NOT FOR QUOTATION
WITHOUT PERMISSION
OF THE AUTHOR

ON A DEFINITION OF THE DESCRIPTION COMPLEXITY OF FINITE SYSTEMS

Ashot Nersisian

January 1984
WP-84-4

INTERNATIONAL INSTITUTE FOR APPLIED SYSTEMS ANALYSIS
2361 Laxenburg, Austria

PREFACE

Most of the systems studied at IIASA are characterized by large numbers of components and by complex interactions. To minimize the problems caused by these factors it is often useful to find the most economical (minimal) representation of the system (or of the model used to describe it).

In this paper the author, a participant in the 1983 IIASA Young Scientists' Summer Program, investigates the problem of finding minimal descriptions of finite deterministic systems with a given accuracy. An algorithm for obtaining an asymptotically minimal description of discrete systems with constraints on the modeling accuracy is proposed.

ANDRZEJ WIERZBICKI
Chairman
System and Decision Sciences

ON A DEFINITION OF THE DESCRIPTION COMPLEXITY OF FINITE SYSTEMS

Ashot Nersisian

1. INTRODUCTION

Modern complex systems are characterized by large numbers of components and by complicated interactions. When studying such systems, therefore, it is useful to consider ways of finding the most economical representation of the system. The size of the minimal representation of a system is usually called its *description complexity*.

In practice *models* of complex systems (approximations of a given accuracy) are constructed, and these are then analyzed rather than the systems themselves. In this case the description complexity of a system should be interpreted as the complexity of its feasible approximation.

If methods for the construction of minimal descriptions are to be used in practice, they must be computationally efficient (i.e., they should have polynomial time complexity*).

* The time complexity of an algorithm is the number of computational steps necessary on any formal machine which carries out letter-by-letter operations.

For probabilistic systems the minimal description for a given fidelity criterion can be obtained by means of Shannon's rate-distortion theory (Shannon, 1959), which is described in more detail by Berger (1971). However, Shannon's results cannot be applied directly to practical problems because they involve a time complexity of 2^{2^n}, where n is the size of the system.

This paper investigates the problem of finding minimal descriptions of finite deterministic systems with a given accuracy. Deterministic methods of information compression, developed by Lupanov (1965), Nechiporuk (1965) and Sholomov (1967) for the synthesis of logical networks, are used. A particular case of this problem, which deals with the description complexity of partially-specified systems, has already been analyzed by Nersisian (1981).

An algorithm for obtaining an asymptotically minimal description of discrete systems with constraints on the modeling accuracy is proposed. The time complexity of this algorithm on any formal machine which carries out letter-by-letter operations (see Aho, Hopcroft and Ullman, 1976) does not exceed $n^{1+\gamma}$, where n characterizes the size of the system and $\gamma > 0$ is an arbitrary constant.

2. STATEMENT OF THE PROBLEM AND FORMULATION OF THE RESULTS

Let a system S be given by

$$S = \langle V, P \rangle , \tag{1}$$

where $V = \{1,\dots,v\}$ is a set of objects and $P = \{P_1,\dots,P_n\}$ a collection of relations given on set V.

Let a model M of the system be given by

$$M = < V,Q > ,$$

where $Q = \{Q_1,\dots,Q_g\}$ is a collection of relations "simpler" than P. It can easily be shown (see, for example, Nersisian, 1981), that a system such as (1) can be represented by a sequence of symbols $\tilde{x} = (x_1,\dots,x_n)$ in some finite alphabet $A = \{a_1,\dots,a_s\}$. Similarly, model M of system S can be represented by a sequence $\tilde{y} = (y_1,\dots,y_n)$ in some finite alphabet $B = \{b_1,\dots,b_t\}$. We shall call sequence $\tilde{y}$ an approximation (model) of $\tilde{x}$.

The accuracy of modeling is determined by several criteria $f_1,\dots,f_N$. Each criterion is characterized by a distortion measure $\rho^{(d)}(a_i,b_j) = \alpha_{ij}^{(d)}$ for each letter, where $0 \le \alpha_{ij}^{(d)} \le \infty$ is the "penalty" incurred by replacing symbol a_i by symbol b_j in the model. The overall distortion on going from sequence $\tilde{x}$ to sequence $\tilde{y}$ is the sum of the distortions for the individual letters:

$$\rho^{(d)}(\tilde{x},\tilde{y}) = \sum_{u=1}^{n} \rho^{(d)}(x_u,y_u) = \sum_{i,j} \alpha_{ij}^{(d)} w_{ij} ,$$

where w_{ij} is the number of positions u for which $x_u = a_i$, $y_u = b_j$. We shall consider the case in which the permitted level of relative distortion $\varepsilon^{(d)} \ge 0$, $d = 1,\dots,N$, for each criterion f_d is given, and approximation $\tilde{y} = \tilde{y}(\tilde{x})$ satisfies the constraints:

$$\rho^{(d)}(\tilde{x},\tilde{y}) \le \varepsilon^{(d)} n, \qquad d = 1,\dots,N .$$

This type of approximation will be called *$\tilde{\varepsilon}$-accurate*, where $\tilde{\varepsilon} = (\varepsilon^{(1)},\dots,\varepsilon^{(N)})$. It can easily be seen that to satisfy these conditions it is necessary that for each i there exists a $j = j(i)$ such that $\alpha_{ij}^{(d)} = 0$ for all d (otherwise for some d the distortion introduced by modeling will be greater than $\varepsilon^{(d)}n$). It should be noted that if $\alpha_{ij}^{(d)} = \infty$ then $w_{ij} = 0$ (i.e., no a_i is replaced by b_j).

Let $M_n(k_1,\ldots,k_s)$ denote the class of all sequences of length n in alphabet A such that each sequence contains k_i symbols a_i, $i = 1,\ldots,s$, $k_1 + \cdots + k_s = n$. Let parameters $k_1,\ldots,k_s$ be functions of n, i.e., $k_1 = k_1(n),\ldots,k_s = k_s(n)$. We shall assume that there exists a certain encoding technique $K = K_n$ which associates each sequence $\tilde{x} \in M_n(k_1,\ldots,k_s)$ with a binary sequence (code) $K(\tilde{x})$, and that there exists a decoding technique $D = D_n$ such that $D(K(\tilde{x}))$ is a word of length n in alphabet B. The word $D(K(\tilde{x}))$ will be considered to be the approximation of $\tilde{x}$. Let $l(\tilde{x})$ denote the length of codeword $K(\tilde{x})$. Let us assume that

$$l_n(k_1,\ldots,k_s) = \max_{\tilde{x} \in M_n(k_1,\ldots,k_s)} l(\tilde{x}) .$$

This quantity will be called the *description complexity* of class $M_n(k_1,\ldots,k_s)$.

The accuracy of the model will be characterized by the following quantities:

$$\rho^{(d)}(\tilde{x}) = \rho^{(d)}(\tilde{x}, D(K(\tilde{x}))), \quad d = 1,\ldots,N .$$

The above encoding–decoding method will be called *$\tilde{\varepsilon}$-accurate*, where $\tilde{\varepsilon} = (\varepsilon^{(1)},\ldots,\varepsilon^{(N)})$, $\varepsilon^{(1)} \geq 0,\ldots,\varepsilon^{(N)} \geq 0$, if for each $\tilde{x} \in M_n(k_1,\ldots,k_s)$

$$\rho^{(d)}(\tilde{x}) \leq \varepsilon^{(d)}n, \quad d = 1,\ldots,N .$$

We now define *$\tilde{\varepsilon}$-entropy* for the given $\tilde{\varepsilon} = (\varepsilon^{(1)},\ldots,\varepsilon^{(N)})$ and $P = (p_1,\ldots,p_s)$, $\sum_i p_i = 1$, $p_i \geq 0$, as*

* All logarithms are assumed to be to base two, i.e., $\log_2 x$.

$$H_{\tilde{\varepsilon}}(P) = \min_{||p_{ij}||} \sum_{i,j} p_{ij} \log \frac{p_{ij}}{p_i \sum_u p_{uj}}, \qquad (2)$$

where the minimum is taken over all $(s \times t)$ matrices $||p_{ij}||$ for which

$$\sum_j p_{ij} = p_i, \quad p_{ij} \geq 0, \quad \sum_{i,j} \alpha_{ij}^{(d)} p_{ij} \leq \varepsilon^{(d)} n, \qquad d = 1, \ldots, N.$$

Theorem 1

(1) Let the condition

$$\frac{H_{\tilde{\varepsilon}}\left(\frac{k_1}{n}, \ldots, \frac{k_s}{n}\right) \log n}{\log \log n} \to \infty$$

be satisfied.

Then for an arbitrary function $\alpha(n) \to \infty$ there exist $\tilde{\varepsilon}$-accurate encoding and decoding techniques which ensure that the description complexity satisfies the following relationship:

$$l_n(k_1, \ldots, k_s) \leq nH_{\tilde{\varepsilon}}\left(\frac{k_1}{n}, \ldots, \frac{k_s}{n}\right) + n\,\alpha(n) \frac{\log \log n}{\log n}. \qquad (3)$$

(2) If function $\alpha(n)$ is computable with time complexity n, then the encoding operation K_n and the decoding operation D_n have time complexities not greater than $n^{1+\gamma}$, where $\gamma > 0$ is an arbitrary constant.

(3) For each $\tilde{\varepsilon}$-accurate encoding and decoding technique*

$$l_n(k_1, \ldots, k_s) \geq nH_{\tilde{\varepsilon}}\left(\frac{k_1}{n}, \ldots, \frac{k_s}{n}\right) + c_1 \log n. \qquad (4)$$

* The letter c, possibly with a subscript, denotes a constant here and elsewhere.

3. PROOF OF THEOREM 1

1. Let $W = ||w_{ij}||$ be an $(s \times t)$ matrix satisfying the condition

$$\sum_j w_{ij} = k_i, \quad i = 1,\ldots,s .$$

The function

$$I(W) = n \log n - \sum_i k_i \log k_i - \sum_j m_j \log m_j - \sum_{i,j} w_{ij} \log w_{ij} \tag{5}$$

is then associated with matrix W, where

$$m_j = \sum_i w_{ij}, \quad j=1,\ldots,t .$$

It is easy to show that

$$I(W) = nI(||p_{ij}||) , \tag{6}$$

where

$$I(||p_{ij}||) = \sum_{i,j} p_{ij} \log \frac{p_{ij}}{\sum_u p_{uj} \sum_v p_{iv}} , \quad p_{ij} = \frac{w_{ij}}{n} .$$

Lemma 1

Given both an $(s \times t)$ matrix $W = ||w_{ij}||$ and an $(s \times t)$ matrix $W' = ||w'_{ij}||$ such that

$$|w_{ij} - w'_{ij}| \leq \Delta , \tag{7}$$

then*

$$|I(W) - I(W')| \lesssim \log n . \tag{8}$$

---

* The notation $\varphi_n \lesssim \psi_n$ or $\varphi_n = O(\psi_n)$ means that $\lim\limits_{n\to\infty} \frac{\varphi_n}{\psi_n} < \infty$, while the notation $\varphi_n = o(\psi_n)$ means that $\lim\limits_{n\to\infty} \frac{\varphi_n}{\psi_n} = 0$.

Proof

The inequality (8) follows immediately from expression (5), using condition (7) and the fact that s and t are limited.

2. Let*

$$\Delta = \left[\frac{n}{\log^2 n}\right]. \tag{9}$$

Consider all possible $(s \times t)$ matrices $W = ||w_{ij}||$ which satisfy the conditions

$$\sum_{i,j} a_{ij}^{(d)} w_{ij} \le \varepsilon^{(d)} n, \qquad d = 1,\dots,N. \tag{10}$$

For each $j \ne j(i)$ let

$$w_{ij} = l_{ij}\Delta, \tag{11}$$

where the l_{ij} are integers; for each $j = j(i)$ let

$$w_{ij(i)} = k_i - \sum_{j \ne j(i)} w_{ij}. \tag{12}$$

Now, from these, let us find the matrix W_0 with the smallest value of $I(W)$.

Lemma 2

(1) The following condition holds for matrix W_0

$$I(W_0) = nH_{\tilde{\varepsilon}}\left(\frac{k_1}{n},\dots,\frac{k_s}{n}\right) + O\left(\frac{n}{\log n}\right). \tag{13}$$

(2) The time complexity of the algorithm for finding matrix W_0 is limited by the value $n(\log n)^{2st+4}$.

---

* $[Z]$ $(]Z[)$ denotes the largest (smallest) integer which does not exceed Z.

Proof

Let the minimum in expression (2) be achieved on collection $\{p_{ij}\}$. Define

$$w_{ij}^{*} = p_{ij} n$$

and introduce matrix $W^{*} = ||w_{ij}^{*}||$. From (5) it follows that

$$nH_{\bar{\varepsilon}}(\frac{k_1}{n},\dots,\frac{k_s}{n}) = I(W^{*}) . \tag{14}$$

Let

$$l_{ij} = \left[\frac{w_{ij}^{*}}{\Delta}\right], \quad j \neq j(i) .$$

Then if w_{ij} is found from (9)–(12), the inequalities

$$|w_{ij} - w_{ij}^{*}| < \Delta, \quad j \neq j(i) ,$$
$$|w_{ij(i)} - w_{ij(i)}^{*}| < \Delta(s-1)$$

are satisfied, and from Lemma 1

$$|I(W) - I(W^{*})| \leq \Delta \log n . \tag{15}$$

Combining (7), (9), (14) and (15) leads to the expression

$$|I(W) - I(W')| \leq \Delta \log n . \tag{16}$$

Since the condition

$$w_{ij} \leq w_{ij}^{*} = p_{ij} n$$

holds for all $j \neq j(i)$ and condition (10) is valid for matrix W^{*}, we have

$$\sum_{i,j} \alpha_{ij}^{(d)} w_{ij} = \sum_{i,j \neq j(i)} w_{ij}\, \alpha_{ij}^{(d)} \leq \sum_{i,j \neq j(i)} w_{ij}^{*}\, \alpha_{ij}^{(d)} = \sum_{i,j} \alpha_{ij}^{(d)} w_{ij}^{*} \leq \varepsilon^{(d)} n ,$$

and therefore matrix W also satisfies condition (10). Inequality (13) follows from (16) and from the relation

$$I(W_0) \leq I(W) .$$

Let us now evaluate the complexity of finding matrix W_0. Instead of function $I(W)$ we can consider the monotonically related function

$$2^{I(W)} = \frac{n^n \prod_{i,j} w_{ij}^{w_{ij}}}{\prod_i k_i^{k_i} \prod_j n_j^{n_j}} . \tag{17}$$

This expression contains numbers of the form s^s, $s \leq n$, whose dimensions (numbers of binary digits) do not exceed $n \log n$. To compute s^s requires only $\log s \leq \log n$ multiplications (see, for example, Valski, 1959). According to Schönhage and Strassen (1971), not more than $n \log^3 n$ elementary operations are needed to multiply together two $(n \log n)$-digit numbers and the general number of operations required to compute s^s, $s \leq n$, is of the order of $n \log^4 n$, or less. After finding all the numbers of the form s^s involved in (17) a finite number of multiplication and division operations must be performed. This requires of the order of $n \log^3 n$ operations or less. To verify condition (10) requires no more than $\log n$ operations. Since there are at most $((\log n)^2 + 1)^{st}$ versions of $(s \times t)$ matrix W, the overall time complexity of finding matrix W_0 is of the order of $n(\log n)^{2st+4}$ or less.

3. Let us introduce quantities

$$q^0(j/i) = \frac{w_{ij}^0}{k_i}, \quad i = 1, \ldots, s, \quad j = 1, \ldots, t \tag{18}$$

and construct the $(s \times t)$ matrix $Q^0 = ||q^0(j/i)||$. Assume that an integer $\nu \leq n$ is given. We shall consider an arbitrary collection $\chi_1, \ldots, \chi_s$ such that $\sum_i \chi_i = \nu$. Let $M_\nu(\chi_1, \ldots, \chi_s)$ be the class of all sequences containing χ_i symbols a_i. We shall use the following notation:

$$\omega_{ij} = [\chi_i q^0(j/i)], \quad j \neq j(i), \tag{19}$$
$$\omega_{ij(i)} = \chi_i - \sum_{j \neq j(i)} \omega_{ij},$$
$$\sum_i \omega_{ij} = \mu_j .$$

Next let us form an $(s \times t)$ matrix $\Omega = ||\omega_{ij}||$. Since

$$|\omega_{ij} - \chi_i q^0(j/i)| < c_2 ,$$

we deduce from Lemma 1 that

$$I(\Omega) - \nu I(\frac{\chi_1}{\nu},\ldots,\frac{\chi_s}{\nu};\ Q^0) \leq \log \nu , \tag{20}$$

where

$$I(\frac{\chi_1}{\nu},\ldots,\frac{\chi_s}{\nu};\ Q^0) = \sum_{i,j} \frac{\chi_i}{\nu} q^0(j/i) \log \frac{\frac{\chi_i}{\nu} q^0(j/i)}{\frac{\chi_i}{\nu} \sum_i \frac{\chi_i}{\nu} q^0(j/i)} .$$

It is also evident that

$$\sum_{i,j} \alpha_{ij}^{(d)} \omega_{ij} \leq \sum_{i,j} \alpha_{ij}^{(d)} \chi_i q^0(j/i), \quad d = 1,\ldots,N . \tag{21}$$

We shall say that a set N of sequences in alphabet B $\tilde{\varepsilon}$-accurately approximates a set M of sequences in alphabet A if for each sequence from M there is a corresponding $\tilde{\varepsilon}$-accurate approximation in the N. Let us denote by $T_{\tilde{\varepsilon}}(\chi_1,\ldots,\chi_s)$ the minimum cardinality of a set which $\tilde{\varepsilon}$-accurately approximates $M_\nu(\chi_1,\ldots,\chi_s)$.

Lemma 3

(1) The following relation holds:

$$\log T_{\tilde{\varepsilon}}(\chi_1,\ldots,\chi_s) \leq \nu I(\frac{\chi_1}{\nu},\ldots,\frac{\chi_s}{\nu};\ Q^0) + O(\log \nu) , \tag{22}$$

where $\tilde{\delta} = (\delta^{(1)},\ldots,\delta^{(N)})$ and

$$\delta^{(d)} = \sum_{i,j} \alpha_{ij}^{(d)} \omega_{ij}, \quad d = 1,\ldots,N. \tag{23}$$

(2) A set N which $\tilde{\varepsilon}$-accurately approximates $M_\nu(\chi_1,\ldots,\chi_s)$ and satisfies estimate (22) can be found with time complexity not greater than $c_3^\nu + c_4 \log^2 n$.

Proof

Estimate (22) can be obtained using the gradient procedure suggested by Sholomov (1967); this is a modification of the procedure proposed by Nechiporuk (1965). A table is formed with $\frac{\nu!}{\prod_i \chi_i!} \leq s^\nu$ columns corresponding to the sequences $\tilde{\zeta} \in M_\nu(\chi_1,\ldots,\chi_s)$ and with $\frac{\nu!}{\prod_j \mu_j!} \leq t^\nu$ rows corresponding to the sequences $\tilde{\eta}$ containing μ_j symbols b_j. At the intersection of row $\tilde{\eta}$ and column $\tilde{\zeta}$ we put a "1" if we have ω_{ij} $(i = 1,\ldots,s\ ,\ j = 1,\ldots,t)$ positions with symbol a_i in sequence $\tilde{\zeta}$ and symbol b_j in sequence $\tilde{\eta}$; otherwise we insert a "0". It is evident that if a "1" is found at the intersection of row $\tilde{\eta}$ and column $\tilde{\zeta}$ then $\tilde{\eta}$ is a $\tilde{\delta}$-accurate approximation of $\tilde{\zeta}$. The gradient procedure is then used. At each step of this procedure it is necessary to find the row with the maximum number of ones in the current table, and delete this row and the columns containing the ones. The procedure terminates when all the columns have been deleted, and the desired set N is formed by the sequences corresponding to the deleted rows. Calculations similar to those described by Sholomov (1967) give the estimate

$$\log T_{\tilde{\delta}}(\chi_1,\ldots,\chi_s) \le I(\Omega) + O(\log \nu) ,$$

which, taken together with (20), yields (22).

The upper bound of the size of the table is $(s,t)^{\nu}$. The time complexity of constructing the table and performing the gradient procedure cannot exceed the polynomial of the table size, and is limited by the value of c_3^{ν}. The complexity of finding parameters $\omega_{ij} = \left[\frac{\chi_i w_{ij}}{k_i}\right]$ and $\mu_j = \sum_j \omega_{ij}$ does not exceed the value of $c_4 \log^2 n$. The resulting total complexity is

$$c_3^{\nu} + c_4 \log^2 n .$$

4. Let there be a sequence $\tilde{x} \in M_n(k_1,\ldots,k_s)$ in alphabet A. Assume that a natural parameter ν is given and that the sequence $\tilde{x}$ is broken into pieces of length ν:

$$\tilde{x} = \tilde{\zeta}_1 \cdots \tilde{\zeta}_M, \quad M = \left]\frac{n}{\nu}\right[.$$

Let $\tilde{\zeta}_u$ belong to class $M_{\nu}(\chi_1^{(u)},\ldots,\chi_s^{(u)})$. We shall denote by $\tilde{\eta}_u$ any $\tilde{\delta}$-accurate approximation of $\tilde{\zeta}_u$, where $\tilde{\delta}$ is determined using (19) and (23).

Lemma 4

The sequence

$$\tilde{\eta} = \tilde{\eta}_1,\ldots,\tilde{\eta}_M \tag{24}$$

is an $\tilde{\varepsilon}$-accurate approximation of sequence $\tilde{x}$.

Proof

The overall distortion between the sequences $\tilde{x}$ and $\tilde{y}$ is equal to the sum of the distortions between the pieces $\tilde{\zeta}_u$ and their $\tilde{\delta}$-accurate approximations $\tilde{\eta}_u$:

$$\rho^{(d)}(\tilde{x},\tilde{y}) = \sum_{u=1}^{M} \rho^{(d)}(\tilde{\zeta}_u,\tilde{\eta}_u) = \sum_{u,i,j} \alpha_{ij}^{(d)}\omega_{ij}^{(u)} . \tag{25}$$

Substituting in (25) the value of $\omega_{ij}^{(u)}$ from (19) we obtain

$$\rho^{(d)}(\tilde{x},\tilde{y}) \le \sum_{u,i,j\ne j(i)} \alpha_{ij}^{(d)}\chi_i q^0(j/i) = \sum_{i,j\ne j(i)} \alpha_{ij}^{(d)} q^0(j/i)\sum_u \chi_i^{(u)} .$$

From the obvious equality $\sum_{u=1}^{M} \chi_i^{(u)} = k_i$ and relations (10) and (18), we find that

$$\rho^{(d)}(\tilde{x},\tilde{y}) \le \sum_{i,j} \alpha_{ij}^{(d)} w_{ij}^0 \le \varepsilon^{(d)} n ,$$

which proves the Lemma.

5. We shall now describe the encoding procedure.

The codeword $K(\tilde{x})$ for a sequence $\tilde{x} \in M_n(k_1,\dots,k_s)$ consists of three parts:

$$K(\tilde{x}) = \Lambda\, \Sigma\, \Xi .$$

These are known as the reference, main and auxiliary parts, respectively. We shall consider first the main part Σ of the codeword $K(\tilde{x})$. Sequence $\tilde{x} \in M_n(k_1,\dots,k_s)$ is assumed to have been broken down into pieces of length n. We group the pieces with the same parameters $\chi_1,\dots,\chi_s$, $\chi_1+\cdots+\chi_s = \nu$ into separate classes labelled $M_1,\dots,M_R$. For each class M_i the δ-accurate approximations (where δ is determined from (23)) are found with the help of the gradient procedure described earlier (see the proof of Lemma 3). We shall denote the number of approximations by $\hat{T}_\delta(\chi_1,\dots,\chi_s)$. Now arrange all these δ-approximations $\tilde{\eta}$ for sequences from $M_\nu(\chi_1,\dots,\chi_s)$ in a certain (e.g., lexicographical) order and number them using binary sequences $\tilde{\pi}(\tilde{\eta})$ of length

$$\tau_\nu =]\log T_\nu[\ . \tag{26}$$

Let $\tilde{i} = (i_1,\dots,i_p)$ denote the binary representation of number i and $\tilde{i}^*$ be the corresponding binary sequence, where

$$\tilde{i}^* = (i_1 i_1,\dots,i_p i_p 01) \ .$$

It is obvious that numbers i and j can be found uniquely from sequence $\tilde{i}^*\tilde{j}$. The code $\tilde{\sigma}(\tilde{\zeta}_i)$ for the piece $\tilde{\zeta}_i \in M_v = M^{(\nu)}(\chi_1,\dots,\chi_s)$ will then be of the form

$$\tilde{\sigma}(\tilde{\zeta}_i) = \tilde{v}^*\tilde{\pi}(\tilde{\eta}_j) \ , \tag{27}$$

where $\tilde{\eta}_j$ denotes a $\tilde{\delta}$-accurate approximation of $\tilde{\zeta}_i$. A sequence of codes of adjacent pieces $\tilde{\zeta}_i$ constitutes the main part of the codeword

$$\Sigma = \tilde{\sigma}(\tilde{\zeta}_1) \ \cdots \ \tilde{\sigma}(\tilde{\zeta}_M) \ .$$

The auxiliary part of the codeword is a list of $\tilde{\delta}$-approximations. Let the $\tilde{\delta}$-approximations for the class $M_v = M_\nu(\chi_1,\dots,\chi_s)$ be

$$\tilde{\eta}_{v1},\dots,\tilde{\eta}_{v\hat{T}_v} \ ,$$

where

$$\hat{T}_v = \hat{T}_{\tilde{\delta}}(\chi_1,\dots,\chi_s) \ .$$

Assume that

$$\tilde{\xi}(M_v) = \hat{\eta}_{v1},\dots,\hat{\eta}_{v\hat{T}_v} \ ,$$

where $\hat{\eta}_{v\hat{T}_v}$ is obtained from $\tilde{\eta}_{v\hat{T}_v}$ by replacing each symbol $b_j \in B$ by a binary representation $\tilde{j}$ of length $]\log t[$.

The auxiliary part Ξ will be of the form

$$\Xi = \tilde{\xi}(M_1) \cdots \tilde{\xi}(M_R) ,$$

where R is the number of classes $M_\nu(\chi_1,\ldots,\chi_s)$. We shall denote the length of $\tilde{\xi}(M_i)$ by J_i and the length of the auxiliary part Ξ by L_Ξ.

The reference part Λ contains the numerical parameters required to decode the main and auxiliary parts. It is of the form

$$\Lambda = \tilde{n}^* \tilde{k}_1^* \cdots \tilde{k}_s^* \tilde{L}_\Sigma^* \tilde{L}_\Xi^* \tilde{\tau}_1^* \cdots \tilde{\tau}_R^* \tilde{J}_1^* \cdots \tilde{J}_R^* .$$

Let the length of the reference part be L_Λ. It is evident that the codeword $K(\tilde{x})$ can be decoded into $\tilde{\delta}$-accurate approximations $\tilde{\eta}_u$ of pieces $\tilde{\zeta}_u$, which since they are adjacent (according to Lemma 4) give an $\tilde{\varepsilon}$-accurate approximation $\tilde{y}$ of the sequence $\tilde{x}$.

8. The length $L(K(\tilde{x}))$ of codeword $K(\tilde{x})$ is given by:

$$L(K(\tilde{x})) = L_\Lambda + L_\Sigma + L_\Xi .$$

Let us estimate the length L_Σ of the main part Σ of the codeword. This is

$$L_\Sigma = \sum_{i=1}^{M} l(\tilde{\sigma}(\tilde{\zeta}_i)) , \tag{28}$$

where $l(\tilde{\sigma}(\tilde{\zeta}_i))$ is the length of the piece of code $\tilde{\sigma}(\tilde{\zeta}_i)$. From (26) and (27) it follows that

$$l(\tilde{\sigma}(\tilde{\zeta}_i)) \leq 2 \,]\log R\,[\, + \,]\log \hat{T}_{\mathfrak{z}}(\chi_1,\ldots,\chi_s)\,[\, + 2 . \tag{29}$$

Since the number R of classes cannot exceed $(\nu+1)^s$, we may write (taking into account Lemma 3)

$$l(\tilde{\sigma}(\tilde{\zeta})) \leq \nu I(\frac{\chi_1}{\nu},\ldots,\frac{\chi_s}{\nu};\, Q^0) + O(\log \nu) . \tag{30}$$

One more Lemma must be proved before we can estimate the value of L_Σ. For $\tilde{\zeta} \in M_\nu(\chi_1,\ldots,\chi_s)$ we introduce the notation

$$\varphi(\tilde{\zeta}) = \nu I\left(\frac{\chi_1}{\nu},\ldots,\frac{\chi_s}{\nu};\ Q^0\right). \tag{31}$$

Lemma 5

$$\varphi(\tilde{\zeta}_1 \cdots \tilde{\zeta}_g) \geq \sum_{u=1}^{g} \varphi(\tilde{\zeta}_u). \tag{32}$$

Proof

It is sufficient to consider the case $g = 2$ (as will be seen later). Let $\tilde{\zeta}_i \in M_\nu(\chi_1^{(i)},\ldots,\chi_s^{(i)})$, $i=1,\ldots,M$. Using the upward convexity property of the average mutual information function, which holds when the collection Q^0 of transitional probabilities is fixed (see Gallager, 1968), we can write

$$\frac{\nu_1}{\nu_1+\nu_2} I\left[\frac{\chi_1^{(1)}}{\nu_1},\ldots,\frac{\chi_s^{(1)}}{\nu_1};\ Q^0\right] + \frac{\nu_2}{\nu_1+\nu_2} I\left[\frac{\chi_1^{(2)}}{\nu_2},\ldots,\frac{\chi_s^{(2)}}{\nu_2};\ Q^0\right] \leq$$

$$\leq I\left[\frac{\chi_1^{(1)}+\chi_1^{(2)}}{\nu_1+\nu_2},\ldots,\frac{\chi_s^{(1)}+\chi_s^{(2)}}{\nu_1+\nu_2};\ Q^0\right].$$

Multiplying both parts of the inequality by $\nu_1+\nu_2$ and making use of (31) and $\tilde{\zeta} = \tilde{\zeta}_1\tilde{\zeta}_2 \in M_{\nu_1+\nu_2}(\chi_1^{(1)}+\chi_1^{(2)},\ldots,\chi_s^{(1)}+\chi_s^{(2)})$, we obtain inequality (32), which proves the Lemma. Substituting (29) into (28), using Lemmas 2 and 5 and the relation $I(W_0) = nI\left[\frac{k_1}{n},\ldots,\frac{k_s}{n};\ Q^0\right]$, we obtain

$$L_\Sigma \leq nI\left[\frac{k_1}{n},\ldots,\frac{k_s}{n};\ Q^0\right] + O(n \log \nu) \leq$$

$$\leq nH_{\tilde{\varepsilon}}\left(\frac{k_1}{n},\ldots,\frac{k_s}{n}\right)+O\left(\frac{n}{\log n}\right)+O\left(\frac{n\log\nu}{\nu}\right).$$

Let us now estimate the length of the auxiliary part of the codeword. The binary length of one approximation $\hat{\eta}_u$ is not more than $\nu]\log t[$. From the fact that the number of $\tilde{\delta}$-approximations for one class does not exceed s^ν and the number of classes is not more than $(\nu+1)^s$, we arrive at the estimate

$$L_{\tilde{\Xi}}\leq O((\nu+1)^s\nu s^\nu)\leq c_5^\nu .$$

The value of L_Λ is the sum of the lengths of the $\tilde{\iota}^*$ form representations. Since the length of each parameter does not exceed n, it is not difficult to see that

$$L_\Lambda\leq R\log n\leq(\nu+1)^s\log n\leq\nu^{c_6}\log n .$$

Finally we have

$$L(K(\tilde{x}))\leq nH_{\tilde{\varepsilon}}\left(\frac{k_1}{n},\ldots,\frac{k_s}{n}\right)+O\left(\frac{n}{\log n}\right)+ \tag{33}$$

$$+\,O\left(\frac{n\log\nu}{\nu}\right)+c_5^\nu+\nu^{c_6}\log n .$$

9. Let $\alpha=\alpha(n)\to\infty$ be an arbitrary function which satisfies the condition

$$\frac{\alpha n\log n}{\log\log n}\to\infty ,$$

and assume that

$$\nu=\left[\frac{\log n}{\alpha}\right].$$

It can be shown directly that substitution of this value into (33) yields the estimate (3).

10. Computing experience shows that the time complexity of encoding and decoding techniques satisfies the estimate given in point (2) of Theorem 1.

11. Using the technique described in Sholomov (1967), the following inequality can be obtained:

$$\log T_{\tilde{\varepsilon}}(k_1,\ldots,k_s) \geq nH_{\tilde{\varepsilon}}\left[\frac{k_1}{n},\ldots,\frac{k_s}{n}\right] + c_1 \log n \ ,$$

from which, by means of "power" considerations, we arrive at estimate (4).

4. CONCLUDING REMARKS

1. The encoding technique proposed in the present paper may be used to obtain a proof of Shannon's Theorem of encoding of discrete sources with a fidelity criterion without resorting to the random encoding technique.

2. The proposed encoding technique and codeword construction can be used to compress large arrays of information if a certain distortion of the initial array is allowed in decoding. A data-compression algorithm for the case $A = \{0,1,*\}$, $B = \{0,1,\}$, $\alpha_{ij} \in \{0,\infty\}$ was constructed at IIASA by the author. Here * represents an unspecified symbol which can be replaced arbitrarily by 0 or 1. This algorithm has been implemented on the VAX computer at IIASA by Z. Fortuna.

ACKNOWLEDGMENTS

The author would like to express profound gratitude to Zenon Fortuna and Alexander Sarkisov for their useful suggestions, and to thank Helen Gasking for editorial support, and Nora Avedisian for preparing this paper on IIASA's computerized text-processing system.

REFERENCES

Aho, A., Y. Hopcroft, and U. Ullman (1976). *The Design and Analysis of Computer Algorithms*. Addison Wesley, Massachusetts.

Berger, T. (1971). *Rate Distortion Theory: A Mathematical Basis for Data Compression*. Prentice-Hall, Englewood Cliffs, New Jersey.

Gallager, R.G. (1968). *Information Theory and Reliable Communication.* John Wiley and Sons, New York and London.

Lupanov, O.B. (1965). On one approach to the synthesis of control systems: the principle of local encoding. *Problemi Kibernetiki*, Vol. 14, Moscow, Nauka. (In Russian).

Nechiporuk, E.I. (1965). On the complexity of valve schemes that realize Boolean matrices with undefined elements. *Dokladi Akademii Nauk USSR*, Vol. 163. (In Russian).

Nersisian, A.L. (1981). Asymptotically optimal representation of partially specified systems. In *Metodi Issledovania Slojnikh Sistem*, All-Union Institute of Systems Studies (VNIISI), Moscow. (In Russian).

Schönhage, A. and V. Strassen (1971). Schnelle Multiplikation grosser Zahlen. *Computing*, Vol. 7, No. 3-4.

Shannon, C.E. (1959). Coding theorems for a discrete source with fidelity criterion. *IRE Nat. Con. Record.*

Sholomov, L.A. (1967). On functionals that characterize the complexity of systems of incompletely specified Boolean functions. *Problemi Kibernetiki*, Vol. 19, Moscow, Nauka. (In Russian).

Valski, P.E. (1959). On the minimum number of multiplications necessary to raise to a given power. *Problemi Kibernetiki*, Vol. 2, Moscow, Nauka. (In Russian).