Štefan Šujan

Some stationary source and joint source-channel coding theorems with a fidelity criterion

Kybernetika, Vol. 22 (1986), No. 6, 461--470

Persistent URL: http://dml.cz/dmlcz/124452

KYBERNETIKA — VOLUME 22 (1986), NUMBER 6

SOME STATIONARY SOURCE AND JOINT SOURCE-CHANNEL CODING THEOREMS WITH A FIDELITY CRITERION

ŠTEFAN ŠUJAN, Posthumous[1]

A simple direct proof of the strong Sinai's theorem involving an average distortion constraint is given. Stationary information transmission theorems are established for ergodic continuous alphabet sources and ergodic and weakly continuous channels, generalizing previous results of Gray (for B-sources and noiseless channels) and of Gray and Ornstein (for ergodic sources and discrete memoryless channels). In the special case of finite alphabet B-sources the use of rather complex sliding-block source coding theorem can be replaced by simple considerations based on the strong Sinai's theorem mentioned above.

1. INTRODUCTION

The information transmission theorem for block codes; cf. [1, Thm. 7.2.6] asserts that a source can be transmitted over a channel of capacity $C > R(D)$ in such a way that the channel output process reproduces the original process with fidelity D, provided the block length is sufficiently large and, on the other hand, if $C < R(D)$, then fidelity D is not achievable whatever the block length will be. Here, $R(\cdot)$ stands for the distortion-rate function of the source.

The main problem addressed in this paper is a dual form of the information transmission theorem. That is, we are given a source $[A, \mu]$, having the distortion-rate function (DRF) relative to some distortion measure, and a channel $[B, \nu, C]$ with capacity $\mathscr{C}$. Then we ask which are the connections between the optimum performance theoretically attainable using encoder-decoder pairs $(\bar{\varphi}, \bar{\psi})$; $\bar{\varphi}: A^\infty \to B^\infty$, $\bar{\psi}: C^\infty \to A^\infty$, and the capacity $\mathscr{C}$. We are interested in the case when $\bar{\varphi}$, $\bar{\psi}$ are stationary (finite or infinite) codes. Denoting the corresponding OPTA by δ^*, the result we seek for is

$$(*) \qquad \delta^* = D(\mathscr{C}) .$$

This result was previously obtained by Gray [2] for B-sources and noiseless

[1] Obituary of Štefan Šujan appeared in Kybernetika vol. *21* (1985), No. 6, pp. 482—483.

channels, by Gray and Ornstein [3] for ergodic sources and discrete memoryless channels, and by Gray, Ornstein, and Dobrushin [4] for B-sources and totally ergodic $\bar{d}$-continuous channels. We extend it to arbitrary ergodic sources and ergodic weakly continuous channels. The proof makes use of the sliding-block coding theorem with a fidelity criterion which was obtained in [5] using somewhat complex arguments. In the finite alphabet case and for B-sources it is possible to replace it by a strong form of Sinai's theorem (a slight generalization of Theorem 1 in [6]). Its formulation and a simple direct proof will be given in the next section.

2. STRONG SINAI'S THEOREM

Let A, $\hat{A}$ be finite sets, and suppose that ϱ: $(A \cup \hat{A}) \times (A \cup \hat{A}) \to [0, \infty)$ is a finite-valued metric on $A \cup \hat{A}$. Let $\varrho_M = \max\{\varrho(a', a''): a', a'' \in A \cup \hat{A}\}$. A process is defined to be a bilateral sequence $X = \{X_i\}_{i=-\infty}^{\infty}$ of random variables defined on some common probability space ($(\Omega, \mathscr{P}, \mathsf{P})$, say), and taking values in a common finite set (called the state space of X). If the state space of X is A then, by regarding X as a mapping from Ω to A^∞, the measurable space of all bilateral sequences from A, we let $\operatorname{dist}(X)$ denote the induced probability measure on A^∞. If $\operatorname{dist}(X) = \mu$, then we shall write $[A, \mu]$ or $[A, \mu, X]$ for the corresponding source [7]. A process X or, a source $[A, \mu]$ is said to be stationary (ergodic) if $\operatorname{dist}(X)$ or, μ, is invariant (ergodic) with respect to the shift T on A^∞ (we shall use T for the shift in any space of the form C^∞, the state space C being always clear from the context). Similar comments apply to pair processes (X, Y), triple processes (X, Y, Z), etc. [8].

If X is a process with state space A, Y a process with state space $\hat{A}$, let $X \vee Y$ denote the set of all jointly invariant measures on $(A \times \hat{A})^\infty$ having $\operatorname{dist}(X)$ and $\operatorname{dist}(Y)$ as marginals. If the pair process (X, Y) is ergodic then we define the $\bar{\varrho}$-distance between X and Y as

$$\bar{\varrho}(X, Y) = \inf_{p \in X \vee Y} \mathsf{E}_p\, \varrho(X_0, Y_0)$$

(see [6, (12)]; more on $\bar{\varrho}$-distance see in [9]). In particular, if ϱ is a metric, then so is $\bar{\varrho}$.

A process X with a finite state space (a source $[A, \mu]$ with a finite alphabet A) is said to be a B-process (a B-source) if it admits an isomorphic representation in the form of a bilateral sequence of independent, identically distributed (i.i.d.) random variables (cf., e.g., [10, 11, 7]).

Theorem 1 (Strong Sinai's Theorem). Let A, $\hat{A}$, and ϱ be specified as above. Let X be a stationary, ergodic, and aperiodic (meaning that $\mathsf{P}[X = x] = 0$, $x \in A^\infty$) process with state space A, and let Y be a B-process with state space $\hat{A}$ such that their entropies satisfy $h(X) \geqq h(Y)$. Then for any $\gamma > 0$ there is a stationary code $\bar{f}$: $A^\infty \to \hat{A}^\infty$ such that

(i) $\bar{f}X = Y$, and

(ii) $\mathsf{E}_\mu\, \varrho(X_0, (\bar{f}X)_0) \leqq \bar{\varrho}(X, Y) + \gamma$; $\quad \mu = \operatorname{dist}(X)$.

Recall from [7] that a measurable mapping $\bar{f}$: $A^\infty \to \hat{A}^\infty$ is said to be a stationary (or, infinite) code if $\bar{f} \circ T = T \circ \bar{f}$. A stationary code $\bar{f}$ is said to be a sliding-block (or, finite) code if there exists a positive integer N and a mapping f: $A^{2N+1} \to \hat{A}$ such that $(\bar{f}x)_0 = f(x_{-N}^N)$; $x \in A^\infty$. The window length N will be sometimes indicated by writing $\bar{f}^{(N)}$ instead of $\bar{f}$.

Assertion (i) is the usual form of Sinai's theorem: each ergodic aperiodic finite state process X has Bernoulli factors with entropy H for any $0 < H \leqq h(X)$ [10, 7]. The expectation in (ii) is usually denoted by $\varrho_X(\bar{f})$ (or, $\varrho_\mu(\bar{f})$ if $\mu = \text{dist}(X)$) – it is the average distortion when $\bar{f}$ is used to code X.

The special case when Y is an i.i.d. process is Theorem 1 in [6]. The proof from [6] easily extends to our case, however, it is based on a rather complex construction from [5]. Here we shall give a simple direct proof based on the following lemma:

Lemma 2. Let (X, Y) be a stationary and ergodic pair process with X and Y each having a finite state space such that X is aperiodic and $h(X) \geqq h(Y)$. Then there is a sequence $Y^{(N)}$ of processes having the same state space as Y for which

(i) each $Y^{(N)}$ is a finite coding of X;

(ii) $(X, Y^{(N)}) \to (X, Y)$ in distribution; and

(iii) $h(Y^{(N)}) \to h(Y)$ as $N \to \infty$.

This is a special case of Lemma 1 in [8] (cf. also [12]). As shown in [8, Sect. IV], Lemma 2 follows easily from the Shannon-McMillan theorem using a standard construction of good sliding-block codes from good block codes (see [13, 7, 11]).

Proof of Theorem 1. By redefining the underlying probability space, if necessary, we can and do assume that (X, Y) is jointly stationary and ergodic. By Lemma 2 we find a sequence $Y^{(N)}$ of sliding-block codings of X such that

(a) $(X, Y^{(N)}) \to (X, Y)$ in distribution,

(b) $h(Y^{(N)}) \to h(Y)$.

Since Y is a B-process, it is finitely determined [10, 11]. Consequently, (a) and (b) yield

(c) $\bar{d}(Y^{(N)}, Y) \to 0$,

where $\bar{d}$ is Ornstein's $\bar{d}$-distance ([10], $\bar{d}$-distance is the special case of $\bar{\varrho}$-distance when the role of ϱ is played by the Hamming distance d_H). Since ϱ is a finite-valued metric, for any $b', b'' \in \hat{A}$, $\varrho(b', b'') \leqq \varrho_M d_H(b', b'')$. Hence also $\bar{\varrho}(Y^{(N)}, Y) \leqq \leqq \varrho_M \bar{d}(Y^{(N)}, Y)$ so that (c) implies

(d) $\bar{\varrho}(Y^{(N)}, Y) \to 0$.

Using (d) we may pick a good initial coding of X. Take some $\varepsilon > 0$ and pick a sequence $\varepsilon_1, \varepsilon_2, \ldots$ of positive numbers such that $\sum \varepsilon_i < \varepsilon$. By (d), choose some stationary coding Z of X for which $\bar{\varrho}(Z, Y) < \varepsilon_1$. Since $\varepsilon_1 < \sum \varepsilon_i$, by taking $\varepsilon < \gamma/3$ we can find (from the definition of $\bar{\varrho}$) a process $\hat{Y}$ having the same state space as Y, such that $(X, \hat{Y})$ is jointly ergodic, $\text{dist}(\hat{Y}) = \text{dist}(Y)$,

(e) $\mathsf{E}_\mu\, \varrho(\hat{Y}_0, Z_0) < \varepsilon_1$, and

(f) $\mathsf{E}_\mu\, \varrho(X_0, \hat{Y}_0) < \bar{\varrho}(X, Y) + \gamma/3$.

Consider the pair process $((X, Z), \hat{Y})$. Since $\hat{Y}$ is again a B-process, using the property of finitely determined we find a stationary coding $Z^{(1)}$ of X such that $((X, Z), Z^{(1)})$ is as close in distribution to $((X, Z), \hat{Y})$, and $h(Z^{(1)})$ so close to $h(\hat{Y})$ that

$$\mathsf{E}_\mu\, \varrho(Z_0, Z_0^{(1)}) < \varepsilon_1\,, \quad \bar{\varrho}(Z^{(1)}, Y) < \varepsilon_2\,.$$

(In fact, finitely determined implies the latter inequalities with d_H and $\bar{d}$ in place of ϱ and $\bar{\varrho}$, but we can pass to ϱ and $\bar{\varrho}$ as above from (c) to (d)). By making use of this argument repeatedly we find a sequence $Z^{(1)}, Z^{(2)}, \ldots$ of stationary codings of X such that

(g) $\bar{\varrho}(z^{(i)}, Y) < \varepsilon_{i+1}$,

(h) $\mathsf{E}_\mu\, \varrho(Z_0^{(i-1)}, Z_0^{(i)}) < \varepsilon_i$.

Since we know by construction that also

$$\sum_i \mathsf{E}_\mu\, d_H(Z_0^{(i-1)}, Z_0^{(i)}) = \sum_i \operatorname{Prob}\left[Z_0^{(i-1)} \neq Z_0^{(i)}\right] < \Sigma\varepsilon_i < \infty\,,$$

there exists a.s. limit, $\hat{Z}$, of the sequence $Z^{(i)}$. By (g) and (h) we see that

(j) $\bar{\varrho}(\hat{Z}, Y) = 0$ (in particular, $\operatorname{dist}(\hat{Z}) = \operatorname{dist}(Y)$);

(k) $\mathsf{E}_\mu\, \varrho(\hat{Z}_0, Z_0) < \sum \varepsilon_i < \frac{1}{3}\gamma$.

Since $\varepsilon_1 < \frac{1}{3}\gamma$, from (k), (f), and (e) we get

$$\begin{aligned} (1) \qquad & \mathsf{E}_\mu\, \bar{\varrho}(X_0, \hat{Z}_0) \leqq \mathsf{E}_\mu\, \varrho(X_0, \hat{Y}_0) + \mathsf{E}_\mu\, \varrho(\hat{Y}_0, Z_0) + \\ & + \mathsf{E}_\mu\, \varrho(Z_0, \hat{Z}_0) < \bar{\varrho}(X, Y) + \tfrac{1}{3}\gamma + \varepsilon_1 + \tfrac{1}{3}\gamma < \bar{\varrho}(X, Y) + \gamma\,. \end{aligned}$$

It remains to show that $\hat{Z}$ is a stationary coding of X, too. To see this, let $\bar{f}^{(i)}$ denote the stationary code with $\bar{f}^{(i)}X = Z^{(i)}$. Define the distance

$$\begin{aligned} |\bar{f}^{(i)} - \bar{f}^{(i-1)}|_X &= \operatorname{Prob}\left[(\bar{f}^{(i)}X)_0 \neq (f^{(i-1)}X)_0\right] = \\ &= \operatorname{Prob}\left[Z_0^{(i-1)} \neq Z_0^{(i)}\right]. \end{aligned}$$

As shown in [4], the space of all stationary codes $\bar{f}$ from A^∞ to $\hat{A}^\infty$ is complete relative to that metric so that $|\bar{f}^{(i)} - \bar{f}|_X \to 0$ for some $\bar{f}$. Since $\hat{Z}$ is an a.s. limit of $Z^{(i)} = \bar{f}^{(i)}X$, it follows that $\hat{Z} = \bar{f}X$. By (k) and (1), $\bar{f}$ has the desired properties. □

We have proved actually a little bit more in the course of passing from (d) till the end of the proof:

Corollary 3. Let the hypotheses of Theorem 1 prevail. Suppose there is a stationary coding Z of X such that $\bar{\varrho}(Z, Y) < \varepsilon$. Then there is a stationary coding $\hat{Z}$ of X for which

(i) $\operatorname{dist}(\hat{Z}) = \operatorname{dist}(Y)$, and

(ii) $\mathsf{E}_\mu\, \varrho(\hat{Z}_0, Z_0) < \varepsilon$.

This form of Sinai's theorem (with d_H replacing ϱ and the relative $\bar{d}$-distance replacing $\bar{d}$) was obtained in [8]. In case of B-processes, Theorem 1 makes it possible to get a different expression of the DRF in the spitit of the "topological" approach to source coding with a fidelity criterion [5, 15].

Corollary 4. Let A, $\hat{A}$, and ϱ be specified as at the beginning of this section. Let X be a B-process with the state space A. Then for any $R \in [0, h(X)]$,

$$D(R) = \inf\{\bar{\varrho}(X, Y): h(Y) \leqq R, Y \text{ is a } B\text{-process}\},$$

where $D(\cdot)$ is the DRF of the process X.

Proof. This follows from the following chain of equalities:

$$D(R) =^{(a)} \inf\{\varrho_X(\bar{f}): h(\bar{f}X) \leqq R\} =^{(b)}$$

$$=^{(b)} \inf\{\varrho_X(\bar{f}): h(\bar{f}X) \leqq R, \bar{f}X \text{ is a } B\text{-process}\} =^{(c)}$$

$$=^{(c)} \inf\{\bar{\varrho}(X, Y): h(Y) \leqq R, Y \text{ is a } B\text{-process}\}.$$

Equation (a) is but of the process definitions of the DRF (valid, by [14], for any ergodic process X). Equality (b) is trivial for, if X is a B-process then so is $\bar{f}X$ for any stationary code $\bar{f}$ [10, 11]. Equality (c) was proved in [6] in case when $\bar{f}X$ and Y were restricted to i.i.d. processes. Using our Theorem 1 instead of [6, Thm. 1] we can repeat the quoted proof word by word, and thereby get (c). □

In particular, we get the following strengthening of the (stationary) source coding theorem for B-processes:

Corollary 5. Let A, $\hat{A}$, ϱ, and X be as in the preceding corollary. Let $0 \leqq R \leqq h(X)$. Then for any $\varepsilon > 0$ there exists a B-process Y with the state space $\hat{A}$ and a stationary code $\bar{f}: A^\infty \to \hat{A}^\infty$ such that
(i) $\bar{f}X = Y$, and
(ii) $\varrho_X(\bar{f}) \leqq D(R) + \varepsilon$.

Proof. Given R and $\varepsilon > 0$, by Corollary 4 find a B-process Y with $h(Y) \leqq R$ such that
(a) $\bar{\varrho}(X, Y) \leqq D(R) + \frac{1}{2}\varepsilon$.
Since $h(Y) \leqq R \leqq h(X)$, we may apply Theorem 1 (with $\gamma = \frac{1}{2}\varepsilon$) in order to find a stationary code $\bar{f}$ such that (i) is true and
(b) $\varrho_X(\bar{f}) \leqq \bar{\varrho}(X, Y) + \frac{1}{2}\varepsilon$.
Assertion (ii) follows from (a) and (b). □

3. THE INFORMATION TRANSMISSION THEOREM

Throughout this section we are given an ergodic source $[A, \mu, U]$ and a finite alphabet channel $[B, \nu, C]$ such that either

(a) $A \cup B$ is finite and ϱ is a pseudo-metric distortion measure (i.e., satisfies the triangle inequality),

or,

(b) $A \cup B$ is a complete separable metric space under a metric ϱ and there exists a reference letter $a_0 \in A$ for which

$$\mathsf{E}_\mu\, \varrho(U_0, a_0) < \infty .$$

Recall that a channel $[B, v, C]$ is a measurable family $(v_x(\cdot); x \in B^\infty)$ of probability measures on C^∞. We suppose that v is stationary, i.e.,

$$\gamma_{Tx}(TF) = v_x(F); x \in B^\infty, F \subset C^\infty$$

(by writing inclusions like $F \subset C^\infty$ we shall automatically assume F is measurable). A channel $[B, v. C]$ is said to be ergodic if for any stationary and ergodic input source $[B, \lambda]$ the joint input-output source $[B \times C, \lambda v]$ is also stationary and ergodic, where

$$\lambda v(E \times F) = \int_E v_x(F)\, \lambda(\mathrm{d}x);\; E \subset B^\infty,\; F \subset C^\infty .$$

Recall from [13] that $[B, v, C]$ is weakly continuous if for any sequence of stationary and ergodic input sources $[B, \lambda^{(n)}]$ such that $\lambda^{(n)}$ weakly converges to a stationary and ergodic source $[B, \lambda]$, the measures $\lambda^{(n)}v$ weakly converge to λv.

Given $[A, \mu]$ and $[B, v, C]$, a quadruple (U, X, Y, V) is said to be a stationary (infinite or finite) hookup of the given source and channel, if U has the state space A, the state space of V is contained in A, X has the state space B, Y has the state space C, and if (U, X, Y, V) forms a Markov chain in the sense that there exist stationary (infinite or finite) codes $\bar{\varphi}: A^\infty \to B^\infty$, $\bar{\psi}: C^\infty \to A^\infty$ such that $\operatorname{dist}(U) = \mu$, $X = \bar{\varphi}U$, $\operatorname{dist}(X, Y) = (\mu\bar{\varphi}^{-1})\, v$, and $V = \bar{\psi}Y$ (note that $\bar{\varphi}$ is well-defined also for continuous alphabet A).

Let $\mathscr{C}(v)$ denote the Shannon (= information rate) capacity of $[B, v, C]$ (cf., e.g., [16]). Thus,

$$\mathscr{C}(v) = \sup I(X, Y) ,$$

where the supremum is over all stationary and ergodic processes X with the state space B, Y is the output process of v given X at the input, and $I(X, Y)$ is the average mutual information rate [1, 14, 16].

Theorem 6 (Information Transmission Theorem). Let A, B, C, and ϱ be given as specified above. Let $[A, \mu]$ be a stationary and ergodic source with the DRF $D(\cdot)$. Let $[B, v, C]$ be an ergodic and weakly continuous channel with the Shannon capacity $\mathscr{C}(v)$. Given stationary codes $\bar{\varphi}: A^\infty \to B^\infty$, and $\bar{\psi}: C^\infty \to A^\infty$, let us denote by (U, X, Y, V) the corresponding hookup. Then

$$\inf_{(\bar{\varphi}, \bar{\psi})} \mathsf{E}_\mu\, \varrho(U_0, V_0) = D[\mathscr{C}(v)] .$$

Proof of the negative part. Since (U, X, Y, V) forms a Markov chain, $I(U, V) \leqq I(X, Y)$ by the information processing theorem [17]. Next, $I(X, Y) \leqq \mathscr{C}(v)$, by definition. Using the process definition of the DRF [14],

$$\mathsf{E}_\mu\, \varrho(U_0, V_0) \geqq \inf \mathsf{E}_p\, \varrho(U_0, V_0) = D[\mathscr{C}(v)]\,,$$

where the infimum is over all stationary and ergodic joint distributions p of (U, V) under which dist $(U) = \mu$ and $I_p(U, V) \leqq \mathscr{C}(v)$. □

Before turning to the proof of the positive part observe the following. If $\bar{\varphi}: A^\infty \to B^\infty$ is an infinite code, where B is a finite set, and if $[A, \mu, U]$ is a stationary source, then for any $\varepsilon > 0$ we can find a finite approximation $\bar{\varphi}^{(N)}$ to $\bar{\varphi}$ in the sense that

$$\text{Prob}\,[(\bar{\varphi}U)_0 \neq (\bar{\varphi}^{(N)}U)_0] \leqq \varepsilon$$

(see Thm. 3.1 of [2]). Since

$$\bar{d}(\bar{\varphi}U, \bar{\varphi}^{(N)}U) \leqq \text{Prob}\,[(\bar{\varphi}U)_0 \neq (\bar{\varphi}^{(N)}U)_0]\,,$$

and since $\bar{\varphi}U$, $\bar{\varphi}^{(N)}U$ each have the finite state space B, as in the proof of Theorem 1 we get from the above inequalities that

$$\bar{\varrho}(\bar{\varphi}U, \bar{\varphi}^{(N)}U) \leqq \varrho_M \varepsilon\,.$$

Thus, we get the following result:

Corollary 7. Let the hypotheses of Theorem 6 prevail. Then, if

$$\varrho(\bar{\varphi}^{(N)}, \bar{\psi}^{(M)}) = \mathsf{E}_\mu\, \varrho(U_0, [\bar{\psi}^{(M)}(\bar{\varphi}^{(N)}U)]_0)\,,$$

we have the formula

$$\inf_{N,M}\ \inf_{(\bar{\varphi}^{(N)}, \bar{\psi}^{(M)})} \varrho(\bar{\varphi}^{(N)}, \bar{\psi}^{(M)}) = D[\mathscr{C}(v)]\,.$$

Corollary 7 thus represents a generalization of Theorem 6.1 in [2] and of Theorem 2 in [3], the sliding-block information transmission theorems known up to now.

The idea of the proof of the positive part is motivated by the proof of Gray [2] for the case of noiseless channel. One first uses the sliding-block source coding theorem with a fidelity criterion [5] in order to achieve the goal of entropy compression to a level under the Shannon capacity of $[B, v, C]$. As shown in [5], proof of Theorem 2, this can be done using a finite reproduction alphabet, $\tilde{A}$ say. Since a finite coding of an ergodic source is again ergodic, we arrive at the situation of having a finite alphabet source $[\tilde{A}, \tilde{\mu}, \tilde{U}]$ such that $h(\tilde{U}) < \mathscr{C}(v)$ (care must be taken of the case $h(\tilde{U}) = \mathscr{C}(v)$). This makes it possible to use the zero-error transmission theorem of Kieffer [18] in order to achieve the goal of transmission of the compressed process $\tilde{U}$ over the channel. Since this second step is perfectly noiseless, all the distortion comes from the first step, which is easy to see giving the optimum distortion $D[\mathscr{C}(v)]$.

First let us recall two useful concepts from [4]. A source $[A, \mu]$ (A finite!) is said to be zero-error transmissible over a channel $[B, \nu, C]$ if there exists a stationary hookup (U, X, Y, V) (see Sect. 2) for which

$$\text{Prob}\,[U_0 \neq V_0] = 0 .$$

A source $[B, \lambda]$ is said to be ν-invulnerable (in symbols, $\lambda \in I(\nu)$; we shall use also the notation $X \in I(\nu)$ if $\lambda \in I(\nu)$, where $\lambda = \text{dist}(X)$), if there is a Markov chain (X, Y, V) and a stationary code $\xi: C^\infty \to B^\infty$ such that X, V each have the state space B, Y has the state space C, $\text{dist}(X) = \lambda$, $\text{dist}(X, Y) = \lambda\nu$, and $V = \xi Y$ satisfies

$$\text{Prob}\,[X_0 \neq V_0] = 0 .$$

As shown in [4], a source $[A, \mu]$ is zero-error transmissible over $[B, \nu, C]$ if and only if it is isomorphic to a ν-invulnerable source $[B, \lambda]$.

Proof of the positive part. Since $I(\nu)$ necessarily contains processes X with $h(X) < \mathscr{C}(\nu)$, but not those for which $h(X) = \mathscr{C}(\nu)$, we have to work with rates below capacity. Since $D(\cdot)$ is continuous [1], for any $\varepsilon > 0$ there is a $\delta > 0$ (we may assume $\delta \leq \frac{1}{4}\varepsilon$) so that $D(\mathscr{C}(\nu) - \delta) \leq D[\mathscr{C}(\nu)] + \frac{1}{4}\varepsilon$. Take $R = \mathscr{C}(\nu) - \delta < \mathscr{C}(\nu)$. By Theorem 2 of [5] we find a finite set $\tilde{A} \subset A$ and a sliding-block code $\bar{f}^{(N)}: A^\infty \to \tilde{A}^\infty$ such that (a)

(a) $h(\bar{f}^{(N)}U) \leq R$;

(b) $\varrho(\bar{f}^{(N)}) \leq D(R) + \frac{1}{4}\varepsilon \leq D[\mathscr{C}(\nu)] + \frac{1}{2}\varepsilon$.

Let $\tilde{U} = \bar{f}^{(N)}U$. Then $h(\tilde{U}) < \mathscr{C}(\nu)$, and $\tilde{U}$ is an ergodic process, so that $\tilde{U}$ is zero-error transmissible over $[B, \nu, C]$. Find $[B, \lambda] \in I(\nu)$ and a stationary invertible code $\bar{f}: \tilde{A}^\infty \to B^\infty$ such that $\bar{f}\tilde{U} = X$, $\text{dist}(X) = \lambda$. Since $\lambda \in I(\nu)$, we find a stationary code $\xi: C^\infty \to B^\infty$ such that

(c) $\text{Prob}\,[X_0 \neq (\xi Y)_0] = 0$,

where $\text{dist}(X, Y) = \lambda\nu$. Since $\bar{f}$ is an isomorphism, it follows from $X = \bar{f}[\bar{f}^{(N)}U]$ that $\bar{f}^{(N)}U = \bar{f}^{-1}X$. Hence

(d) $\text{Prob}\,[(\bar{f}^{(N)}U)_0 \neq (\bar{f}^{-1}X)_0] = 0.$

Now put $\bar{\varphi} = \bar{f} \circ \bar{f}^{(N)}$, $\bar{\psi} = \bar{f}^{-1} \circ \xi$, and let V denote the corresponding decoded process. Since $\bar{f}$ is an isomorphism, all the distortion in $\bar{\varphi}U$ comes from $\bar{f}^{(N)}$. Hence

(e) $\mathsf{E}_\mu\, \varrho(U_0, (\bar{f}^{(N)}U)_0) = \mathsf{E}_\mu\, \varrho(U_0, (\bar{\varphi}U)_0)$

and this, together with (b), shows that

(f) $\varrho(\bar{\varphi}) \leq D[\mathscr{C}(\nu)] + \frac{1}{2}\varepsilon < D[\mathscr{C}(\nu)] + \varepsilon$.

Since $V = \bar{\psi}Y = \bar{f}^{-1}(\xi Y)$, and since $\bar{f}^{-1}$ is an isomorphism, we get

$$\begin{aligned}\mathsf{E}_\mu\, \varrho((\bar{f}^{(N)}U)_0, V_0) &= \mathsf{E}_\mu\, \varrho((\bar{f}^{(N)}U)_0, (\bar{f}^{-1}\xi Y)_0) = \\ &= \mathsf{E}_\mu\, \varrho((\bar{f}[\bar{f}^{(N)}U])_0, (\xi Y)_0) = \mathsf{E}_\mu\, \varrho(X_0, (\xi Y)_0) \leq \\ &\leq \varrho_M\, \text{Prob}\,[X_0 \neq (\xi Y)_0] .\end{aligned}$$

Using (c) we get

(g) $\mathsf{E}_\mu\, \varrho((\bar{f}^{(N)}U)_0, V_0) = 0$.

Let

$$\bar{\varrho}(\bar{\varphi}, \bar{\psi}) = \mathsf{E}_\mu \, \varrho(U_0, [\bar{\psi}(\bar{\varphi}U)]_0) .$$

Then (e), (f), (g), and the triangle inequality together yield

$$\bar{\varrho}(\bar{\varphi}, \bar{\psi}) < D[\mathscr{C}(v)] + \varepsilon .$$

Since ε has been arbitrary, the proof is complete. □

If $[A, \mu, U]$ is a B-source with a finite alphabet A, then we may refer to Corollary 5 instead of the source coding theorem of [5]. This gives (b) and a strengthening of (a) in the above proof to the effect that not only (a) is true but we may find a B-process X and a stationary code $\bar{f}$: $A^\infty \to B^\infty$ such that $\bar{f}U = X$.

In conclusion, observe that the positive part of Theorem 6 can be formulated, with the aid of Corollary 5, as follows:

Corollary 8. Let $[A, \mu, U]$ be a stationary, ergodic, and aperiodic source, where A is finite. Let B, ϱ, and the channel $[B, \gamma, C]$ be specified as in Theorem 6. If $h(U) > \mathscr{C}(v)$ then for any $\varepsilon > 0$ there exists a B-process $X \in I(v)$ and a stationary code $\bar{f}$: $A^\infty \to B^\infty$ such that
(i) $\bar{f}U = X$, and
(ii) $\varrho_U(\bar{f}) \leqq D[\mathscr{C}(v)] + \varepsilon$.

Thus, if $h(U) < \mathscr{C}(v)$, then U is isomorphic with a process $X \in I(v)$, while if $h(U) > \mathscr{C}(v)$, then U has a v-invulnerable factor which is a B-process and the corresponding factor homomorphism gives distortion as close to the optimal one as we please.

ACKNOWLEDGMENT

The above simple proof of Theorem 1 was suggested to the author by John Kieffer.

(Received December 18, 1984.)

REFERENCES

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

[2] R. M. Gray: Sliding-block source coding. IEEE Trans. Inform. Theory *21* (1975), 357–368.

[3] R. M. Gray and D. S. Ornstein: Sliding-block joint source/noisy-channel coding theorems. IEEE Trans. Inform. Theory *22* (1976), 682–690.

[4] R. M. Gray, D. S. Ornstein, and R. L. Dobrushin: Block synchronization, sliding-block coding, invulnerable sources, and zero-error codes for discrete noisy channels. Ann. Probab. *8* (1980), 638–674.

[5] R. M. Gray, D. L. Neuhoff, and D. S. Ornstein: Non-block source coding with a fidelity criterion. Ann. Probab. *3* (1975), 478–491.

[6] Š. Šujan: Sinai's theorem and entropy compression. Problems Control Inform. Theory *12* (1983), 419 – 428.

[7] Š. Šujan: Ergodic theory, entropy, and coding problems of information theory. Kybernetika *19* (1983), supplement, 58 pp.

[8] J. C. Kieffer: A simple development of the Thouvenot relative isomorphism theory. Ann. Probab. *12* (1984), 204—211.
[9] R. M. Gray, D. L. Neuhoff, and P. C. Shields: A generalization of Ornstein's $\bar{d}$-distance with applications to information theory. Ann. Probab. *3* (1975), 315—328.
[10] D. S. Ornstein: Ergodic Theory, Randomness, and Dynamical Systems. Yale Univ. Press, New Haven, Conn. 1974.
[11] P. C. Shields: The Theory of Bernoulli Shifts. Univ. of Chicago Press, Chicago 1973.
[12] J. C. Kieffer: A method for proving multiterminal source coding theorems. IEEE Trans. Inform. Theory *27* (1981), 565—570.
[13] J. C. Kieffer: On the transmission of Bernoulli sources over stationary channels. Ann. Probab. *8* (1980), 942 — 961.
[14] R. M. Gray, D. L. Neuhoff, and J. K. Omura: Process definitions of distortion-rate functions and source coding theorem. IEEE Trans. Inform. Theory *21* (1975), 524—532.
[15] D. L. Neuhoff, R. M. Gray, and L. D. Davisson: Fixed rate universal block source coding with a fidelity criterion. IEEE Trans. Inform. Theory *21* (1975), 511—523.
[16] R. M. Gray and D. S. Ornstein: Block coding for discrete stationary $\bar{d}$-continuous noisy channels. IEEE Trans. Inform. Theory *25* (1979), 292—306.
[17] R. G. Gallager: Information Theory and Reliable Communication. J. Wiley, New York—London—Sydney—Toronto 1968.
[18] J. C. Kieffer: Stationary coding over stationary channels. Z. Wahrsch. verw. Gebiete *56* (1981), 113—126.

†*RNDr. Štefan Šujan, CSc., Joint Institute for Nuclear Research, Laboratory of Computing Techniques and Automation (LCTA), Head Post Office, P. O. Box 79, 101000 Moscow, U.S.S.R.* Permanent address: *Ústav merania a meracej techniky CEFV SAV (Institute of Measurement and Measuring Technique, Electro-Physical Research Centre, Slovak Academy of Sciences), 842 19 Bratislava, Czechoslovakia.*