Hence, $s = |J| = 1$ and

$$\begin{aligned} m_J &= (f_1^{(1)})^{-1}(\{1\}) = 1 \\ m_a &= (f_2^{(1)})^{-1}((11)) = 2 \\ m_b &= (f_3^{(1)})^{-1}((10)) = 2 \\ m' &= 4 + (1-1)\cdot 2 \cdot 2 + (2-1)\cdot 2 + (2-1) + 1 = 8 \end{aligned}$$

where (14) and (15) were used.

References

[1] R. Ahlswede, "Multi-way communication channels," in *2nd Int. Symp. Information Theory* (Tsahkadzor, Armenian SSR, 1971). Budapest, Hungary: Publishing House of the Hungarian Academy of Sciences, 1973, pp. 23–52.

[2] ——, "The capacity region of a channel with two senders and two receivers," *Ann. Probab*., vol. 2, no. 5, pp. 805–814, 1974.

[3] T. Kasami, S. Lin, V. K. Wei, and S. Yamamura, "Graph theoretic approaches to the code construction for the two-user multiple-access binary adder channel," *IEEE Trans. Inform. Theory*, vol. IT-29, pp. 114–130, Jan. 1983.

[4] E. J. Weldon, "Coding for a multiple-access channel," *Inform. Contr.*, vol. 36, pp. 256–274, Mar. 1978.

[5] G. G. Khachatrian, "On the construction of codes for noiseless synchronized 2-user channel," *Probl. Contr. Inform. Theory*, vol. 11, no. 4, pp. 319–324, 1982.

[6] P. A. B. M. Coebergh van den Braak and H. C. A. van Tilborg, "A family of good uniquely decodable code pairs for the two-access binary adder channel," *IEEE Trans. Inform. Theory*, vol. IT-31, pp. 3–9, Jan. 1985.

[7] V. F. Babkin, "A universal encoding method with nonexponential work expediture for a source of independent messages," *Probl. Pered. Inform.*, vol. 7, no. 4, pp. 13–21, Oct.–Dec. 1971. English translation: *Probl. Inform. Transm.*, pp. 288–294.

[8] J. P. M. Schalkwijk, "An algorithm for source coding," *IEEE Trans. Inform. Theory*, vol. IT-18, pp. 395–399, May 1972.

[9] T. M. Cover, "Enumerative source coding," *IEEE Trans. Inform. Theory*, vol. IT-19, pp. 73–77, Jan. 1973.

Hierarchical Guessing with a Fidelity Criterion

Neri Merhav, *Senior Member, IEEE*, Ron M. Roth, *Senior Member, IEEE*, and Erdal Arikan, *Senior Member, IEEE*

***Abstract*— In an earlier paper, we studied the problem of guessing a random vector $\boldsymbol{X}$ within distortion D, and characterized the best attainable exponent $E(D, \rho)$ of the ρth moment of the number of required guesses $G(X)$ until the guessing error falls below D. In this correspondence, we extend these results to a multistage, hierarchical guessing model, which allows for a faster search for a codeword vector at the encoder of a rate-distortion codebook. In the two-stage case of this model, if the target distortion level is D_2, the guesser first makes guesses with respect to (a higher) distortion level D_1, and then, upon his/her first success, directs the subsequent guesses to distortion D_2. As in the above-mentioned earlier paper, we provide a single-letter characterization of the best attainable guessing exponent, which relies heavily on well-known results on the successive refinement problem. We also relate this guessing exponent function to the source-coding error exponent function of the two-step coding process.**

***Index Terms*— Guessing, rate-distortion theory, source-coding error exponent, successive refinement.**

I. Introduction

In [1], we studied the basic problem of guessing a random vector with respect to (w.r.t.) a fidelity criterion. In particular, for a given information source, a distortion measure d, and distortion level D, this problem is defined as follows. The source generates a sample vector $\boldsymbol{x} = (x_1, \cdots, x_N)$ of a random N-vector $\boldsymbol{X} = (X_1, \cdots, X_N)$. Then, the guesser, who does not have access to $\boldsymbol{x}$, provides a sequence of N-vectors (guesses) $\boldsymbol{y}_1, \boldsymbol{y}_2, \cdots$ until the first success of guessing $\boldsymbol{x}$ within per-letter distortion D, namely, $d(\boldsymbol{x}, \boldsymbol{y}_i) \le ND$ for some positive integer i. Clearly, for a given list of guesses, this number of guesses i is solely a function of $\boldsymbol{x}$, denoted by $G_N(\boldsymbol{x})$. The objective of [1] was to characterize the best achievable asymptotic performance and to devise good guessing strategies in the sense of minimizing moments of $G_N(\boldsymbol{X})$. It has been shown in [1], that for a finite-alphabet, memoryless source P and an additive distortion measure d, the smallest attainable asymptotic exponential growth rate of $\boldsymbol{E}\{G_N(\boldsymbol{X})^\rho\}$ $(\rho > 0)$ with N, is given by

$$E(D, \rho) = \max_{P'} [\rho R(D, P') - \mathcal{D}(P' \| P)] \tag{1}$$

where the maximum w.r.t. P' is over the set of all memoryless sources with the same alphabet as P, $R(D, P')$ is the rate-distortion function of P' w.r.t. distortion measure d at level D, and $\mathcal{D}(P' \| P)$ is the relative entropy, or the Kullback–Leibler information divergence, between P' and P, i.e., the expectation of $\ln [P'(X)/P(X)]$ w.r.t. P'.

Manuscript received December 1, 1996. The work of N. Merhav was supported in part by the Israel Science Foundation founded by the Israel Academy of Sciences and Humanities.

N. Merhav was with Hewlett-Packard Laboratories, Palo Alto, CA 94304 USA. He is now with the Department of Electrical Engineering, Technion–Israel Institute of Technology, Haifa 32000, Israel (e-mail: merhav@ee.technion.ac.il).

R. M. Roth was with Hewlett-Packard Laboratories, Palo Alto, CA 94304 USA. He is now with the Department of Computer Science, Technion–Israel Institute of Technology, Haifa 32000, Israel (e-mail: ronny@cs.technion.ac.il).

E. Arikan is with the Electrical-Electronics Engineering Department, Bilkent University, 06533 Ankara, Turkey (e-mail: arikan@ee.bilkent.edu.tr).

Communicated by R. Laroia, Associate Editor for Source Coding.

Publisher Item Identifier S 0018-9448(99)00067-X.

One of the motivations of the guessing problem in its above described basic form, is that a good guessing strategy tells us how to order the codebook vectors of a rate-distortion block encoder, so as to minimize the typical search effort until a satisfactory codeword is found. As explained in [1], however, the guessing performance is an indication to the search complexity only under a very simple search model where the codewords are scanned in a fixed order, without taking advantage of the full information available from earlier unsuccessful search trials or guesses.

In this correspondence, we take one step towards the improvement of this search model. This is done by examining families of guessing strategies that are induced by hierarchical, multistage codebook structures, in particular, successive refinement codes (see, e.g., [2]–[7]). From the rate-distortion coding point of view, these structures are motivated by progressive transmission applications since they allow for simultaneous operation at more than one point in the rate-distortion plane, sometimes without loss of rate-distortion optimality at either point. From the searching, or guessing aspects considered here, these structures are attractive because they provide considerably more efficient and faster search for the first codeword that satisfies the distortion constraint w.r.t. a given source vector. In the two-stage case of the successive refinement structure, in order to encode a source vector $\boldsymbol{X}$ within a given target per-letter distortion level D_2, one first seeks, in a first-layer codebook, the first codeword $\boldsymbol{y}_i$ within distance ND_1 from $\boldsymbol{x}$ (which is a relatively fast search), and then seeks the first codeword $\boldsymbol{z}_{ij}$ at the target distance ND_2 from $\boldsymbol{x}$ along a second-layer codebook that corresponds to $\boldsymbol{y}_i$. As a simple example, if the first-layer code operates at rate-$R/2$ and each second-layer code is at rate-$R/2$, then the total rate is R but the number of guesses, or search trials grows exponentially as $2^{NR/2}$, and not 2^{NR} which would be the case if the code had only one stage.

Analogously to [1], our main result in this correspondence, is in characterizing the best attainable two-stage guessing exponent for memoryless sources, additive distortion measures, and two given distortion levels. We first derive a lower bound $E_2(D_1, D_2, \rho)$ on the exponent of the ρth-order moment of the guessing effort associated with the intermediate distortion level D_1 and the target distortion level D_2. Clearly, if only the target distortion level D_2 is specified, it would be natural to select D_1 so as to minimize $E_2(D_1, D_2, \rho)$. We are able to demonstrate the achievability of $E_2(D_1, D_2, \rho)$ under the assumption that the guesser knows in advance the *type class*, or equivalently, the empirical probability mass function (PMF) of the given source vector $\boldsymbol{x}$. There are several justifications for this assumption. First, in source-coding applications, which serve as the main motivation for the two-stage guessing problem, it is conceivable that the empirical PMF information is easily accessible to the guesser (or the encoder). Secondly, similarly as in the single-stage case, the validity of $E_2(D_1, D_2, \rho)$ as a lower bound is unaffected by knowledge of the type class. For the same reason, this setting still serves as an extension of [1]. Finally, and perhaps most importantly, under this assumption, the guesser has the flexibility to choose the first-layer distortion level D_1 depending on the empirical PMF. This, in general, gives better guessing performance than if D_1 were fixed. We also show that the successively refinable case gives the best possible guessing exponent, which can be easily expressed in terms of the single-stage guessing exponent $E_1(D_2, \cdot)$. The achievability of $E_2(D_1, D_2, \rho)$ without knowing the empirical PMF, however, remains an open problem, and we shall elaborate on this later on.

Another aspect of the guessing exponent is its interesting relation to the source-coding exponent. In the single-stage setting, the source-coding error exponent $F_1(R, D)$, is defined as the best exponential decay rate of the probability of failing to encode a source vector $\boldsymbol{X}$ with a rate-R codebook at distortion D. In [1], it has been shown that the guessing exponent $E_1(D, \rho)$ as a function of ρ, and the source-coding error exponent $F_1(R, D)$ as a function of R, are a Fenchel–Legendre transform (FLT) pair. We show that this result extends to the two-stage case merely in a partial manner: The two-stage guessing exponent is lower-bounded by the FLT of the two-stage error exponent and *vice versa*.

Finally, a general comment is in order: although we confine our attention, in this correspondence, to strategies with two levels of guessing lists, it should be understood that the results extend fairly easily to any fixed and finite number of levels, while the concept remains the same. Our exposition is limited to the two-level case for reasons of simplicity.

The outline of this correspondence is as follows. In Section II, we define notation conventions and provide some background on the problem of interest. Section III is devoted to the lower bound on the guessing exponent. In Section IV, we discuss the conditions for the achievability of the lower bound. In Section V, we focus on the successively refinable case. Section VI discusses the relation to the two-step source-coding error exponent. Finally, Section VII concludes the correspondence.

II. Notation, Problem Description, and Preliminaries

Consider a memoryless information source P emitting symbols from a finite alphabet $\mathcal{X}$, and let $\mathcal{Y}$ and $\mathcal{Z}$ denote two finite reproduction alphabets. Let d_1: $\mathcal{X} \times \mathcal{Y} \to [0, \infty)$ and d_2: $\mathcal{X} \times \mathcal{Z} \to [0, \infty)$, denote two single-letter distortion measures. Let $\mathcal{X}^N$, $\mathcal{Y}^N$, and $\mathcal{Z}^N$ denote the Nth-order Cartesian powers of $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{Z}$, respectively. The distortion between a source vector $\boldsymbol{x} = (x_1, \cdots, x_N) \in \mathcal{X}^N$ and a reproduction vector $\boldsymbol{y} = (y_1, \cdots, y_N) \in \mathcal{Y}^N$ is defined as

$$d_1(\boldsymbol{x}, \boldsymbol{y}) = \sum_{i=1}^{N} d_1(x_i, y_i).$$

Similarly, for $\boldsymbol{z} = (z_1, \cdots, z_n)$, we define

$$d_2(\boldsymbol{x}, \boldsymbol{z}) = \sum_{i=1}^{N} d_2(x_i, z_i).$$

Throughout the correspondence, scalar random variables will be denoted by capital letters while their sample values will be denoted by the respective lower case letters. A similar convention will apply to random N-dimensional vectors and their sample values, which will be denoted by boldface letters. Thus for example, $\boldsymbol{X}$ will denote a random N-vector $(X_1, \cdots, X_N)$, and $\boldsymbol{x} = (x_1, \cdots, x_N)$ is a specific vector value in $\mathcal{X}^N$. Sources and channels will be denoted generically by capital letters (sometimes indexed by the names of the corresponding random variables), e.g., P, Q_{XYZ}, W, V, etc., where these entities denote the set of (conditional or unconditional) letter probabilities, e.g., P is understood as a vector of letter probabilities $\{P(x), x \in \mathcal{X}\}$. For auxiliary random variables $(X, Y, Z) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$, that will be used throughout the sequel, the joint PMF will be denoted by

$$Q_{XYZ} = \{Q_{XYZ}(x, y, z), x \in \mathcal{X}, y \in \mathcal{Y}, z \in \mathcal{Z}\}.$$

Marginal and conditional PMF's that are derived from Q_{XYZ} will be denoted also by Q with an appropriate subscript, e.g., Q_X is the marginal PMF of X, $Q_{Z|XY}$ is the conditional PMF of Z given X and Y, and so on. For N-vectors, the probability of $\boldsymbol{x} \in \mathcal{X}^N$ will be denoted by

$$P^N(\boldsymbol{x}) = \prod_{i=1}^{N} P(x_i).$$

The probability of an event $A \subseteq \mathcal{X}^N$ will be denoted by $P^N\{A\}$, or by $\Pr\{A\}$ whenever there is no room for ambiguity regarding the underlying probability measure. The cardinality of a finite set A will be denoted by $|A|$. The operator $\boldsymbol{E}\{\cdot\}$ will denote expectation w.r.t. the underlying source P. Expectation w.r.t. Q_{XYZ} will be denoted by $\boldsymbol{E}_Q\{\cdot\}$.

For a given source vector $\boldsymbol{x} \in \mathcal{X}^N$, the empirical probability mass function (EPMF) is the vector $P_{\boldsymbol{x}} = \{P_{\boldsymbol{x}}(a),\, a \in \mathcal{X}\}$, where $P_{\boldsymbol{x}}(a) = N_{\boldsymbol{x}}(a)/N$, $N_{\boldsymbol{x}}(a)$ being the number of occurrences of the letter a in the vector $\boldsymbol{x}$. The type class T_P associated with a given PMF P, is the set of all vectors $\boldsymbol{x} \in \mathcal{X}^N$ such that $P_{\boldsymbol{x}} = P$. For two positive sequences $\{a_N\}_{N\geq 1}$ and $\{b_N\}_{N\geq 1}$, the notation $a_N \geq b_N$ means that $N^{-1}\ln(a_N/b_N) \to 0$ as $N \to \infty$, and in words, a_N is said to be *exponentially equal* to b_N. Similarly, $a_N \geq b_N$ means that

$$\liminf_{N\to\infty} N^{-1}\ln(a_N/b_N) \geq 0$$

and in words, a_N is said to be *exponentially at least as large as* b_N, or, b_N is *exponentially no larger* than a_N, and so on.

For two memoryless sources P and P', let

$$\mathcal{D}(P'\|P) = \sum_{x\in\mathcal{X}} P'(x)\ln\frac{P'(x)}{P(x)} \tag{2}$$

denote the relative entropy between P' and P. For a given random pair (X, Y) governed by Q_{XY}, let $I(X;Y)$ denote the mutual information between X and Y. Let $R(D, P')$ denote the rate-distortion function of P', w.r.t. d_1, i.e.,

$$R(D, P') = \inf\{I(X, Y)\colon Q_X = P',\, \boldsymbol{E}_Q d_1(X, Y) \leq D\}. \tag{3}$$

In [1] we defined the following terminology for the basic, single-stage guessing problem. We provide here definitions that are slightly simpler than in [1], but they are equivalent in the finite-alphabet case considered here. Let

$$S_1(\boldsymbol{y}, D) \triangleq \{\boldsymbol{x}\colon d_1(\boldsymbol{x}, \boldsymbol{y}) \leq ND\}.$$

Definition 1: A D-admissible guessing strategy is an ordered list $\mathcal{G}_N = \{\boldsymbol{y}_1, \boldsymbol{y}_2, \cdots\}$ of vectors in $\mathcal{Y}^N$, henceforth referred to as guessing words, such that

$$\bigcup_i S_1(\boldsymbol{y}_i, D) = \mathcal{X}^N. \tag{4}$$

Definition 2: The guessing function $G_N(\cdot)$ induced by a D-admissible guessing strategy $\mathcal{G}_N$, is the function that maps each $\boldsymbol{x} \in \mathcal{X}^N$ into a positive integer, which is the index j of the first guessing codeword $\boldsymbol{y}_j \in \mathcal{G}_N$ such that $d(\boldsymbol{x}, \boldsymbol{y}_j) \leq ND$.

Definition 3: The optimum ρth-order single-stage guessing exponent theoretically attainable at distortion level D is defined, whenever the limit exists, as

$$\mathcal{E}_1(D, \rho) \triangleq \lim_{N\to\infty}\frac{1}{N}\min_{\mathcal{G}_N}\ln \boldsymbol{E}\{G_N(\boldsymbol{X})^\rho\} \tag{5}$$

where the minimum is taken over all D-admissible guessing strategies, and the subscript "1" indicates the fact that the class of single-stage guessing strategies is considered.

The main result of [1] is that for a memoryless source, and an additive distortion measure, $\mathcal{E}_1(D, \rho)$ exists and has a single-letter characterization given by

$$\mathcal{E}_1(D, \rho) = E_1(D, \rho) \triangleq \max_{P'}[\rho R(D, P') - \mathcal{D}(P'\|P)]. \tag{6}$$

Note that $E_1(D, \rho)$ depends on the source P. However, since the underlying source P is fixed, and to avoid cumbersome notation, the dependency of E_1 on P is not denoted explicitly.

We now turn to the two-stage guessing problem, which in its basic form, is defined as follows. A memoryless source P randomly draws a realization $\boldsymbol{x} \in \mathcal{X}^N$ of a random vector $\boldsymbol{X}$. For a given intermediate distortion level D_1 w.r.t. distortion measure d_1, and a given target distortion level D_2 w.r.t. distortion measure d_2, the guesser first presents a sequence of guesses $\boldsymbol{y}_1, \boldsymbol{y}_2, \cdots$, until the first time that $d_1(\boldsymbol{x}, \boldsymbol{y}_i) \leq ND_1$, and then is temporarily scored by the number of guesses thus far $G_N^1(\boldsymbol{x}) = i$. In the second stage of the guessing process, the guesser provides another sequence of guesses $\boldsymbol{z}_{i1}, \boldsymbol{z}_{i2}, \cdots$, corresponding to i, until the first j such that $d_2(\boldsymbol{x}, \boldsymbol{z}_{ij}) \leq ND_2$, and the score increases by the additional number of guesses $G_N^2(\boldsymbol{x}) = j$. The question is: What is the best one can do in designing the guessing lists so as to minimize the exponential growth rate of the ρth moment of the total number of guesses $G_N^1(\boldsymbol{X}) + G_N^2(\boldsymbol{X})$? Clearly, the approach of using an intermediate search makes sense only if $E_1(D_1, \rho)$ w.r.t. distortion measure d_1 is smaller than $E_1(D_2, \rho)$ w.r.t. distortion measure d_2. If $d_1 = d_2$, this simply means that $D_1 > D_2$.

We next provide definitions for the two-stage case which are analogous to our earlier definitions for the single-stage case. In addition to the above definition of $S_1(\boldsymbol{y}, D)$, for a given $\boldsymbol{z} \in \mathcal{Z}^N$, let $S_2(\boldsymbol{z}, D) = \{\boldsymbol{x}\colon d_2(\boldsymbol{x}, \boldsymbol{z}) \leq ND\}$.

Definition 4: Given a source P, an intermediate distortion level D_1, and a target distortion level D_2, an admissible two-stage guessing strategy $\mathcal{G}_N$ comprises a D_1-admissible guessing strategy $\mathcal{G}_N^1 = \{\boldsymbol{y}_i,\, i = 1, 2, \cdots\}$, referred to as a first-layer guessing list, with a guessing function $G_N^1(\cdot)$, and a set of lists, $\{\mathcal{G}_N(i),\, i = 1, 2, \cdots\}$, $\mathcal{G}_N(i) = \{\boldsymbol{z}_{ij},\, j = 1, 2, \cdots\}$, $\boldsymbol{z}_{ij} \in \mathcal{Z}^N$, $i, j = 1, 2, \cdots$, referred to as second-layer guessing lists, such that for all i

$$\bigcup_j S_2(\boldsymbol{z}_{ij}, D_2) \supseteq S_1(\boldsymbol{y}_i, D_1) \bigcap \left[\bigcap_{k=1}^{i-1} S_1(\boldsymbol{y}_k, D_1)^c\right].$$

Comment: This set inequality takes into account the fact that if $G_N^1(\boldsymbol{x}) = i$, then $\boldsymbol{x}$ is in $S_1(\boldsymbol{y}_i, D_1)$, but not in any of the spheres associated with earlier examined guesses $S_1(\boldsymbol{y}_k, D_1)$, $k = 1, \cdots, i-1$. Hence, the second-layer guessing list corresponding to i must cover only source vectors with these properties.

Definition 5: The guessing function induced by a given admissible two-layer guessing strategy is given by

$$G_N(\boldsymbol{x}) = G_N^1(\boldsymbol{x}) + G_N^2(\boldsymbol{x}) \tag{7}$$

where $G_N^1(\cdot)$ is the guessing function induced by the associated first-layer guessing strategy $\mathcal{G}_N^1$, and $G_N^2(\boldsymbol{x})$ is the index j of the first codeword $\boldsymbol{z}_{ij} \in \mathcal{G}_N(i)$, such that $d_2(\boldsymbol{x}, \boldsymbol{z}_{ij}) \leq ND_2$, where $i = G_N^1(\boldsymbol{x})$.

Before we turn to characterize the best attainable two-stage guessing exponent, we review some known results on the multistage source coding problem [2], [6], [7] (see also [4] and [8]), which are intimately related to the two-stage guessing problem considered here. We first present some definitions associated with two-stage source codes.

A rate-R_1 block code of length N consists of an encoder

$$f_N^1\colon \mathcal{X}^N \to \{1, 2, \cdots, 2^{NR_1}\}$$

and a decoder

$$g_N^1\colon \{1, 2, \cdots, 2^{NR_1}\} \to \mathcal{Y}^N.$$

A refined rate-R_2 block code of length N $(R_2 > R_1)$ consists of an encoder

$$f_N^2\colon \mathcal{X}^N \to \{1, 2, \cdots, 2^{N(R_2-R_1)}\}$$

and a decoder

$$g_N^2\colon \{1, 2, \cdots, 2^{NR_1}\} \times \{1, 2, \cdots, 2^{N(R_2-R_1)}\} \to \mathcal{Z}^N.$$

A quadruple (R_1, R_2, D_1, D_2) is referred to as an *achievable quadruple* w.r.t. a source P if for every $\epsilon > 0$, $\delta > 0$, and N sufficiently large, there exists a length-N block code (f_N^1, g_N^1) of rate not exceeding $R_1 + \delta$, and a refined length-N block code (f_N^2, g_N^2) of rate not exceeding $R_2 + \delta$, such that

$$\begin{aligned}\Pr\big\{&d_1(\boldsymbol{X}, g_N^1(f_N^1(\boldsymbol{X})) \le ND_1,\\ &d_2(\boldsymbol{X}, g_N^2(f_N^1(\boldsymbol{X}), f_N^2(\boldsymbol{X}))) \le ND_2\big\} \ge 1-\epsilon. \end{aligned} \tag{8}$$

To characterize the region of achievable quadruples (R_1, R_2, D_1, D_2), consider an auxiliary random vector (X, Y, Z) governed by a PMF Q_{XYZ}, and let $I(X; YZ)$ denote the mutual information between X and (Y, Z).

Theorem 1 ([2], [6], [7]): For a memoryless source P, two additive distortion measures d_1 and d_2, and two distortion levels D_1 and D_2, respectively, a quadruple (R_1, R_2, D_1, D_2) is achievable w.r.t. P if and only if there exists a PMF Q_{XYZ} such that $Q_X = P$, $I(X;Y) \le R_1$, $I(X; YZ) \le R_2$, $\boldsymbol{E}_Q d_1(X,Y) \le D_1$, and $\boldsymbol{E}_Q d_2(X, Z) \le D_2$.

An immediate corollary [7, Corollary 1] to this theorem states that given D_1, D_2, and R_1, the minimum achievable R_2, denoted by $\boldsymbol{R}(R_1, D_1, D_2, P)$, is given by $\min I(X; YZ)$ over all $\{Q_{XYZ}\}$ such that $Q_X = P$, $I(X;Y) \le R_1$, $\boldsymbol{E}_Q d_1(X,Y) \le D_1$, and $\boldsymbol{E}_Q d_2(X,Z) \le D_2$.

III. A Lower Bound

We are now ready to present our main result, which is a single-letter characterization of a lower bound on the best two-stage guessing exponent theoretically attainable. Let d_1 and d_2 be two given distortion measures as above, and let D_1 and D_2 be two given distortion levels, respectively. For a given memoryless source P', let

$$K(D_1, D_2, P') = \min_{\mathcal{S}} \max\{I(X;Y), I(X; Z|Y)\} \tag{9}$$

where $I(X; Z|Y)$ is the conditional mutual information between X and Z given Y, and

$$\mathcal{S} \triangleq \big\{Q_{XYZ}\colon Q_X = P', \boldsymbol{E}_Q d_1(X,Y) \le D_1, \boldsymbol{E}_Q d_2(X,Z) \le D_2\big\}. \tag{10}$$

Now, let

$$E_2(D_1, D_2, \rho) = \max_{P'}[\rho K(D_1, D_2, P') - \mathcal{D}(P'\|P)]. \tag{11}$$

The following theorem tells us that $E_2(D_1, D_2, \rho)$ is a lower bound on the best attainable two-stage guessing exponent.

Theorem 2: Let P be a finite-alphabet memoryless source, d_1 and d_2 two additive distortion measures, D_1 an intermediate distortion level, and D_2 a target distortion level. Then

$$\liminf_{N\to\infty} \frac{1}{N} \min_{\mathcal{G}_N} \ln \boldsymbol{E}\{G_N(\boldsymbol{x})^\rho\} \ge E_2(D_1, D_2, \rho). \tag{12}$$

Discussion: The intuitive interpretation of the expression of $K(D_1, D_2, P')$ is that at each level, the number of guesses is exponential, i.e., exponentially $e^{NI(X;Y)}$ guesses in the first level and $e^{NI(X;Z|Y)}$ in the second. Thus the exponential order of the total number of guesses is dominated by the larger exponent. This is different from the two-step source-coding problem, where the codebook sizes of the two levels multiply, and so, their exponents (the rates) sum up to $I(X;Y) + I(X; Z|Y) = I(X; YZ)$.

The remaining part of this section is devoted to the proof of Theorem 2.

Proof: For a given positive integer N, let $\mathcal{G}_N$ be an arbitrary two-stage guessing scheme with distortion levels D_1 and D_2. Similarly as in the proof of [1, Theorem 1], we begin with the following chain of inequalities for an arbitrary auxiliary memoryless source P':

$$\begin{aligned}\boldsymbol{E}\{G_N(\boldsymbol{X})^\rho\} &= \boldsymbol{E}_{P'}\left[G_N(\boldsymbol{X})^\rho \prod_{i=1}^N \frac{P(X_i)}{P'(X_i)}\right]\\ &= \boldsymbol{E}_{P'} \exp\left[\rho \ln G_N(\boldsymbol{X}) + \sum_{i=1}^N \ln \frac{P(X_i)}{P'(X_i)}\right]\\ &\ge \exp\left[\rho \boldsymbol{E}_{P'} \ln G_N(\boldsymbol{X}) - N\mathcal{D}(P'\|P)\right]\\ &\ge \exp\big[\rho \max\{\boldsymbol{E}_{P'} \ln G_N^1(\boldsymbol{X}),\\ &\qquad \boldsymbol{E}_{P'} \ln G_N^2(\boldsymbol{X})\} - N\mathcal{D}(P'\|P)\big] \end{aligned} \tag{13}$$

where for the first inequality, we have used Jensen's inequality together with the convexity of the exponential function, and for the second inequality, we have used the fact that

$$G_N(\boldsymbol{X}) = G_N^1(\boldsymbol{X}) + G_N^2(\boldsymbol{X}) \ge \max\{G_N^1(\boldsymbol{X}), G_N^2(\boldsymbol{X})\}.$$

Since P' is an arbitrary memoryless source, the proof will be complete if we show that

$$\max\left\{\frac{1}{N}\boldsymbol{E}_{P'} \ln G_N^1(\boldsymbol{X}), \frac{1}{N}\boldsymbol{E}_{P'} \ln G_N^2(\boldsymbol{X})\right\} \ge K(D_1, D_2, P') - o(N) \tag{14}$$

for every P'. Now, let us define

$$R_N = \frac{1}{N}\boldsymbol{E}_{P'} \ln G_N^1(\boldsymbol{X}) \tag{15}$$

and

$$\Delta_N = \frac{1}{N}\boldsymbol{E}_{P'} \ln G_N^2(\boldsymbol{X}). \tag{16}$$

Intuitively, the functions $L_1(\boldsymbol{X}) = \ln G_N^1(\boldsymbol{X})$ and $L_2(\boldsymbol{X}) = \ln G_N^2(\boldsymbol{X}) + \ln G_N^2(\boldsymbol{X})$ are (within negligible terms for large N) legitimate code length functions (in nats) for lossless entropy coding of the locations of the guessing codewords, and so, one would expect $(R_N, R_N + \Delta_N, D_1, D_2)$ to be "essentially" an achievable quadruple in the sense used in Theorem 1. However, this theorem cannot be used as is to establish such an argument because it deals with fixed-rate coding, without allowing for variable-length entropy coding. Nevertheless, in the Appendix, we prove that there exists a constant $c = c(|\mathcal{Y}|, |\mathcal{Z}|)$ such that for all N

$$(R_N + c \ln(N+1)/N, R_N + \Delta_N + c\ln(N+1)/N, D_1, D_2)$$

is an achievable quadruple w.r.t. P'. This is done by constructing a fixed-rate length-l block code ($l \gg N$) that satisfies (8) with less than $e^{l(R_N + 0.5c \ln(N+1)/N)}$ codewords at the first level, and less than $e^{l(\Delta_N + 0.5c\ln(N+1)/N)}$ second-level codewords for each first-level codeword.

Using the same sphere covering arguments as in [6, Lemma 1], the existence of such a code implies that there must exist a PMF $Q_{XYZ} \in \mathcal{S}$ such that

$$R_N + \frac{c\ln(N+1)}{N} \ge I(X;Y) \tag{17}$$

and, at the same time,

$$\Delta_N + \frac{c\ln(N+1)}{N} \ge I(X; Z|Y) \tag{18}$$

and so

$$\begin{aligned}\max\{R_N, \Delta_N\} &\geq \max\{I(X;Y), I(X;Z|Y)\} - \frac{c\ln(N+1)}{N} \\ &\geq \min_{\mathcal{S}} \max\{I(X;Y), I(X;Z|Y)\} \\ &\quad - \frac{c\ln(N+1)}{N} \\ &= K(D_1, D_2, P') - \frac{c\ln(N+1)}{N} \end{aligned} \tag{19}$$

completing the proof of Theorem 2. □

IV. Achievability

The expression of $E_2(D_1, D_2, \rho)$ strongly suggests that the key to the achievability of $E_2(D_1, D_2, \rho)$ lies in the two-stage covering lemma (see, e.g., [7]), which is a straightforward extension of the ordinary single-stage covering lemma [9]. This two-stage covering lemma is the following.

Lemma 1 [7, Lemma 1]: If (R_1, R_2, D_1, D_2) is an achievable quadruple w.r.t. P', then there exist

i) A set $\mathcal{C}_1 \subset \mathcal{Y}^N$ such that

$$\frac{1}{N}\ln|\mathcal{C}_1| \leq R_1 + o(N) \tag{20}$$

and

$$\bigcup_{\boldsymbol{y}\in\mathcal{C}_1} S_1(\boldsymbol{y}, D_1) \supseteq T_{P'}. \tag{21}$$

ii) Sets $\mathcal{C}_2(\boldsymbol{y}) \subset \mathcal{Z}^N$, $\boldsymbol{y} \in \mathcal{C}_1$, such that

$$\frac{1}{N}\ln\sum_{\boldsymbol{y}\in\mathcal{C}_1}|\mathcal{C}_2(\boldsymbol{y})| \leq R_2 + o(N) \tag{22}$$

and

$$T_{P'}\bigcap\bigcup_{\boldsymbol{z}\in\mathcal{C}_2(\boldsymbol{y})} S_2(\boldsymbol{z}, D_2) \supseteq T_{P'}\bigcap S_1(\boldsymbol{y}, D_1) \quad \forall\,\boldsymbol{y}\in\mathcal{C}_1. \tag{23}$$

The construction of $\mathcal{C}_1$ and $\{\mathcal{C}_2(\boldsymbol{y})\}$ in [7] is as follows: since (R_1, R_2, D_1, D_2) is an achievable quadruple by assumption, the set

$$\{Q_{XYZ}: Q_{XYZ} \in \mathcal{S}, I(X;Y) \leq R_1, I(X;YZ) \leq R_2\}$$

is nonempty: first, it is shown that for any Q_{XYZ} in this set, a random selection of $M = e^{NI(X;Y)}$ vectors $\boldsymbol{y}_1, \cdots, \boldsymbol{y}_{M_1} \in T_{Q_Y}$, forming $\mathcal{C}_1$, satisfies (21) with high probability. Secondly, for each $\boldsymbol{y}_i \in \mathcal{C}_1$, let $\mathcal{C}_2(\boldsymbol{y}_i)$ be a randomly selected set of $M' \geq e^{NI(X;Z|Y)}$ vectors $\boldsymbol{z}_{i1}, \cdots \boldsymbol{z}_{i,M'}$ which, conditioned on $\boldsymbol{y}_i$, are in the type class associated with $Q_{Z|Y}$; then $\mathcal{C}_2(\boldsymbol{y}_i)$ satisfies (23) with high probability.

Using this lemma and its proof in [7], it is easy to see that $E_2(D_1, D_2, \rho)$ is achievable at least when the guesser is informed of the EPMF of the input sequence $\boldsymbol{x}$. This is done in the following manner. Let Q^*_{XYZ} attain $K(D_1, D_2, P_{\boldsymbol{X}})$. By applying the proof of Lemma 1 with $P' = P_{\boldsymbol{X}}$, $Q_{XYZ} = Q^*_{XYZ}$, $R_1 = I(X;Y)$, and $R_2 = I(X;YZ)$ (corresponding to Q^*_{XYZ}), one can create a first-layer guessing list $\boldsymbol{y}_1, \boldsymbol{y}_2, \cdots$ of size $\dot{\leq} e^{NI(X;Y)}$ that covers $T_{P_{\boldsymbol{x}}}$, and for each $\boldsymbol{y}_i$, a second-layer guessing list of size $\dot{\leq} e^{NI(X;Z|Y)}$ consisting of second-layer guessing codewords that cover $T_{P_{\boldsymbol{x}}} \cap S_1(\boldsymbol{y}_i, D_1)$. Thus regardless of the order of the guessing words at both levels, the total number of guesses $G^1_N(\boldsymbol{x}) + G^2_N(\boldsymbol{x})$ is exponentially at most

$$e^{NI(X;Y)} + e^{NI(X;Z|Y)} \doteq e^{NK(D_1, D_2, P_{\boldsymbol{x}})}.$$

Averaging the ρth power of this quantity w.r.t. the ensemble of EPMF's $\{P_{\boldsymbol{X}}\}$, we obtain by the method of types [9], the exponential order of $e^{NE_2(D_1, D_2, \rho)}$. The difference between this and the construction of an optimal two-stage code is that the optimum PMF Q_{XYZ} that minimizes the guessing exponent $\max\{I(X;Y), I(X;Z|Y)\}$ might be different than the one that minimizes the total coding rate $I(X;Y)+I(X;Z|Y) = I(X;YZ)$. Thus guessing words may have, in general, different compositions than optimal rate-distortion codewords.

Unfortunately, we were unable to construct a guessing strategy that achieves $E_2(D_1, D_2, \rho)$ without prior knowledge of the EPMF of $\boldsymbol{X}$. The difficulty lies in the fact that the guessing codebooks (at both levels) for different EPMF's may partially intersect. Therefore, no matter how the guessing lists for all EPMF's are integrated, there is no guarantee that the first-layer guessing word $\boldsymbol{y}_i$ for a given $\boldsymbol{x}$, will belong to the guessing codebook that corresponds to the EPMF of $\boldsymbol{x}$. Consequently, $\boldsymbol{x}$ may not be covered in the second-stage guessing list, or may require exponentially more than $e^{NI(X;Z|Y)}$ guesses.

Nevertheless, the assumption of prior knowledge of the EPMF of $\boldsymbol{X}$ is fairly reasonable as explained in Section I: first, in source-coding applications, which serve as the main motivation for the two-stage guessing problem, it is conceivable that the empirical PMF information is easily accessible to the guesser (or the encoder). Secondly, similarly as in the single-stage case, the validity of $E_2(D_1, D_2, \rho)$ as a lower bound is asymptotically unaffected by knowledge of the EPMF. This is true because asymptotically, the EPMF information is of zero rate. For the same reason, this setting still serves as an extension of [1].

More generally, consider a scenario where instead of one guesser we have L_N independent parallel guessers (or search machines) with guessing functions $G^{(j)}_N(\boldsymbol{x})$, $j = 1, \cdots, L_N$, and the guessing process stops as soon as one of the guessers succeeds. Thus the natural relevant performance criterion of interest is some moment of the guessing time $\boldsymbol{E}\{\min_j G^{(j)}_N(\boldsymbol{X})^\rho\}$. Again, it is easy to see that the validity of the lower bound $E_2(D_1, D_2, \rho)$ is asymptotically unaffected as long as $L_N \doteq 1$, that is, L_N grows subexponentially with N. Thus an asymptotically optimal solution to this problem would again suggest that each guesser will be responsible for one EPMF as described above, and so, $L_N \leq (N+1)^{|\mathcal{X}|-1}$.

In summary, it will be safe to argue that the lower bound $E_2(D_1, D_2, \rho)$ is achievable provided that we slightly extend the scope of the problem.

Furthermore, this assumption of knowing the EPMF has even deeper consequences. It provides the guesser with the flexibility to choose the first-layer distortion level D_1 depending on the EPMF.[1] This in general gives better guessing performance than that can be achieved if D_1 was fixed. Specifically, if only the target distortion D_2 is specified and D_1 is a design parameter subjected to optimization, then in the absence of prior information on $P_{\boldsymbol{X}}$, the best performance is bounded from below by

$$\begin{aligned}E^*_2(D_2, \rho) &= \inf_{D_1} E_2(D_1, D_2, \rho) \\ &= \inf_{D_1}\max_{P'}[\rho K(D_1, D_2, P') - \mathcal{D}(P'\|P)]. \end{aligned} \tag{24}$$

On the other hand, if $P_{\boldsymbol{x}}$ is known ahead of time, it is possible to achieve

$$E^{**}_2(D_2, \rho) = \max_{P'}\inf_{D_1}[\rho K(D_1, D_2, P') - \mathcal{D}(P'\|P)] \tag{25}$$

and clearly, $E^{**}_2(D_2, \rho) \leq E^*_2(D_2, \rho)$.

V. Successively Refinable Sources

Obviously, from the viewpoint of rate-distortion source coding, the best possible situation is when the rate-distortion function can be attained at both distortion levels. A source for which this can

[1] Furthermore, the first-level distortion measure d_1 may also be subjected to optimization.

be achieved is referred to as a *successively refinable* source in the literature (see, e.g., [5]). It turns out, as we show in this section, that the successively refinable case in this rate-distortion coding sense is also the best we can hope for from the viewpoint of guessing. Although this is fairly plausible, it is not quite obvious since the guessing performance criterion is somewhat different than that of coding.

To show this, we begin with a simple lower bound on $E_2^{**}(D_2, \rho)$ in terms of the single-stage guessing exponent function $E_1(D_2, \cdot)$.

Lemma 2: For every memoryless source P

$$E_2^{**}(D_2, \rho) \ge E_1(D_2, \rho/2).$$

Proof:

$$\begin{aligned} K(D_1, D_2, P') &= \min_{\mathcal{S}} \max\{I(X;Y), I(X;YZ) - I(X;Y)\} \\ &\ge \min_{\mathcal{S}} \max\{I(X;Y), I(X;Z) - I(X;Y)\} \\ &\ge \min_{\mathcal{S}} \max\{I(X;Y), R(D_2, P') - I(X;Y)\} \\ &\ge \frac{1}{2} R(D_2, P'). \end{aligned} \tag{26}$$

Since the rightmost side is independent of D_1, then

$$\inf_{D_1} K(D_1, D_2, P') \ge \tfrac{1}{2} R(D_2, P') \tag{27}$$

and so

$$E_2^{**}(D_2, \rho) \ge \max_{P'} \left[\frac{\rho}{2} R(D_2, P') - \mathcal{D}(P' \| P)\right] = E_1\left(D_2, \frac{\rho}{2}\right) \tag{28}$$

completing the proof of Lemma □

As we show next, in the successively refinable case, this lower bound is met.

Lemma 3: If the distortion measures d_1 and d_2 are such that every memoryless source P' is successively refinable for every D_1 together with the given target distortion level D_2, then for every memoryless source P, $E_2^{**}(D_2, \rho) = E_1(D_2, \rho/2)$.

Comment: If $d_1 = d_2$ is the Hamming distortion measure, the condition of Lemma 3 is met. Another case is where d_1 and d_2 are arbitrary distortion measures and $D_2 = 0$.

Proof of Lemma 3: Consider a guesser that is informed of the EPMF $P_{\boldsymbol{x}}$ of $\boldsymbol{x}$, and chooses $D_1 = D_1(P_{\boldsymbol{x}})$ such that

$$R(D_1, P_{\boldsymbol{x}}) = R(D_2, P_{\boldsymbol{x}})/2.$$

Since $P_{\boldsymbol{x}}$ is assumed successively refinable, the quadruple $(R(D_2, P_{\boldsymbol{x}})/2, R(D_2, P_{\boldsymbol{x}}), D_1, D_2)$ is achievable w.r.t. $P_{\boldsymbol{x}}$, and so there exists a PMF Q_{XYZ} for which $Q_X = P_{\boldsymbol{x}}$, $\boldsymbol{E}_Q d_1(X, Y) \le D_1$, $\boldsymbol{E}_Q d_2(X, Z) \le D_2$, and $I(X;Y) = I(X;Z|Y) = R(D_2, P_{\boldsymbol{x}})/2$. Thus for every sequence $\boldsymbol{x}$, $G_N(\boldsymbol{x}) \dot{\le} e^{NR(D_2, P_{\boldsymbol{x}})/2}$, and so, $E_2^{**}(D_2, \rho) = E_1(D_2, \rho/2)$, completing the proof of Lemma 3. □

Discussion: Intuitively, the successively refinable case reflects a situation where for each $P_{\boldsymbol{X}}$, the guessing complexity is divided evenly between the two levels. More generally, in a k-stage guessing system this would suggest that for a target distortion level D_k, the best guessing exponent is $E_1(D_k, \rho/k)$, which by the convexity of E_1 in ρ [1], cannot be larger than $E_1(D_k, \rho)/k$ (with strict inequality unless $R(D_k, P) = \max_P R(D_k, P)$). Returning to the case $k = 2$, this means that the effect of two-stage guessing, in the successively refinable case, is even better than halving the exponent.

For the sake of comparison, consider another form of a two-stage guessing list, where the first stage makes guesses on the first $N/2$ coordinates of $\boldsymbol{X}$ (until distortion D is achieved on these coordinates) and the second stage then makes guesses on the second half of the coordinates. In this case, we get

$$\begin{aligned} &\boldsymbol{E}\{[G_{N/2}(X_1, \cdots, X_{N/2}) + G_{N/2}(X_{N/2+1}, \cdots, X_N)]^\rho\} \\ &\qquad \ge \exp[N(E_1(D, \rho)/2] \end{aligned} \tag{29}$$

which means exactly halving the exponent. Thus the earlier proposed two-stage guessing mechanism has better guessing performance. However, the difference between the two approaches vanishes as the number of hierarchy levels k grows.

VI. Relation to the Two-Stage Source Coding Error Exponent

Consider now a situation where both D_1 and D_2 are specified (e.g., good guessing exponents are required at two specified distortion levels), and again, the guesser knows in advance the EPMF of $\boldsymbol{X}$. In this case, as we already proved, the best guessing exponent achievable is $E_2(D_1, D_2, \rho)$. We will now relate this to the two-stage source-coding error exponent, characterized in [7].

For an achievable quadruple (R_1, R_2, D_1, D_2), the two-stage source-coding error exponent $F(R_1, R_2, D_1, D_2)$ is defined as the best attainable exponential decay rate of the probability of the event

$$\begin{aligned} \mathcal{B} = \{\boldsymbol{x}&: d_1(\boldsymbol{x}, g_N^1(f_N^1(\boldsymbol{x}))) > ND_1, \\ &\text{or } d_2(\boldsymbol{x}, g_N^2(f_N^1(\boldsymbol{x}), f_N^2(\boldsymbol{x}))) > ND_2\}. \end{aligned}$$

Kanlis and Narayan [7] have shown that

$$F(R_1, R_2, D_1, D_2) = \min \mathcal{D}(P' \| P) \tag{30}$$

where the minimum is over the set

$$\begin{aligned} &\mathcal{K}(R_1, R_2, D_1, D_2) \\ &\quad = \{P': R(D_1, P') \ge R_1 \text{ or } \boldsymbol{R}(R_1, D_1, D_2, P') \ge R_2\} \end{aligned}$$

where $\boldsymbol{R}(R_1, D_1, D_2, P')$ is defined as in the last paragraph of Section II.

Let $R_0(D_1, D_2)$ be defined as the solution to the equation

$$R = \boldsymbol{R}(R, D_1, D_2, P') - R$$

with R being the unknown, provided that a solution exists. If a solution does not exist, i.e., if

$$R(D_1, P') > 0.5\, \boldsymbol{R}(R(D_1, P'), D_1, D_2, P'),$$

then $R_0(D_1, D_2) \triangleq 0$. It is easy to see that there is at most one solution to this equation. Now

$$\begin{aligned} &E_2(D_1, D_2, \rho) \\ &\quad = \max_{P'} \min_{\mathcal{S}} [\rho \max\{I(X;Y), \\ &\qquad I(X;YZ) - I(X;Y)\} - \mathcal{D}(P' \| P)] \\ &\quad \ge \max_{P'} \min_{\mathcal{S}} [\rho \inf_{R > I(X;Y)} \max\{R, I(X;YZ) - R\} \\ &\qquad - \mathcal{D}(P' \| P)] \\ &\quad \ge \max_{P'} [\rho \inf_{R > R(D_1, P')} \max\{R, \boldsymbol{R}(R, D_1, D_2, P') - R\} \\ &\qquad - \mathcal{D}(P' \| P)] \\ &\quad = \max_{P'} [\rho \max\{R(D_1, P'), R_0(D_1, D_2)\} \\ &\qquad - \mathcal{D}(P' \| P)] \\ &\quad = \max_{P'} \sup_{R < \max\{R(D_1, P'), R_0(D_1, D_2)\}} [\rho R - \mathcal{D}(P' \| P)] \\ &\quad = \sup_{R > 0} \max_{P' \in \mathcal{K}(R, 2R, D_1, D_2)} [\rho R - \mathcal{D}(P' \| P)] \\ &\quad = \sup_{R > 0} [\rho R - F(R, 2R, D_1, D_2)]. \end{aligned} \tag{31}$$

Thus for fixed D_1 and D_2, the guessing exponent$E_2(D_1, D_2, \rho)$ as a function of ρ, is lower-bounded by the one-sided Fenchel—Legendre transform (FLT) of $F(R, 2R, D_1, D_2)$ as a function of R. In [1], we established an analogous *equality* relation between the single-stage guessing exponent and the FLT of the single-stage source-coding exponent. Here, we cannot claim that the inequality is met with equality, in general. As for the inverse relation, note that (31) is equivalent to the statement

$$E_2(D_1, D_2, \rho) + F(R, 2R, D_1, D_2) \geq \rho R, \qquad \forall\, \rho > 0,\ R > 0 \tag{32}$$

which also means that

$$F(R, 2R, D_1, D_2) \geq \sup_{\rho>0} [\rho R - E_2(D_1, D_2, \rho)].$$

It should be pointed out that in [1] equality for all R is not guaranteed either. While the right-hand side is clearly a convex function of R, the function $F(R, 2R, D_1, D_2)$ is not necessarily so. This is demonstrated in the following example.

Example: Let P be a binary memoryless source with letter probabilities p and $1-p$, and let $d_1 = d_2$ be the Hamming distortion measure. Let $h(p) = -p \ln p - (1-p) \ln(1-p)$ denote the binary entropy function. Since $R(D, P') = h(p') - h(D)$ and binary sources with the Hamming distortion measure are successively refinable [5], then in this case

$$\begin{aligned}\mathcal{K}(R, 2R, D_1, D_2) \\ = \{P'\colon h(p') \geq \min\{R + h(D_1), 2R + h(D_2)\}\}.\end{aligned} \tag{33}$$

Now, let $R^* = h(D_1) - h(D_2)$, assume that $h(p) < R^* + h(D_1)$, and define

$$U(t) \triangleq \min_{x\colon h(x)\geq t} \left[x \ln \frac{x}{p} + (1-x) \ln \frac{1-x}{1-p}\right] \tag{34}$$

which for $p < 1/2$, $t > h(p)$, can be also written as

$$U(t) = h^{-1}(t) \ln \frac{h^{-1}(t)}{p} + (1 - h^{-1}(t)) \ln \frac{1 - h^{-1}(t)}{1-p} \tag{35}$$

where $h^{-1}(\cdot)$ is the inverse of $h(\cdot)$ in the range where the argument of $h(\cdot)$ is less than $1/2$. Clearly, $U(t)$ is a monotonically increasing, differentiable function in the above range, and let $U'(t)$ denote the derivative. Now, it is easy to see that

$$\begin{aligned}F(R, 2R, D_1, D_2) &= U(\min\{R + h(D_1), 2R + h(D_2)\}) \\ &= \begin{cases} U(2R + h(D_2)) & R \leq R^* \\ U(R + h(D_1)) & R > R^*. \end{cases}\end{aligned} \tag{36}$$

This means that the derivative of $F(R, 2R, D_1, D_2)$ w.r.t. R, which is positive, jumps at $R = R^*$ from $2U'(R^* + h(D_1))$ down to $U'(R^* + h(D_1))$, which, in turn, means that $F(R, 2R, D_1, D_2)$ cannot be convex in this case.

VII. Conclusion

We have derived a lower bound on the two-level guessing exponent, and discussed the conditions for its achievability. It has been also shown that the successively refinable case is the ideal case from the viewpoint of guessing as well as coding. Finally, we have shown that the two-level guessing exponent can be lower-bounded in terms of the two-level source-coding error exponent function with $R_2 = 2R_1$. However, this bound is not always tight.

Some open problems for future research are the following: i) Devise a two-level guessing strategy that is not informed of the EPMF but still attains $E_2(D_1, D_2, \rho)$. ii) Alternatively, find a tighter lower bound that can be achieved in the absence of knowledge of the EPMF. iii) Characterize the optimum performance for classes of more sophisticated guessing/searching mechanisms (e.g., take advantage of the full information carried by unsuccessful guesses thus far). These issues are currently under investigation.

Appendix

In this Appendix, we prove that for some constant c, that depends only on the reproduction alphabet sizes, the quadruple

$$(R_N + c \ln(N+1)/N,\ R_N + \Delta_N + c \ln(N+1)/N,\ D_1,\ D_2)$$

is achievable w.r.t. P'. We begin with the following simple auxiliary result.

Lemma 4: Let $\mathcal{J} = \{1, \cdots, J\}$ (J a positive integer), and for a given positive integer n, let

$$\mathcal{T}_n = \{(u_1, \cdots, u_n) \in \mathcal{J}^n\colon \sum_{i=1}^n \ln u_i \leq nR\}$$

for some positive real R. Then

$$|\mathcal{T}_n| \leq (n+1)^{J-1} \exp\{n[R + \ln(2 \ln J + 2)]\}. \tag{A.1}$$

Proof of Lemma 4: First, observe that by the method of types [9], we have

$$|\mathcal{T}_n| \leq (n+1)^{J-1} e^{Bn} \tag{A.2}$$

where $B = \max\{H(V)\colon \boldsymbol{E} \ln V \leq R\}$, $H(V)$ and $\boldsymbol{E}$ being, respectively, the entropy and the expectation w.r.t. a random variable V. Thus it remains to show that $B \leq R + \ln(2 \ln J + 2)$. Consider the following PMF on $\mathcal{J}$:

$$F(v) = \frac{1}{Cv}, \qquad v = 1, \cdots, J \tag{A.3}$$

where

$$C \triangleq \sum_{v=1}^J 1/v \leq 1 + \ln J.$$

Now, let us examine the codeword length function (in nats)

$$L(v) = \frac{\lceil -\log_2 F(v) \rceil}{\log_2 e} \leq \ln v + \ln C + \ln 2. \tag{A.4}$$

Then, we have

$$\begin{aligned}H(V) &\leq \boldsymbol{E} L(V) \\ &\leq \boldsymbol{E} \ln V + \ln C + \ln 2 \\ &\leq R + \ln(2 \ln J + 2)\end{aligned} \tag{A.5}$$

completing the proof of Lemma 4.

Consider now sequences $(\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n) \in \mathcal{X}^l$, where $l = nN$ (n a positive integer), $\boldsymbol{x}_i \in \mathcal{X}^N$, $i = 1, \cdots, n$. Now, for a given $\delta > 0$, let

$$\begin{aligned}\mathcal{A}_l = \Bigg\{&\sum_{i=1}^n \ln G_N^1(\boldsymbol{x}_i) \leq Nn(R_N + \delta), \\ &\sum_{i=1}^n \ln G_N^2(\boldsymbol{x}_i) \leq Nn(\Delta_N + \delta)\Bigg\}.\end{aligned} \tag{A.6}$$

Let us consider a two-stage, fixed-rate block code for l-vectors that operates as follows: if $(\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n) \in \mathcal{A}_l^c$, then the all-zero codeword is assigned at both levels. Else, $(\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n)$ is encoded by codewords that are formed by concatenating the respective guessing words at both levels. Since $\mathcal{A}_l$ is fully covered by codewords within distortion levels D_1 and D_2, at both levels, respectively, and since, by the weak law of large numbers, the probability of A_l under P'

tends to unity as $n \to \infty$ (while N is kept fixed), we have constructed a sequence of fixed-rate block codes that satisfies (8).

To estimate the number of codewords (and hence the rate) at the first level code, we apply Lemma 4 by setting $R = N(R_N + \delta)$, $u_i = G_N^1(\boldsymbol{x}_i)$, and $J = |\mathcal{Y}|^N$, where the latter assignment expresses the fact that in the finite reproduction alphabet case, the guessing list size need not exceed the total number of possible reproduction vectors. Thus we can upper-bound the number of codewords in the first level by

$$M_1 \le (n+1)^{|\mathcal{Y}|^N} \exp\{n[N(R_N+\delta) + \ln(2\ln|\mathcal{Y}|^N + 2)]\} = \exp\left\{Nn\left[R_N + \delta + \frac{|\mathcal{Y}|^N \ln(n+1)}{Nn} + \frac{\ln(2N\ln|\mathcal{Y}|+2)}{N}\right]\right\}. \tag{A.7}$$

Letting $n \to \infty$ for fixed N, we see that the exponent of this expression tends to $R_N + \delta + \ln(2N \ln |\mathcal{Y}| + 2)/N$. In the same manner, one can verify that the total number of codewords at the second level satisfies

$$\limsup_{n\to\infty} \frac{1}{nN} \ln M_2 \le R_N + \Delta_N + 2\delta + \frac{1}{N}[\ln(2N\ln|\mathcal{Y}|+2) + \ln(2N\ln|\mathcal{Z}|+2)].$$

Clearly, there exists a constant c (that depends solely on $|\mathcal{Y}|$ and $|\mathcal{Z}|$) such that $c\ln(N+1)/N$ upper-bounds the $O(\log N/N)$ terms in the exponents of both M_1 and M_2, for all N. Finally, since δ is arbitrarily small, this implies that

$$(R_N + c\ln(N+1)/N,\ R_N + \Delta_N + c\ln(N+1)/N,\ D_1,\ D_2)$$

is an achievable quadruple w.r.t. P' by definition.

References

[1] E. Arikan and N. Merhav, "Guessing subject to distortion," *IEEE Trans. Inform. Theory*, vol. 44, pp. 1041–1056, May 1998; see also, EE-1015, Dept. Elec. Eng., Tech. Rep. CC-141, Technion, Haifa, Israel, Feb. 1996.

[2] V. Koshélev, "Hierarchical coding of discrete sources," *Probl. Pered. Inform.*, vol. 16, no. 3, pp. 31–49, 1980.

[3] ——, "Estimation of mean error for a discrete successive approximation scheme," *Probl. Pered. Inform.*, vol. 17, no. 3, pp. 20–33, July–Sept. 1981.

[4] H. Yamamoto, "Source coding theory for cascade and branching communication systems," *IEEE Trans. Inform. Theory*, vol. IT-27, pp. 299–308, May 1981.

[5] W. H. Equitz and T. M. Cover, "Successive refinement of information," *IEEE Trans. Inform. Theory*, vol. 37, pp. 269–275, Mar. 1991.

[6] B. Rimoldi, "Successive refinement of information: Characterization of achievable rates," *IEEE Trans. Inform. Theory*, vol. 40, pp. 253–259, Jan. 1994.

[7] A. Kanlis and P. Narayan, "Error exponents for successive refinement by partitioning," *IEEE Trans. Inform. Theory*, vol. 42, pp. 275–282, Jan. 1996.

[8] A. El Gamal and T. M. Cover, "Achievable rates for multiple descriptions," *IEEE Trans. Inform. Theory*, vol. IT-28, pp. 851–857, Nov. 1982.

[9] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems.* New York: Academic, 1981.

Almost-Sure Variable-Length Source Coding Theorems for General Sources

Jun Muramatsu and Fumio Kanaya

***Abstract*—Source coding theorems for general sources are presented. For a source μ, which is assumed to be a probability measure on all strings of infinite-length sequence with a finite alphabet, the notion of almost-sure sup entropy rate is defined; it is an extension of the Shannon entropy rate. When both an encoder and a decoder know that a sequence is generated by μ, the following two theorems can be proved: 1) in the almost-sure sense, there is no variable-rate source coding scheme whose coding rate is less than the almost-sure sup entropy rate of μ. and 2) in the almost-sure sense, there exists a variable-rate source coding scheme whose coding rate achieves the almost-sure sup entropy rate of μ.**

***Index Terms*—Almost-sure sup entropy rate, general sources, source coding theorems.**

I. Introduction

Throughout this correspondence, let $\hat{\mathcal{A}}$ be a finite set and $(\hat{\mathcal{A}}^\infty, \mathcal{F})$ a measurable space, where $\hat{\mathcal{A}}^\infty$ is the set of all strings of infinite length that can be formed from the symbols in $\hat{\mathcal{A}}$, and $\mathcal{F}$ is a σ-field of subsets of $\hat{\mathcal{A}}^\infty$.

Let μ be a probability measure defined on $(\hat{\mathcal{A}}, \mathcal{F})$. Then, we call $(\hat{\mathcal{A}}, \mathcal{F}, \mu)$ a probability space. We call μ a *general source* or simply a *source*. It should be noted that μ satisfies consistency restrictions. Traditionally, a source is defined as a sequence of random variables $\hat{X} \equiv \{\hat{X}n\}_{n=1}^\infty$, but if $\hat{X}$ satisfies consistency restrictions

$$\sum_{\hat{x}^{n+1}\in\hat{A}} \mathrm{Prob}(\hat{X}^{n+1} = \hat{x}^{n+1}) = \mathrm{Prob}(\hat{X}^n = \hat{x}^n), \qquad \forall \hat{x}^n \in \hat{A}^n, \quad \forall n \in \mathbb{N}$$

we can construct the probability measure $\mu_{\hat{X}}$ satisfying

$$\mu_{\hat{X}}^n(\hat{x}^n) \equiv \mathrm{Prob}(\hat{X}^n = \hat{x}^n)$$

where $\mu_{\hat{X}}^n$ is a probability distribution on $\hat{\mathcal{A}}^n$ induced by $\mu_{\hat{X}}$. Then, $\mu_{\hat{X}}$ can be considered as a general source.

We will prove almost-sure source coding theorems for general sources, placing no assumption on sources except consistency restrictions. To this end, we define the almost-sure sup entropy rate of a general source μ. Assuming that an encoder and a decoder know that a string is produced by μ, we can make the following two statements:

1) There is no variable-length code such that the coding rate of this code is less than the almost-sure sup entropy rate of the source with probability 1.

2) There exists a variable-length code such that the coding rate of this code is equal to the almost-sure sup entropy rate of the source with probability 1.

Manuscript received July 17, 1997; revised April 26, 1998. The material in this correspondence was presented in part at the 19th Symposium on Information Theory and Its Applications, Hakone, Japan, December 3–6, 1996 (in Japanese).

J. Muramatsu is with the NTT Communication Science Laboratories, Kyoto 619-0237, Japan.

F. Kanaya is with the Department of Information Science, Shonan Institute of Technology, Kanagawa 251-0046, Japan.

Communicated by M. Feder, Associate Editor for Source Coding.

Publisher Item Identifier S 0018-9448(99)00625-2.