationary sources. By definition, $\mu \in \mathscr{M}_b(A)$ if $\mu \in \mathscr{M}^n(A)$ for some $n \geqq 1$.

Let $\mathscr{V}_A$ stand for the field od all finite dimensional cylinders in A^∞. $V \in \mathscr{V}_A$ if there exist an integer i, a nonnegative integer n, and a set $E \in \mathscr{A}^n$ such that

$$V = C_i^n(E) = \{z \in A^\infty \mid (z_i, \ldots, z_{i+n-1}) \in E\} .$$

We write z_i^n for $(z_i, \ldots, z_{i+n-1})$ and, as above, z^n means z_0^n. Similarly, $C^n(\cdot) = C_0^n(\cdot)$. Let μ^n denote the restriction of μ to $\mathscr{A}^n$, that is,

$$\mu^n(E) = \mu[C^n(E)] , \quad E \in \mathscr{A}^n .$$

The field $\mathscr{V}_A$ plays a distinguished role when A is a countable discrete space so that $\mathscr{A}$ consists of all subsets of A. In this case, $\mathscr{V}_A$ is a base for the topology of A^∞ which, being a field, consists of clopen sets so that A^∞ becomes a totally disconnected space. In particular, the uniform closure of the convex hull of the set of all indicator functions of sets in $\mathscr{V}_A$ is just the space $C(A^\infty)$ of all bounded continuous functions on A^∞. This considerably simplifies notions like weak convergence of probability measures on spaces with discrete alphabets.

Let $(B, \mathscr{B})$ and $(C, \mathscr{C})$ be two standard spaces. The coordinate processes on B^∞ and C^∞ will be denoted by $X = \{X_i\}_{i=-\infty}^{\infty}$ and $Y = \{Y_i\}_{i=-\infty}^{\infty}$. A *channel* with input alphabet B and output alphabet C is denoted by $[B, \nu, C]$, where $\nu = \{\nu_x \mid x \in B^\infty\}$ is class of probability measures ν_x on $(C^\infty, \mathscr{C}^\infty)$ such that for each $F \in \mathscr{C}^\infty$, the map $x \mapsto \nu_x(F) : B^\infty \to [0, 1]$ is measurable. We assume that all channels throughout the paper are stationary, that is,

$$\nu(T_C F \mid T_B x) = \nu_x(F) ; \quad x \in B^\infty , \quad F \in \mathscr{C}^\infty ,$$

where we use the symbol $\nu(\cdot \mid x)$ as alternate for $\nu_x(\cdot)$. Given a source $[B, \lambda]$ we define the joint input/output distribution or, the double source, $[B \times C, \lambda\nu]$ by the properties that

$$\lambda\nu(E \times F) = \int_E \nu_x(F)\, \lambda(\mathrm{d}x) ; \quad E \in B^\infty , \quad F \in \mathscr{C}^\infty .$$

Also let $(X, Y)_i\,(x, y) = (x_i, y_i)$ and $(T_{B\times C}(x, y))_i = (x_{i+1}, y_{i+1})$, respectively. By joining a source $[B, \lambda] \in \mathscr{M}^n(B)$ with a stationary channel $[B, \nu, C]$ we get that $[B \times C, \lambda\nu] \in \mathscr{M}^n(B \times C)$, $n \geqq 1$.

If A is a finite set, let $\|A\|$ denote its cardinality. Our next considerations relate to finite alphabet channels unless otherwise stated. Following [5], a (M, n, ε)-*channel code* for a stationary channel $[B, \nu, C]$ is a collection $\mathscr{C}^{(n)} = \{(w_i, W_i) \mid 1 \leqq i \leqq M\}$ of M distinct n-tuples $w_i \in B^n$ called code words, and of M pairwise disjoint decoding sets $W_i \in \mathscr{C}^n$ (as $\|C\| < \infty$, the W_i's are simply subsets of C^n) such that for every $[B, \lambda] \in \mathscr{M}^n(B)$ satisfying $\lambda^n\{w_1, \ldots, w_M\} = 1$, we have

$$\lambda\nu[(X^n, Y^n) \in \bigcup_{i=1}^{M} \{w_i\} \times W_i] \geqq 1 - \varepsilon .$$

Recall that a (M, n, ε)-channel code in the usual sense of Wolfowitz [4] is a collection $\mathscr{C}^{(n)}$ such that

$$\max_{1 \leqq i \leqq M} \sup_{x \in C^n(w_i)} \nu_x^n(W_i^c) \leqq \varepsilon .$$

Observe that Kieffer's notion is similar to the notion of a Feinstein code as introduced in [3]. This suggests the kind of restrictions we must impose upon the channels in order that we can get a coding theorem. Indeed, if a (M, n, ε)-channel code was designed for some input source $[B, \lambda]$ then its performance should remain nearly the same for sources $[B, \lambda']$ which are not too apart from $[B, \lambda]$. From the point of view of the error probability this means that a small change in λ should produce only a small change in $\lambda\nu$. A natural tool for measuring the extent of such changes is provided by the weak topology. Accordingly, a stationary channel $[B, \nu, C]$ is said to be *weakly continuous* if, for any $\lambda_n, \lambda \in \mathscr{S}(B)$ such that $\lambda_n \to \lambda$ weakly, we have that $\lambda_n\nu \to \lambda\nu$ weakly. Recall that $\lambda_n \to \lambda$ weakly if and only if

$$\int f \, d\lambda_n \to \int f \, d\lambda \quad \text{for} \quad f \in C'B^\infty) .$$

If $\|B\| < \infty$ then $\lambda_n \to \lambda$ weakly if and only if $\lambda_n(V) \to \lambda(V)$ for each $V \in \mathscr{V}_B$ (cf. [8] for more about weak convergence).

The next definitions also concern finite alphabets. A map $\Phi : A^\infty \to B^\infty$ is said to be a block code of order N if $\Phi \circ T_A^N = T_B^N \circ \Phi$ and for some $\Phi' : A^N \to B^N$, $\Phi(z)^N = \Phi'(z^N)$, $z \in A^\infty$. The process X is said to be a N-block coding of the process Z if there is a block code of order N such that $X = \Phi(Z)$. Observe that if $[A, \mu, Z] \in \mathscr{M}(A)$ then $[B, \mu\Phi^{-1}, X] \in \mathscr{M}^N(B)$ provided Φ is of order N. Following [9] we call a sequence Z, X, Y, V of processes with alphabets A, B, C, A a *hookup* of the source $[A, \mu]$ to the channel $[B, \nu, C]$ if Z, X, Y, V form a Markov chain, the distribution of Z is μ, and the distribution of the conditioned process $Y | X$ is determined by ν. We say that $[A, \mu]$ is *block transmissible* over $[B, \nu, C]$ if, given $\varepsilon > 0$, there are a N and a hookup Z, X, Y, V for which $X = \Phi_1(Z)$ and $V = \Phi_2(Y)$ for some block codes of order N, the *encoder* $\Phi_1 : A^\infty \to B^\infty$ and the *decoder* $\Phi_2 : C^\infty \to A^\infty$, such that

$$\text{Prob}\,[V^N \neq Z^N] \leqq \varepsilon .$$

Observe that block transmissibility in this sense is termed strong block transmissibility in [9], whereas block transmissibility refers to the weaker requirement that

$$N^{-1} \sum_{i=0}^{N-1} \text{Prob}\,[V_i \neq Z_i] \leqq \varepsilon .$$

It is clear that knowing the distribution of Z and $Y | X$ does the same job as knowing those of Z and (X, Y), within the context of (block) transmission of the process Z over the channel $[B, \nu, C]$ with input process X and output process Y.

3. QUANTIZATION, MUTUAL INFORMATION, CAPACITY

For a moment assume that $(A, \mathscr{A})$ is an abstract measurable space. A triple $(\mathscr{P}, \hat{A}(\mathscr{P}), p)$ is said to be a *quantizer* of the alphabet A if $\mathscr{P} = \{P_i \mid 0 \leqq i \leqq \|\mathscr{P}\| - 1\}$ is a finite partition of A with $P_i \in \mathscr{A}$, $\hat{A}(\mathscr{P}) = \{\hat{a}_i(\mathscr{P}) \mid 0 \leqq i \leqq \|\mathscr{P}\| - 1\}$ is an arbitrary finite set satisfying $\hat{a}_i(\mathscr{P}) \in P_i$ for all i, and p is the corresponding quantizer map from A onto $\hat{A}(\mathscr{P})$, i.e.

$$p(a) = \hat{a}_i(\mathscr{P}) \quad \text{if} \quad a \in P_i\,.$$

Also let $\pi : A \to J(\|\mathscr{P}\|) = \{0, 1, \ldots, \|\mathscr{P}\| - 1\}$ denote the index function

$$\pi(a) = i \quad \text{if} \quad a \in P_i\,.$$

The properties of the index function have been studied in depth in [10]. We extend p and π to A^n and A^∞ in a single symbol way and thereby obtain measurable maps which will be denoted by the same symbols. This will cause no difficulties as the domain on which these maps will act will always be clear from the context. Let $\mathscr{J}(\|\mathscr{P}\|)$ stand for the σ-field of all subsets of $J(\|\mathscr{P}\|)$. Given $[A, \mu] \in \mathscr{M}^n(A)$ $(\in \mathscr{E}^n(A))$ we have that $[\hat{A}(\mathscr{P}), \mu p^{-1}] \in \mathscr{M}^n(\hat{A}(\mathscr{P}))$ $(\in \mathscr{E}^n(\hat{A}(\mathscr{P})))$ and $[J(\|\mathscr{P}\|), \mu\pi^{-1}] \in \mathscr{M}^n(J(\|\mathscr{P}\|))$ $(\in \mathscr{E}^n(J(\|\mathscr{P}\|)))$, respectively. Moreover, if $[A, \mu] \in \mathscr{M}^n(A)$ then the shifts on $\hat{A}(\mathscr{P})^\infty$ and $J(\|\mathscr{P}\|)^\infty$ have isomorphic n-th powers so that

$$H(\mu p^{-1}) = H(\mu\pi^{-1})\,,$$

where $H(\cdot)$ stands for the entropy rate functional (cf. [11] for $n = 1$). If $[A, \mu]$ is stationary, then the quantity

$$H(\mu) = \sup H(\mu p^{-1})\,,$$

where the supremum is taken over all quantizers of the alphabet A is called the *entropy rate* of the source $[A, \mu]$. As pointed out in [10], $H(\mu)$ is the Kolmogorov-Sinai invariant for the shift (T_A, μ) provided the σ-field $\mathscr{A}^\infty$ is countably generated.

Gray and Kieffer [6] have shown that if $(A, \mathscr{A})$ is a standard space then there is a sequence of quantizers $\{(\mathscr{P}(m), \hat{A}(m), p_m)\}_{m=1}^{\infty}$ which is asymptotically accurate in that

$$\lim_{m\to\infty} d_A(a, p_m(a)) = 0\,, \quad a \in A\,;$$

and

$$\mathscr{P}(m) \uparrow \mathscr{A} \ \text{(i.e., } \mathscr{A} = \sigma(\bigcup_{m=1}^{\infty} \sigma(\mathscr{P}(m))))\,.$$

Moreover, this sequence have been constructed by them in some precisely defined way, therefore we call it a Gray-Kieffer sequence, for later reference. Using the latter property one can prove as in [11] that

$$H(\mu p^{-1}) \uparrow H(\mu)\,, \quad H(\mu\pi^{-1}) \uparrow H(\mu)$$

as $m \to \infty$. This is in fact the well known Rohlin-Sinai approximation theorem for entropy. Gray and Kieffer established analogous results for the mutual information rate.

Let $\{(\mathscr{R}(m), \hat{B}(m), r_m)\}_{m=1}^{\infty}$ and $\{(\mathscr{S}(m), \hat{C}(m), s_m)\}_{m=1}^{\infty}$ denote Gray-Kieffer sequences of quantizers of the channel alphabets B and C which are assumed to be standard spaces. Let $m \geqq 1$ and $\hat{E} \in \hat{\mathscr{B}}(m)^{\infty}$, $\hat{F} \in \hat{\mathscr{C}}(m)^{\infty}$ be given. If $[B, \lambda]$ is a source, we define the double source $[\hat{B}(m) \times \hat{C}(m)], \widehat{\lambda v}^{(m)}]$ by the properties that

$$\widehat{\lambda v}^{(m)}(\hat{E} \times \hat{F}) = \lambda v(r_m^{-1}\hat{E} \times s_m^{-1}\hat{F}) .$$

Let

$$i[\widehat{\lambda v}^{(m)}](\hat{x}, \hat{y}) = \\ = \lim_{n \to \infty} n^{-1} \log \frac{\widehat{\lambda v}^{(m)}\{(\hat{\xi}, \hat{\eta}) | \hat{\xi}^n = \hat{x}^n, \hat{\eta}^n = \hat{y}^n\}}{\widehat{\lambda v}^{(m)}\{(\hat{\xi}, \hat{\eta}) | \hat{\xi}^n = \hat{x}^n\} \widehat{\lambda v}^{(m)}\{(\hat{\xi}, \hat{\eta}) | \hat{\eta}^n = \hat{y}^n\}},$$

where $\hat{x} \in \hat{B}(m)^{\infty}$, $\hat{y} \in \hat{C}(m)^{\infty}$, and $\log = \log_2$; the convergence being in $L^1(\widehat{\lambda v}^{(m)})$-norm. The average mutual information rate of the double source $\widehat{\lambda v}^{(m)}$ is defined as the expectation

$$I(\widehat{\lambda v}^{(m)}) = \int i[\widehat{\lambda v}^{(m)}](\hat{x}, \hat{y}) \, \widehat{\lambda v}^{(m)}(\mathrm{d}\hat{x}, \mathrm{d}\hat{y}) .$$

Then $I(\widehat{\lambda v}^{(m)}) \uparrow I(\lambda v)$, where $I(\lambda v)$ is defined as the supremum over all possible (single symbol) quantizers of the corresponding finite alphabet average mutual information rates. In other words, a Rohlin-Sinai approximation theorem is valid (cf. [6], Theorem 3). Observe that usually one works with reversal of supremum and limit [12]. However, this alternate approach has many advantages, in particular, the ergodic decomposition theory known for entropy (cf. [13] and [14]) extends easily to the average mutual information rate (cf. [6] for details). Let

$$C_m^* = \lim_{\varepsilon \to 0+} [\sup_{\lambda \in \mathscr{E}(B)} \sup \{R | \widehat{\lambda v}^{(m)}[i[\widehat{\lambda v}^{(m)}] \leqq R] < \varepsilon\}] .$$

As in [7] one can prove that

$$\lim_{m \to \infty} C_m^* = \sup_{m \geqq 1} C_m^* = C^*(v) ,$$

where

$$C^*(v) = \lim_{\varepsilon \to 0+} \sup_{\lambda \in \mathscr{E}(B)} \sup \{R | \lambda v[i[\lambda v] \leqq R] < \varepsilon\}$$

is the information quantile capacity of the stationary channel $[B, v, C]$. Here, the ergodic decomposition is vital. In fact, the key step of the proof is the identification of $i[\lambda v](x, y)$ with $I(\omega_{xy})$, where ω_{xy} is the ergodic component in the Krylov-Bogoljubov ergodic decomposition of the double source λv (cf. [14] for the details concerning the ergodic decomposition for sources having alphabets standard spaces).

So far, our considerations included only easily verifiable stationarity and ergodicity properties of quantized sources and double sources. Now we shall analyze quantizers from the point of view of weak convergence.

Let $(A, \mathscr{A})$ be a standard space, and let $\{p_m\}_{m=1}^{\infty}$ be an asymptotically accurate sequence of quantizer maps. Let $\mu_n, \mu \in \mathscr{E}(A)$ and $\mu_n \to \mu$ weakly. The question reads to determine the properties of p_m which imply that $\mu_n p_m^{-1} \to \mu p_m^{-1}$ weakly, $m \geqq 1$. There are two natural conditions:

(a) $p_m : A \to \hat{A}(m)$ is continuous, $m = 1, 2, \ldots$;

(b) for each $E \in \bigcup_{m=1}^{\infty} p_m^{-1} \mathscr{V}_{A(m)}$ and for each $\mu \in \mathscr{E}(A)$, the boundary of E is μ-null.

Note (a) is included in (b) because in the former case each element of the family indicated in (b) is clopen and thus has empty boundary. Condition (b) actually asserts that the maps p_m are nearly continuous in the sense that the sets of discontinuity have probability zero with respect to any source $\mu \in \mathscr{E}(A)$ (see [8]).

Of course, these considerations apply equally well to double sources so that we introduce the following stronger notion of continuity. A stationary channel $[B, v, C]$ is said to *satisfy condition* (Q) if there exist asymptotically accurate sequences of quantizers of its alphabets such that for $\lambda_n, \lambda \in \mathscr{E}(B)$ with $\lambda_n \to \lambda$ weakly, we have that $\widehat{\lambda_n v}^{(m)} \to \widehat{\lambda v}^{(m)}$ weakly for any $m \geqq 1$.

Remark 1. It is an easy task to give an example of a channel which is weakly continuous but fails to possess the above continuity property. This amounts to finding a standard space for which no asymptotically accurate sequence of quantizers can preserve weak convergence. E.g., in case when B and C are the real line, the best we can do is to achieve that the union of all boundary points of all elements of the corresponding partitions is a countable dense set. Then it suffices to choose as $[B, \lambda]$ some periodic ergodic source supported by a subset of D^{∞}, where D is the above dense set.

Thus, condition (Q) is quite strong and, moreover, not easy to verify. On the other hand, there are several cases when (Q) is automatically satisfied. For example, if the channel alphabets are discrete countably infinite spaces then the requirement (a) above is satisfied so that (Q) is satisfied as well. This type of channel alphabets is not of much interest so that it would be desirable to find more general conditions than condition (Q). In the written version of the author's conference talk [15] it is shown that the following idea works well.

To this end observe the following. Let $\{(\mathscr{P}(m), \hat{A}(m), p_m)\}_{m=1}^{\infty}$ denote Gray-Kieffer sequence so that $P_0(m)$ is the unbounded component of $\mathscr{P}(m)$. Since $\mathscr{P}(m+1)$ refines $\mathscr{P}(m)$, given any compact set $K \subset A$ we find a m_0' such that for $m \geqq m_0'$, $P_0(m)$ does not intersect K. Consequently, given a $\varepsilon > 0$ we can find a m_0 such that for $m \geqq m_0$,

$$\sup_{a \in K} d_A(p_m(a), a) < \varepsilon ,$$

i.e., p_m converge uniformly on compact sets to the continuous identity map on A

By a result of Topsøe [16]. if a sequence $\{\varkappa_m\}_{m=1}^{\infty}$ of probability measures on $(A, \mathscr{A})$ converges weakly to $\varkappa$, then $\varkappa_m p_m^{-1} \to \varkappa$ weakly as well. Henceforth, if $\lambda_m, \lambda \in \mathscr{E}(A)$ and $\lambda_m \to \lambda$ weakly, then $(\lambda_m p_m^{-1})^N \to \lambda^N$ weakly on $(A^N, \mathscr{A}^N)$ for all $N \geqq 1$. Thus, for large m, the maps p_m behave as if they were nearly continuous. It is shown in [15] that this idea and the related kind of continuity relative to quantizers suffices in order to carry over all steps of the proofs given below in Section 5. Thus, the only condition really needed for the block transmission theorems is that of weak continuity.

4. MAIN RESULTS

So far, the concept of block transmissibility has been defined only for finite alphabets. In what follows, we put $\hat{Z}(k) = p_k(Z)$, $\hat{X}(m) = r_m(X)$, $\hat{Y}(m) = s_m(Y)$, and $\hat{V}(k) = p_k(V)$, respectively, where Z, X, Y, V are alphabet A, B, C, A processes and p_k, r_m, s_m are the corresponding quantizer maps. A source $[A, \mu]$ is said to be *block transmissible* over $[B, \nu, C]$ if there is a M_0 such that for any $m \geqq M_0$ and any $k \geqq 1$ the following holds: given $\varepsilon > 0$ there are a N and a pair of block codes of order N, say $\hat{\Phi}_1(k, m) : \hat{A}(k)^\infty \to \hat{B}(m)^\infty$ and $\hat{\Phi}_2(m, k) : \hat{C}(m)^\infty \to \hat{A}(k)^\infty$ such that

$$\text{Prob}\,[\hat{B}(k)^N \neq (\hat{\Phi}_2(m, k)\, \hat{Y}(m))^N] \leqq \varepsilon\,.$$

Our first result, the negative part of the coding theorem, requires only stationarity and existence of asymptotically accurate sequences of quantizers of the alphabets A, B, and C. Thus, we have the following assertion.

Theorem 1. Let A, B, and C be standard spaces. If a stationary source $[A, \mu]$ is block transmissible over a stationary channel $[B, \nu, C]$ then $H(\mu) \leqq C^*(\nu)$.

The positive part may be formulated as follows.

Theorem 2. Let $[A, \mu]$ be a stationary and ergodic source with alphabet a standard space. Assume that $[B, \nu, C]$ is a stationary channel satisfying condition (Q) for all positive entropy input sources $[B, \lambda] \in \mathscr{E}(B)$. If $H(\mu) < C^*(\nu)$ then $[A, \mu]$ is block transmissible over $[B, \nu, C]$.

Theorems 1 and 2 show that for channels which satisfy condition (Q) the information quantile capacity has the desired operational meaning in the sense that it can be expressed as some kind of operational joint source/channel block coding capacity (cf. [3] for a discussion concerning operational channel capacities).

5. THE PROOFS

Throughout the rest of the paper we shall use also the process notation for information theoretic quantities, e.g., $H(\mu) = H(X)$, $H(\hat{\mu}^{(k)}) = H(\hat{X}(k))$ for $\hat{\mu}^{(k)} =$

$= \mu p_k^{-1}$, $I(\hat{X}(m), \hat{Y}(m)) = I(\widehat{\lambda v^{(m)}})$, etc. We often do not specify the distribution of the process under consideration, this being always clear from the context.

Proof of Theorem 1. Assume that $[A, \mu]$ is block transmissible over $[B, v, C]$. That is, there is a M_0 such that for all $m \geqq M_0$ and all $k \geqq 1$ we can find block codes $\hat{\Phi}_1(k, m)$ and $\hat{\Phi}_2(m, k)$ of order N with the properties that

$$\text{Prob}\,[\hat{Z}(k)^N \neq \hat{V}(k)^N] \leqq \varepsilon\,,$$

where $\hat{V}(k) = \hat{\Phi}_2(m, k)\,\hat{Y}(m)$. For fixed m and k, N depends only on ε. The last inequality entails that

$$N^{-1} \sum_{i=0}^{N-1} \text{Prob}\,[\hat{Z}(k)_i \neq \hat{V}(k)_i] \leqq \varepsilon\,.$$

Let

$$H(\hat{Z}(k)^N \mid \hat{V}(k)^N) = H(\hat{Z}(k)^N, \hat{V}(k)^N) - H(\hat{V}(k)^N)\,.$$

As $\hat{Z}(k) = p_k(Z)$ and $V(k) = \hat{\Phi}_2(m, k)\,[s_m(Y)]$, the knowledge of μ and v specifies the probability vectors needed to calculate the entropies on the right hand side so that the above conditional entropy is well-defined. Let $K = \|\hat{A}(k)\|$ and, for $L \geqq 1$, let

$$q_L(\varepsilon) = -\varepsilon \log \varepsilon - (1 - \varepsilon) \log (1 - \varepsilon) + \varepsilon \log L\,.$$

By Fano's inequality

$$H(\hat{Z}(k)_i \mid \hat{V}(k)_i) \leqq q_K(\text{Prob}\,[\hat{Z}(k)_i \neq \hat{V}(k)_i])$$

so that

$$H(\hat{Z}(k)^N \mid \hat{V}(k)^N) \leqq N^{-1} \sum_{i=0}^{N-1} H(\hat{Z}(k)_i \mid \hat{V}(k)_i \leqq$$

$$\leqq N^{-1} \sum_{i=0}^{N-1} q_K(\text{Prob}\,[\hat{Z}(k)_i \neq V(k)_i]) \leqq q_K(N^{-1} \sum_{i=0}^{N-1} \text{Prob}\,[\hat{Z}(k)_i \neq \hat{V}(k)_i]) \leqq q_K(\varepsilon)\,.$$

Now $\hat{X}(m) = \hat{\Phi}_1(k, m)\,\hat{Z}(k)$ is N-stationary and we denote by $h(\hat{X}(m))$ the limit function from Shannon-McMillan-Breiman theorem for $\hat{X}(m)$. Also let $h(\hat{X}(m) \mid \hat{Y}(m))$ designate the limit function for the conditioned process $\hat{X}(m) \mid \hat{Y}(m)$ (for this it suffices again to require the knowledge of the joint distribution of $\hat{X}(m)$ and $\hat{Y}(m)$ and of its marginals). Then

$$i(\hat{X}(m), \hat{Y}(m)) = h(\hat{X}(m)) + h(\hat{Y}(m)) - h(\hat{X}(m), \hat{Y}(m)) =$$
$$= h(\hat{X}(m)) - h(\hat{X}(m) \mid \hat{Y}(m))\,.$$

Using standard techniques based on Markov inequality (cf. [3] or the proof of Theorem 10 in [9] we get the estimates

$$\text{Prob}\,[h(\hat{X}(m)) \leqq H(\hat{Z}(k)) - q_K(\varepsilon)^{1/2}] \leqq q_K(\varepsilon)^{1/2}\,;$$
$$\text{Prob}\,[h(\hat{X}(m) \mid \hat{Y}(m)) \geqq q_K(\varepsilon)^{1/2}] \leqq q_K(\varepsilon)^{1/2}$$

so that the density of the mutual information rate $i(\hat{X}(m), \hat{Y}(m))$ cannot be much smaller than the entropy of the source to be transmitted, i.e.,

$$\mathrm{Prob}\left[i(\hat{X}(m), \hat{Y}(m)) \leqq H(\hat{Z}(k)) - 2q_K(\varepsilon)^{1/2}\right] \leqq 2\, q_K(\varepsilon)^{1/2}.$$

Let

$$C_m^*(\varepsilon) = \sup_{\lambda \in \mathscr{E}(B)} \sup \left\{R \middle| \widehat{\lambda\nu}^{(m)}\left[i[\widehat{\lambda\nu}^{(m)}] \leqq R\right] < \varepsilon\right\}.$$

One can easily check (cf., e.g., [17]) that $\mathscr{E}(B)$ can be replaced by $\mathscr{M}_b(B)$ so that we obtain that

$$C_m^*(3\, q_k(\varepsilon)^{1/2}) \geqq H(\hat{\mu}^{(k)}) - 2\, q_K(\varepsilon)^{1/2}.$$

By letting $\varepsilon \to 0+$ we get $C_m^* \geqq H(\hat{\mu}^{(k)})$. But this conclusion is valid for all $m \geqq M_0$ and all $k \geqq 1$ so that we finally get $C^*(\nu) \geqq H(\mu)$. □

Remark 2. Our proof followed the lines of the proof of Theorem 10 in [9]. Alternatively, we could proceed as in [3]. However, the proof would become much more complicated in that it would again constitute a two-step procedure in the sense described in the introduction.

Proof of Theorem 2. The proof will follow the lines of [5], Section VI. For finite alphabets A, B call a map $\Phi : A^\infty \to B^\infty$ a sliding-block code if $\Phi \circ T_A = T_B \circ \Phi$ and there is $\Phi' : A^{2m+1} \to B$ such that

$$\Phi(z)_0 = \Phi'(z_{-m}^{2m+1}), \quad z \in A^\infty,$$

where, in accordance with our notations, z_{-m}^{2m+1} denotes the vector $(z_{-m}, \ldots, z_m)$. As in [5] we start with the following assertion.

(*) Let $[B, \lambda] \in \mathscr{E}(B)$, $R > 0$, $0 < \varepsilon < 1$ be such that

$$\widehat{\lambda\nu}^{(m)}\left[i[\widehat{\lambda\nu}^{(m)}] \leqq R\right] < \varepsilon.$$

Then there exist sliding-block codes $\Phi : \hat{B}(m)^\infty \to \hat{B}(m)^\infty$,

$$\Psi : \hat{B}(m)^\infty \times \hat{C}(m)^\infty \to \hat{B}(m)^\infty \quad \text{such that} \quad H(\hat{X}(m) \mid \Phi(\hat{X}(m))) > R,$$

$$\int d_1(\hat{X}(m)_0, \Psi(\hat{X}(m)), \hat{Y}(m))_0)\, \mathrm{d}\widehat{\lambda\nu}^{(m)} < \varepsilon,$$

where d_1 is the usual Hamming distance. Using the results obtained in [6] on quantization of the mutual information rate one can easily extend well-known results of Jacobs [18] to the abstract alphabet setup. In particular, there is a $T_{B\times C}$-invariant function i on $B^\infty \times C^\infty$ such that for any stationary channel $[B, \tilde{\nu}, C]$,

$$\lambda\tilde{\nu}[i = i[\lambda\tilde{\nu}]] = 1.$$

Now, we know that $i[\lambda\tilde{\nu}^{(m)}] \uparrow i[\lambda\tilde{\nu}]$, and as shown in [6], there are invariant functions $\hat{\imath}(m)$ on $\hat{B}(m)^\infty \times \hat{C}(m)^\infty$ for which

$$\widehat{\lambda\tilde{\nu}}^{(m)}[\hat{\imath}(m) = i[\widehat{\lambda\tilde{\nu}}^{(m)}]] = 1$$

and $\hat{i}(m) \uparrow i$ as $m \to \infty$. By observing that Theorem 4 of [19] makes use only of the double sources and their input marginals we can proceed as in [5], proof of Lemma 1, and get sliding-block codes Φ and Ψ with the desired properties.

So, let $[B, v, C]$ be weakly continuous at an input source $[B, \lambda] \in \mathscr{E}(B)$, and let us assume that

$$\widehat{\lambda v}^{(m)}[i(\widehat{\lambda v}^{(m)}] \leqq R] < \varepsilon .$$

By $(*)$ we find sliding-block codes Φ and Ψ such that

$$H(\hat{X}(m) \mid \Phi(\hat{X}(m))) > R ,$$

$$\int d_1(\hat{X}(m)_0, \Psi(\Phi(\hat{X}(m)), \hat{Y}(m))_0) \, \mathrm{d}\, \widehat{\lambda v}^{(m)} < \varepsilon .$$

By the definition of a sliding-block code there exists a positive integer k such that for each $n > 2k$ there is a map $\Phi_n : \hat{B}(m)^n \to \hat{B}(m)^{n-2k}$ with the property that

$$\Phi(\hat{x})_{k+1}^{n-2k} = \Phi_n(\hat{x}^n) , \quad \hat{x} \in \hat{B}(m)^\infty ,$$

where $\hat{x}_{k+1}^{n-2k} = (\hat{x}_{k+1}, \ldots, \hat{x}_{n-k})$. Using the fact that the map $\tau \mapsto \widehat{\tau v}^{(m)}$ is continuous at λ we can, given $\delta > 0$, find a pair (M, α) such that, if

$$\max_{\hat{b}^M \in \hat{B}(m)^M} \left|\hat{\tau}^{(m)M}(\hat{b}^M) - \hat{\lambda}^{(m)M}(\hat{b}^M)\right| < \alpha$$

then

$$\left|\int d_1(\hat{X}(m)_0, \Psi(\Phi(\hat{X}(m)), \hat{Y}(m))_0) \, \mathrm{d}\, \widehat{\tau v}^{(m)} - \right.$$

$$\left. - \int d_1(\hat{X}(m)_0, \Psi(\Phi(\hat{X}(m)), \hat{Y}(m))_0) \, \mathrm{d}\, \widehat{\lambda v}^{(m)}\right| < \delta .$$

A sequence $\hat{x}^n \in \hat{B}(m)^n$ is said to be (M, α)-typical for a stationary source $[\hat{B}(m), \hat{\tau}^{(m)}]$ if $n \geqq M$ and the distance between $\hat{\tau}^{(m)M}(\hat{b}^M)$ and the relative frequency of $\hat{b}^M$ in $\hat{x}^n$ is less than α, for all $\hat{b}^M \in \hat{B}(m)^M$. Let $[B, \tau] \in \mathscr{M}^n(B)$ so that $[\hat{B}(m), \hat{\tau}^{(m)}] \in \mathscr{M}^n(\hat{B}(m))$. Then the source

$$\overline{\hat{\tau}^{(m)}} = n^{-1} \sum_{i=0}^{n-1} \hat{\tau}^{(m)} T^i_{\hat{B}(m)}$$

is stationary. Moreover, if $[\hat{B}(m), \hat{\tau}^{(m)}] \in \mathscr{E}^n(\hat{B}(m))$ then $\overline{\hat{\tau}^{(m)}} \in \mathscr{E}(\hat{B}(m))$. It follows as in [5] that given (M, α) for $n \geqq N(M, \alpha)$ we have that

$$\max_{\hat{b}^M \in \hat{B}(m)^M} \left|\overline{\hat{\tau}^{(m)}}^M(\hat{b}^M) - \hat{\lambda}^{(m)M}(\hat{b}^M)\right| < \alpha$$

for every $[B, \tau] \in \mathscr{M}^n(B)$ such that $\hat{\tau}^{(m)n}$ places mass one on a set of sequences from $\hat{B}(m)^n$ which are $(M, \alpha/2)$-typical of $\hat{\lambda}^{(m)}$. Moreover

$$E_{\widehat{\bar{\tau} v}^{(m)}} d_1(\hat{X}(m)_0, \Psi(\Phi(\hat{X}(m)), \hat{Y}(m))_0) =$$

$$= \int d_1(\hat{X}(m)_0, \Psi(\Phi(\hat{X}(m)), \hat{Y}(m))_0) \, \mathrm{d}\Big[\widehat{(n^{-1} \sum_{i=0}^{n-1} \tau T^i_B) \, v^{(m)}}\Big] =$$

$$= \int n^{-1} \sum_{i=0}^{n-1} d_1(\hat{X}(m)_0, \Psi(\Phi(\hat{X}(m)), \hat{Y}(m))_0) \, \mathrm{d}[\widehat{(\tau T_B^i) \, v^{(m)}}] =$$

$$= \int n^{-1} \sum_{i=0}^{n-1} d_1(\hat{X}(m)_i, \Psi(\Phi(\hat{X}(m)), \hat{Y}(m))_i) \, \mathrm{d}\widehat{\tau v}^{(m)} =$$

$$= E_{\widehat{\tau v}^{(m)}} d_n(\hat{X}(m)^n, \Psi(\Phi(\hat{X}(m)), \hat{Y}(m))^n) .$$

Finally, given (M, α), for n large enough there is a set $W_n^{(m)} \subset \hat{B}(m)^n$ such that $\|W_n^{(m)}\| > \exp_2(nR)$, Φ_n is constant on $W_n^{(m)}$, and any sequence in $W_n^{(m)}$ is $(M, \alpha/2)$-typical of $\hat{\lambda}^{(m)}$. Using these properties, one can find robust decoding sets in the sense of [3] and thereby obtain a $(\|W_n^{(m)}\|, n, \sqrt{\varepsilon})$-channel code with slightly decreased rate, viz.

$$\|W_n^{(m)}\| \geqq \exp_2 n(R - q_{\|\hat{B}(m)\|}(2\sqrt{\varepsilon})) .$$

Now the proof of Theorem 2 follows standard arguments. Indeed, let $H(\mu) < C^*(\nu)$. We can choose $\gamma > 0$ and $\varepsilon > 0$ so small that

$$R - q_m(2\sqrt{\varepsilon}) > H(\mu) + \gamma$$

for some fixed $m \geqq 1$ satisfying $H(\mu) < C_m^*$ (in this case, $H(\hat{\mu}^{(k)}) < C_m^*$ for all $k \geqq 1$). By the noiseless source coding theorem of [20] we encode $[\hat{A}(k), \hat{\mu}^{(k)}, \hat{Z}(k)]$ into a n-stationary channel input process which puts probability one on the set $W_n^{(m)}$ appearing above. This gives then the desired result. □

Remark 3. A direct proof can be based on the fact that

$$H(\hat{\mu}^{(k)}) = \lim_{n \to \infty} n^{-1} \log L_n(\varepsilon, \hat{\mu}^{(k)}) , \quad 0 < \varepsilon < 1$$

where

$$L_n(\varepsilon, \hat{\mu}^{(k)}) = \min \{ \|E\| \mid E \subset \hat{A}(k)^n, \hat{\mu}^{(k)}[C^n(E)] > 1 - \varepsilon \}$$

(cf. [20] for the proof). The idea is to couple this result directly with the fact that $H(\hat{\mu}^{(k)}) < C_m^*$; Winkelbauer [21] used quite strong conditions on the channel in order to construct the desired joint source/channel code. However, his idea can be modified in the spirit of the above proof. On the other hand, Kieffer's approach is more informative because it gives positive results on transmissibility also in the case when the channel is weakly continuous only locally, i.e., at some fixed input source.

6. CONCLUSION

Kieffer ([9] and [5]) completely clarified the relations between weak continuity and $\bar{d}$-continuity. In fact, he shows in [9] that any $\bar{d}$-continuous channel is weakly continuous, and in [5] he gives an example of a weakly continuous channel which is not $\bar{d}$-continuous in a non-trivial way (this means that the channel is not equivalent to a $\bar{d}$-continuous one; cf. [5]). Moreover, he shows that every stationary channel is "almost" continuous with respect to any input source $[B, \lambda] \in \mathscr{S}(B)$, in the sense that we can construct a weakly continuous channel $[B, \hat{\nu}(\lambda), C]$ for which $\nu_x = \hat{\nu}(\lambda)_x$ for λ almost all $x \in B^\infty$.

In case of metric alphabets we can replace the notion of $\bar{d}$-continuity by that of $\bar{\varrho}$-continuity, where the $\bar{\varrho}$-distance was introduced in [22] as a generalization of Ornstein's $\bar{d}$-distance. Another possibility is to employ the process definition of $\bar{\varrho}$-distance also introduced in [22]. We thus get three notions of continuity and it is an interesting open problem to clarify the relations among them.

ACKNOWLEDGMENT

Thanks are due to John C. Kieffer for his kind submission of the paper [5] prior to publication.

(Received January 22, 1981.)

REFERENCES

[1] K. Winkelbauer: On the coding theorem for decomposable channels I, II. Kybernetika *7* (1971), 109–123, 230–255.

[2] J. C. Kieffer: A general formula for the capacity of a stationary nonanticipatory channel. Inform. and Control *26* (1974), 381–391.

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

[4] J. Wolfowitz: Coding Theorems of Information Theory. 2nd ed. Springer-Verlag, Berlin–Göttingen–New York 1964.

[5] J. C. Kieffer: Block coding for a stationary channel satisfying a weak continuity condition (to appear).

[6] R. M. Gray, J. C. Kieffer: Mutual information rate, distortion and quantization in metric spaces. IEEE Trans. Inform. Theory *26* (1980), 412–422.

[7] Š. Šujan: Channels with additive asymptotically mean stationary noise. Kybernetika *17* (1981), 1, 1–15.

[8] P. Billingsley: Convergence of Probability Measures. J. Wiley, New York–London–Sydney–Toronto 1968.

[9] J. C. Kieffer: On the transmission of Bernoulli sources over stationary channels. Ann. Prob. 8 (1980), 942–961.

[10] Š. Šujan: A generalized coding problem for discrete information sources. Supplement. Kybernetika *13* (1977), 95 pp.

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

[12] R. L. Dobrushin: A general formulation of the basic Shannon theorem of information theory (in Russian). Uspehi mat. nauk *14* (1959), 3–104.

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

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

[15] Š. Šujan: Block transmissibility and quantization (submitted).

[16] F. Topsøe: Preservation of weak convergence under mappings. Ann. Math. Statist. *38* (1967), 1661–1665.

[17] Š. Šujan: On the capacity of asymptotically mean stationary channels. Kybernetika *17* (1981), 3, 222—233.

[18] K. Jacobs: Über die Struktur der mittleren Entropie. Math. Z. *78* (1962), 33—43.

[19] J. C. Kieffer: Some universal noiseless multiterminal source coding theorems. Inform. and Control *46* (1980), 93—107.

[20] K. Winkelbauer: On discrete information sources. Trans. 3rd Prague Conf. Inform. Theory, NČSAV Prague 1964, 765—830.

[21] K. Winkelbauer: On the capactiy of decomposable channels. Trans. 6th Prague Conf. Inform. Theory, Academia, Prague 1973, 903—914.

[22] R. M. Gray, D. L. Neuhoff, P. C. Shields: A generalization of Ornstein's $\bar{d}$-distance with applications to information theory. Ann. Prob. *3* (1975), 315—328.

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