Successive Refinement of Information

William H. R. Equitz, *Member, IEEE* , and Thomas M. Cover, *Fellow, IEEE*

Abstract —The successive refinement of information consists of first approximating data using a few bits of information, then iteratively improving the approximation as more and more information is supplied. The goal is to achieve an optimal description at each stage. In general an ongoing description is sought which is rate-distortion optimal whenever it is interrupted. It is shown that a rate distortion problem is successively refinable if and only if the individual solutions of the rate distortion problems can be written as a Markov chain. This implies in particular that tree structured descriptions are optimal if and only if the rate distortion problem is successively refinable. Successive refinement is shown to be possible for all finite alphabet signals with Hamming distortion, for Gaussian signals with squared-error distortion, and for Laplacian signals with absolute-error distortion. However, a simple counterexample with absolute error distortion and a symmetric source distribution shows that successive refinement is not always achievable.

Index Terms —Rate distortion, refinement, progressive transmission, multiuser information theory, squared-error distortion, tree structure.

I. INTRODUCTION

PROBLEMS are characterized in which optimal descriptions can be considered as refinements of previous optimal descriptions. For example, we may optimally describe a message with a particular amount of distortion and later decide that the message needs to be specified more accurately. Then, when an addendum to the original message is sent we hope that this refinement is as efficient as if the more strict requirements had been known at the start. In general, we ask whether it is possible to interrupt a transmission at any time without loss of optimality.

An example of successive refinement might be image compression in which one briefly describes a gross image and then follows with successive refinements of the description that further refine the image. The goal is to achieve the rate distortion bound at each stage. Similar remarks apply to voice compression.

The difficulty with achieving this goal is that optimal descriptions are not always refinements of one another. Difficulties arise even in the simple case of describing a single random variable X drawn from a standard normal distribution where the problem is to minimize the average squared error resulting from using a few bits to describe $X \sim N(0,1)$. If one bit of description is used, it is obvious that the optimal one bit description will specify whether X is

W. H. R. Equitz is with the IBM Almaden Research Center, 650 Harry Road, Department K52/802, San Jose, CA 95120-6099, e-mail: equitz@ibm.com.

T. M. Cover is with the Departments of Electrical Engineering and Statistics, Durand, Room 121, Stanford University, Stanford, CA 94305, e-mail: cover@isl.stanford.edu.
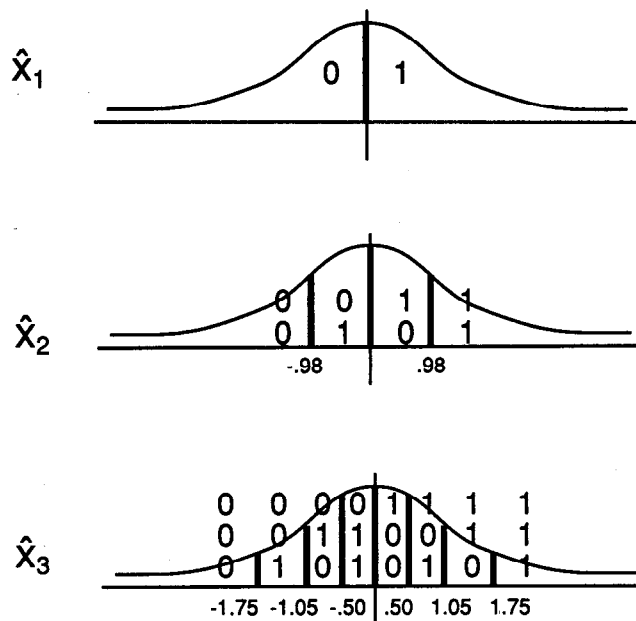
Fig. 1. Nonrefinable description.

positive or negative. For instance one should send a "0" to indicate that X is negative and a "1" otherwise, as indicated in Fig. 1. The reconstruction $\hat{X}_1$ resulting from this 1 bit description will be the centroid of the partition. Thus $\hat{X}_1 = -\sqrt{2/\pi}$ if $X < 0$, and $\hat{X}_1 = \sqrt{2/\pi}$ if $X \geq 0$. The resulting squared error distortion is $D = E(X - \hat{X})^2 = (\pi - 2)/\pi \approx 0.3634$.

If there are two bits available to describe X, then there is an optimal quantization [1] of the interval $(-\infty, \infty)$. Here the interval is quantized into four regions, and $\hat{X}_2$ is given by the centroid of the bin in which X happens to fall. Here it is clear that the two bit description is a refinement of the one bit description in the sense that one can merely append another bit to the optimal one bit description to transmit an optimal two bit description, i.e., the best four-cell partition is a refinement of the best two-cell partition.

However, Fig. 1 shows that the optimal quantization levels for the three bit description is not a refinement of the optimal two bit description. Optimal use of three bits of information about X requires advance knowledge that three full bits will be available.

This failure of successive refinement for the quantization of a single Gaussian random variable suggests that successive refinement is rarely achievable. However, if we consider long blocks of i.i.d. Gaussian random variables, we will see that successive refinement is always possible. Nonetheless, successive refinement, even with large block sizes, is not possible in general unless the solutions to the individual rate distortion problems obey a Markov relationship.

In Theorem 2 we prove that successive refinement from a coarse description $\hat{X}_1$ with distortion D_1 to a finer description $\hat{X}_2$ with distortion D_2 can be achieved if and only if the conditional distributions $p(\hat{x}_1|x)$ and $p(\hat{x}_2|x)$, which achieve $I(X;\hat{X}_i) = R(D_i)$, $i = 1,2$, are Markov compatible in the sense that we can write $\hat{X}_1 \to \hat{X}_2 \to X$ as a Markov chain.

Section IV then uses these necessary and sufficient conditions to exhibit a counterexample for which successive refinement cannot be achieved. In Section V we prove that all finite alphabet distributions with Hamming distortion are successively refinable and also exhibit two specific continuous valued problems in which successive refinement is achievable.

II. BACKGROUND AND STATEMENT OF THE PROBLEM

We recall the definition of the rate distortion function.

Definition 1 (Rate distortion function): For a probability mass function $p(x)$, $x \in \chi$, and distortion function $d(x,\hat{x})$ on $\chi \times \hat{\chi}$, the *rate distortion function* $R(D)$ is given by

$$R(D) = \min_{p(\hat{x}|x)} I(X;\hat{X}), \qquad (1)$$

where the minimum is over all conditional pmf's $p(\hat{x}|x)$ satisfying $\sum_{x,\hat{x}} p(x)p(\hat{x}|x)d(x,\hat{x}) \le D$. The distortion rate function $D(R)$ is the inverse function of $R(D)$, which can be characterized as

$$D(R) = \min_{p(\hat{x}|x)} Ed(X,\hat{X}), \qquad (2)$$

where the minimum is over all conditional pmf's $p(\hat{x}|x)$ satisfying $I(X;\hat{X}) \ge R$.

The rate distortion theorem states that a rate $R(D)$ description of $\{X_i\}$, X_i independent and identically distributed (i.i.d.), suffices to estimate the process to within distortion D. We now describe what we mean by successive refinement.

We consider a sequence of i.i.d. random variables $X_1, X_2, \cdots, X_n$ where each X_i is drawn from a source alphabet χ. We are given a reconstruction alphabet $\hat{\chi} = \chi$, and consider the distortion measure

$$d: \chi \times \hat{\chi} \to \mathcal{R}. \qquad (3)$$

The distortion measure on n-sequences in $\chi^n \times \hat{\chi}^n$ is defined by the average per-letter distortion

$$d(x^n, \hat{x}^n) = \frac{1}{n} \sum_{j=1}^{n} d(x_j, \hat{x}_j), \qquad (4)$$

where $x^n = (x_1, x_2, \cdots, x_n)$, and $\hat{x}^n = (\hat{x}_1, \hat{x}_2, \cdots, \hat{x}_n)$.

We say that we are successively refining a sequence of random variables $X_1, \cdots, X_n$ when we use a two-stage description that is optimal at each stage. First, as shown in Fig. 2 we describe the X sequence at rate R_1 bits per symbol and incur distortion D_1. Then we provide an addendum to the first message at rate $R_2 - R_1$ bits per symbol so that the two-stage resulting message now has distortion D_2. We shall say we have successively refined the sequence $X_1, \cdots, X_n$ if $R_1 = R(D_1)$ and $R_2 = R(D_2)$. In other words, we demand that we achieve the rate distortion limit at each of the two stages. In general, we will demand that we be able to achieve all points on the rate distortion curve.

Definition 2 (Successive refinement from distortion D_1 to distortion D_2): We shall say that *successive refinement* from distortion D_1 to distortion D_2 is achievable $(D_1 \ge D_2)$ if

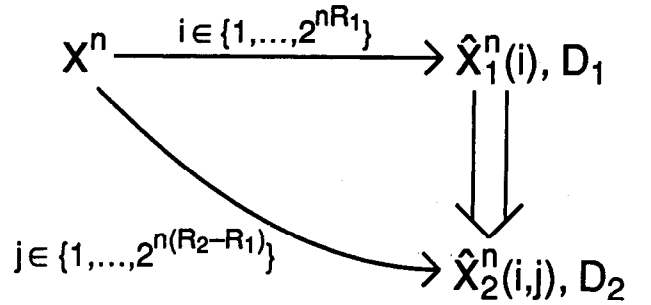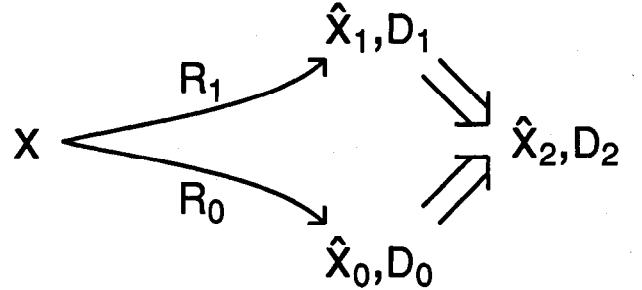Fig. 2. Successive refinement problem.

Fig. 3. Multiple descriptions problem.

there exists a sequence of encoding functions $i_n: \chi^n \to \{1, \cdots, 2^{nR_1}\}$ and $j_n: \chi^n \to \{1, \cdots, 2^{n(R_2-R_1)}\}$, and reconstruction functions $g_{1n}: \{1, \cdots, 2^{nR_1}\} \to \hat{\chi}_1^n$ and $g_{2n}: \{1, \cdots, 2^{nR_1}\} \times \{1, \cdots, 2^{n(R_2-R_1)}\} \to \hat{\chi}_2^n$, such that for $\hat{X}_1^n = g_{1n}(i_n(X^n))$ and for $\hat{X}_2^n = g_{2n}(i_n(X^n), j_n(X^n))$,

$$\limsup_{n \to \infty} Ed(X^n, \hat{X}_1^n) \le D(R_1), \qquad (5)$$

and

$$\limsup_{n \to \infty} Ed(X^n, \hat{X}_2^n) \le D(R_2), \qquad (6)$$

where $D(R)$ is the distortion rate function.

Definition 3 (Successive refinement in general): We say that a problem defined by a source distribution $p(x)$ and distortion measure $d(x,\hat{x})$ is *successively refinable in general* or simply successively refinable if successive refinement from distortion D_1 to distortion D_2 is achievable for every $D_1 \ge D_2$.

A. Related Problems

Our main tool is the achievable rate region for the multiple descriptions problem investigated by Gersho, Witsenhausen, Wolf, Wyner, Ziv, Ozarow, El Gamal, Cover, Berger, Zhang, and Ahlswede [2]–[6]. In this problem a sender wishes to describe the same sequence of random variables $X_1, X_2, \cdots, X_n$ to more than one receiver. The ith receiver will receive description $f_i(X^n) \in \{1, 2, \cdots, 2^{nR_i}\}$, from which it will produce an estimate $\hat{X}_{i1}, \hat{X}_{i2}, \cdots, \hat{X}_{in}$ of the original message. The distortion associated with representing the source symbol x with the symbol $\hat{x}$ is given by $d(x,\hat{x})$ and the distortion between the sequences $x^n = (x_1, x_2, \cdots, x_n)$ and $\hat{x}^n = (\hat{x}_1, \hat{x}_2, \cdots, \hat{x}_n)$ is given by (4).

An important special case is shown in Fig. 3, where there are three receivers, two of which receive individual descriptions and the third of which has access to both descriptions.

Information about the source is transmitted to receivers 0 and 1 at rates R_0 and R_1 respectively, and the two receivers individually generate estimates $\hat{X}_0$ and $\hat{X}_1$ with distortion D_0 and D_1, respectively. When they pool their information, a third estimate $\hat{X}_2$ with distortion D_2 is generated (with $D_2 \leq D_0, D_2 \leq D_1$). The rate distortion region is the set of achievable quintuples $(R_0, R_1, D_0, D_1, D_2)$.

The successive refinement problem is a special case of the multiple descriptions problem in which there is no constraint on $Ed(X, \hat{X}_0)$ and in which we require $R_1 = R(D_1)$ and $R_0 + R_1 = R(D_2)$.

We require the following achievable region established by El Gamal and Cover [5].

Theorem 1: A rate-distortion quintuple is achievable if there exists a probability mass distribution

$$p(x)p(\hat{x}_0, \hat{x}_1, \hat{x}_2 | x)$$

with

$$Ed(X, \hat{X}_m) \leq D_m, \quad m = 0, 1, 2,$$

such that

$$R_0 > I(X; \hat{X}_0), \tag{7}$$

$$R_1 > I(X; \hat{X}_1), \tag{8}$$

$$R_0 + R_1 > I(X; \hat{X}_0, \hat{X}_1, \hat{X}_2) + I(\hat{X}_0; \hat{X}_1). \tag{9}$$

Ahlswede [7] showed in the "no-excess-rate case," i.e., $R_0 + R_1 = R(D_2)$, that the conditions given by El Gamal and Cover are necessary as well as sufficient. (Zhang and Berger [6] exhibit a simple counterexample that shows that the conditions of El Gamal and Cover are sometimes not tight when $R_0 + R_1 > R(D_2)$.) However, successive refinement is indeed the case of no excess rate, which is the constraint under which (7)–(9) yield the entire rate region.

Also relevant to the successive refinement problem is the closely related conditional rate distortion problem formulated by R. M. Gray [8]–[10], which deals with the question of determining the minimum rate needed to describe a source at distortion D, when side information Y is present. See also Gray and Wyner [11]. Gray defines the conditional rate distortion function $R_{X|Y}(D)$ as

$$R_{X|Y}(D) = \inf_{p(\hat{x}|x, y)} I(X; \hat{X}|Y), \tag{10}$$

where the minimum is over all $p(\hat{x}|x, y)$ satisfying $\sum_{\hat{x}, x, y} p(\hat{x}|x, y)p(x, y)d(x, \hat{x}) \leq D$. Conditional rate-distortion theory is relevant to our question because one might say that in successive refinement $R_1 = R(D_1)$ and $R_2 - R_1 = R_{X|\hat{X}_1}(D_2)$. Of course, $R_{X|\hat{X}_1}$ is defined only when $\hat{X}_1$ is an i.i.d. sequence, which in general is not the case. Also relevant is the work of Yamamoto [12] and Kaspi and Berger [13].

III. ACHIEVABILITY OF SUCCESSIVE REFINEMENT

A. The Markov Conditions

Here we prove that successive refinement from coarse description $\hat{X}_1$ with distortion D_1 to fine description $\hat{X}_2$ with distortion D_2 is achievable if and only if the individual rate distortion solutions $p(\hat{x}_1|x)$ and $p(\hat{x}_2|x)$ for $D_1 \geq D_2$ are such that we can write $X \to \hat{X}_2 \to \hat{X}_1$ as a Markov chain. We do this by considering the successive refinement problem as a special case of the multiple descriptions problem and by

applying theorems from [5] and [7]. Let the source distribution $p(x)$ and the distortion $d(x, \hat{x})$ be given. Let $R(D)$ be defined as in (1).

Theorem 2 (Markovian characterization of successive refinement): Successive refinement with distortions D_1 and D_2 $(D_1 \geq D_2)$ can be achieved if and only if there exists a conditional distribution $p(\hat{x}_1, \hat{x}_2 | x)$ with

$$Ed(X, \hat{X}_1) \leq D_1, \tag{11}$$

and

$$Ed(X, \hat{X}_2) \leq D_2, \tag{12}$$

such that

$$I(X; \hat{X}_1) = R(D_1), \tag{13}$$

$$I(X; \hat{X}_2) = R(D_2), \tag{14}$$

and

$$p(\hat{x}_1, \hat{x}_2 | x) = p(\hat{x}_2 | x)p(\hat{x}_1 | \hat{x}_2). \tag{15}$$

Remark: The last condition is equivalent to saying that $X, \hat{X}_1, \hat{X}_2$ can be written as the Markov chain $X \to \hat{X}_2 \to \hat{X}_1$ or, equivalently, as $\hat{X}_1 \to \hat{X}_2 \to X$.

Proof:
(*Sufficiency*)—Let $p(\hat{x}_1|x)$ and $p(\hat{x}_2|x)$ satisfy (11)–(14). Let $\hat{X}_0$ be a dummy symbol (some constant). Fix the joint pmf $p(x)p(\hat{x}_0)(\hat{x}_2|x)p(\hat{x}_1|\hat{x}_2)$. The joint description achievable region of Theorem 1 becomes

$$R_0 > I(X; \hat{X}_0) = 0, \tag{16}$$

$$R_1 > I(X; \hat{X}_1) = R(D_1), \tag{17}$$

$$R_0 + R_1 > I(X; \hat{X}_0, \hat{X}_1, \hat{X}_2) + I(\hat{X}_0; \hat{X}_1) \tag{18}$$

$$= I(X; \hat{X}_2) = R(D_2), \tag{19}$$

where we have used (13) and (14). But the total rate R_2 is given by $R_2 = R_0 + R_1$. Thus $(R_1, R_2) = (R(D_1), R(D_2))$ is achievable.

(*Necessity*)—Successive refinement requires $R_1 = R(D_1)$ and $R_0 + R_1 = R(D_2)$. This is the "no excess rate" condition of Ahlswede for which the region of Theorem 1 is the entire achievable rate region. Thus there must exist a conditional pmf $p(\hat{x}_0, \hat{x}_1, \hat{x}_2 | x)$ with $Ed(X, \hat{X}_1) \leq D_1, Ed(X, \hat{X}_2) \leq D_2$ such that

$$R_1 = R(D_1) \geq I(X; \hat{X}_1), \tag{20}$$

$$R_0 = R(D_2) - R(D_1) \geq I(X; \hat{X}_0), \tag{21}$$

and

$$R_0 + R_1 = R(D_2) \geq I(X; \hat{X}_0, \hat{X}_1, \hat{X}_2) + I(\hat{X}_0; \hat{X}_1). \tag{22}$$

The definition of the rate-distortion function implies that $I(X; \hat{X}_1) \geq R(D_1)$, so (20) is satisfied if and only if $I(X; \hat{X}_1) = R(D_1)$. Expanding (22) by the chain rule yields

$$R(D_2) \geq I(X; \hat{X}_0, \hat{X}_1, \hat{X}_2) + I(\hat{X}_0; \hat{X}_1) \tag{23}$$

$$= I(X; \hat{X}_2) + I(X; \hat{X}_1 | \hat{X}_2)$$

$$\quad + I(X; \hat{X}_0 | \hat{X}_1, \hat{X}_2) + I(\hat{X}_0; \hat{X}_1) \tag{24}$$

$$\geq I(X; \hat{X}_2) \tag{25}$$

$$\geq R(D_2), \tag{26}$$

where the last inequality follows from the definition of the rate-distortion function, and inequality (25) follows from the nonnegativity of mutual information. Since the start and end of the chain are equal, all inequalities must be satisfied with equality, which implies that

$$I(\hat{X}_0; \hat{X}_1) = 0, \qquad (27)$$

$$I(X; \hat{X}_0 | \hat{X}_1, \hat{X}_2) = 0, \qquad (28)$$

$$I(X; \hat{X}_1 | \hat{X}_2) = 0, \qquad (29)$$

and

$$I(X; \hat{X}_2) = R(D_2). \qquad (30)$$

Equation (29) is equivalent to the Markovity of $X \to \hat{X}_2 \to \hat{X}_1$, while (28) forces the Markovity of $X \to (\hat{X}_1, \hat{X}_2) \to \hat{X}_0$. Thus from the above we must have $X \to \hat{X}_2 \to \hat{X}_1 \to \hat{X}_0$. Finally, (27) requires $\hat{X}_0$ to be independent of $\hat{X}_1$. We conclude that the achievability of $(R_1, R_2) = (R(D_1), R(D_2))$ guarantees the existence of a pmf $p(x)p(\hat{x}_1, \hat{x}_2 | x)$ satisfying (15). Thus successive refinement is achievable only by joint pmf's of the form $p(x)p(\hat{x}_2 | x)p(\hat{x}_1 | \hat{x}_2)p(\hat{x}_0)$, i.e., only if there exists $p(\hat{x}_1 | \hat{x}_2)$ such that $X \to \hat{X}_2 \to \hat{X}_1$. □

B. Codes for Successive Refinement

Let $p(\hat{x}_1)$ and $p(\hat{x}_2 | \hat{x}_1)$ be probability mass functions achieving the bound in Theorem 2. To generate the codebook for the first refinement, we draw 2^{nR_1} i.i.d. code vectors according to the distribution $\prod_{i=1}^{n} p(\hat{x}_{1i})$. We index these code vectors $\hat{x}_1^n(i)$, $i \in \{1, \cdots, 2^{nR_1}\}$. Then, for each i we generate a codebook for the second refinement with $2^{n(R_2 - R_1)}$ codewords drawn according to the conditional distribution $\prod_{k=1}^{n} p(\hat{x}_{2k} | \hat{x}_{1k}(i))$. We index these code vectors $\hat{x}_2^n(i,j)$, $i \in \{1, \cdots, 2^{nR_1}\}$, $j \in \{1, \cdots, 2^{n(R_2 - R_1)}\}$.

We describe the first refinement of a source vector x^n with the index i of the codeword that minimizes $d(x^n, \hat{x}_1^n(i))$. Next we describe the second refinement of x^n by the index $j \in \{1, \cdots, 2^{n(R_2 - R_1)}\}$ that minimizes $d(x^n, \hat{x}_2^n(i,j))$. Because successive refinement is a special case of the multiple descriptions problem, the proof of Theorem 1 [5] establishes that this method of encoding will achieve the desired rates and distortions.

We can now see that the codes which achieve successive refinement have a "tree structure," where the coarse descriptions occur near the root of the tree and the finer descriptions near the leaves. Although these tree structured codes will usually only be optimal asymptotically in the limit as the block length n grows to infinity, it is possible to use the idea of tree structured codes in practical finite block length schemes for describing messages with successive refinements. One such method is described in [14], [15].

IV. COUNTEREXAMPLE

In this section we show that not all problems are successively refinable. We now provide a sketch of a counterexample that has its roots in a problem described by Gerrish [16], which forms the basis for an exercise in Berger's textbook [17, p. 61]. A detailed analysis of this counterexample can be found in [18]. Let $\chi = \hat{\chi} = \{1, 2, 3\}$,

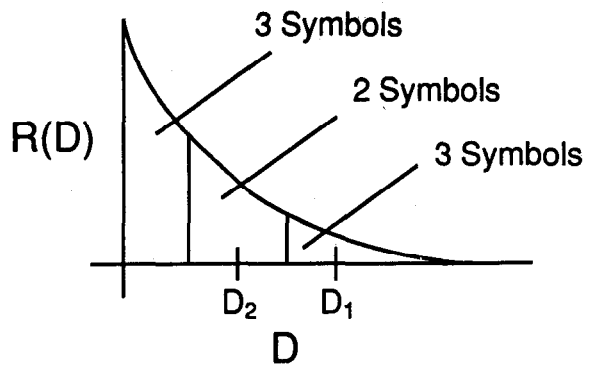$$P_X = \left[\frac{1-p}{2}, p, \frac{1-p}{2} \right], \qquad (31)$$

Fig. 4.   Choice of D_1 and D_2 in counterexample.

with $0 < p < 1$, and $d(x, \hat{x}) = |x - \hat{x}|$. We assume, for this example, that $p < 3 - 2\sqrt{2}$. Let

$$z = e^{R'(D)}, \qquad (32)$$

where $R'(D)$ is the derivative of the rate distortion function at D. Then by the Kuhn Tucker conditions, it can be shown that the solution to the rate distortion problem for distortion D is

$$P_{X|\hat{X}} = \frac{1}{1+z} \begin{bmatrix} 1 & z(1-z) & z^2 \\ z & 1-z & z \\ z^2 & z(1-z) & 1 \end{bmatrix}, \qquad (33)$$

and

$$P_{\hat{X}} = \left[\frac{1-p-z-pz}{2(1-z)^2}, \frac{p-z+pz+z^2}{(1-z)^2}, \frac{1-p-z-pz}{2(1-z)^2} \right], \qquad (34)$$

if $|z - (1/2)(1-p)|^2 \le (1/4)(p^2 - 6p + 1)$; and

$$P_{X|\hat{X}} = \begin{bmatrix} \dfrac{1-p}{1+z^2} & p & \dfrac{(1-p)z^2}{1+z^2} \\ \dfrac{(1-p)z^2}{1+z^2} & p & \dfrac{1-p}{1+z^2} \end{bmatrix}, \qquad (35)$$

and

$$P_{\hat{X}} = [1/2, 0, 1/2], \qquad (36)$$

if $|z - (1/2)(1-p)|^2 > (1/4)(p^2 - 6p + 1)$.

Let D_2 be in the region for which $p(\hat{x}) = (1/2, 0, 1/2)$. Specifically, let $z_2 = e^{R'(D_2)} = (1/2)(1-p)$. Let $D_1 > D_2$ be chosen to lie in the 3-symbol active region, i.e., let $z_1 = e^{R'(D_1)}$ satisfy

$$\frac{1}{2}(1-p) + \sqrt{p^2 - 6p + 1} < z_1 < \frac{1-p}{1+p}. \qquad (37)$$

See Fig. 4.

We shall argue that we cannot find a (necessarily) 3×2 transition matrix $p(\hat{x}_2 | \hat{x}_1)$ such that

$$p(x | \hat{x}_1) = \sum_{\hat{x}_2} p(x | \hat{x}_2) p(\hat{x}_2 | \hat{x}_1). \qquad (38)$$

This is because there is a bottleneck in $\hat{X}_1 \to \hat{X}_2 \to X$, since $\hat{X}_2$ has only two states, thus preventing $p(x | \hat{x}_1)$ from having the degrees of freedom necessary to satisfy (33). We consider

the matrix equation

$$P_{X|\hat{X}_1} = P_{\hat{X}_2|\hat{X}_1} P_{X|\hat{X}_2}, \tag{39}$$

which we rewrite as

$$\frac{1}{1+z_1} \begin{bmatrix} 1 & z_1(1-z_1) & z_1^2 \\ z_1 & 1-z_1 & z_1 \\ z_1^2 & z_1(1-z_1) & 1 \end{bmatrix}$$

$$= \begin{bmatrix} A & 1-A \\ B & 1-B \\ C & 1-C \end{bmatrix} \begin{bmatrix} \dfrac{1-p}{1+z_2^2} & p & \dfrac{(1-p)z_2^2}{1+z_2^2} \\ \dfrac{(1-p)z_2^2}{1+z_2^2} & p & \dfrac{1-p}{1+z_2^2} \end{bmatrix}. \tag{40}$$

Finally, we observe from (40) that $P_{\hat{X}_2|\hat{X}_1} P_{X|\hat{X}_2}$ is of the form

$$\begin{bmatrix} D & E & 1-D-E \\ F & E & 1-F-E \\ G & E & 1-E-G \end{bmatrix}. \tag{41}$$

Note the equal entries in the second column. Thus, by inspecting the left-hand side of (40), we see that $P_{X|\hat{X}_1}$ has the above form only if $z_1(1-z_1) = 1-z_1$, i.e., $z_1 = 1$. But $z_1 = e^{R(D_1)} \le e^{R(D_{\max})} = (1-p)/(1+p) < 1$.

Thus there exists no Markov chain $\hat{X}_1 \to \hat{X}_2 \to X$ satisfying the rate distortion conditional marginals $p(x|\hat{x}_2)$ and $p(x|\hat{x}_1)$ given in (33) and (35). So for $0 < p < 3 - 2\sqrt{2}$, the problem $P_X = [(1-p)/2, p, (1-p)/2]$, $d(x,\hat{x}) = |x - \hat{x}|$ is not successively refinable in general.

We can also characterize exactly when successive refinement is achievable from distortion D_1 to D_2. One interesting case is described in the following theorem, which is true for any $0 < p < 1$. See [18] for a proof.

Theorem 3: Consider the discrete rate-distortion problem with $\chi = \hat{X}_1 = \hat{X}_2 = \{1,2,3\}$, $d(x,\hat{x}) = |x - \hat{x}|$, and

$$X = \begin{cases} 1, & \text{with probability} \quad \dfrac{1-p}{2} \\ 2, & \text{with probability} \quad p \\ 3, & \text{with probability} \quad \dfrac{1-p}{2} \end{cases}.$$

If D_1 and D_2 are such that $\hat{\chi}_{\text{supp}}(D_1) = \hat{\chi}_{\text{supp}}(D_2) = \{1,2,3\}$, then successive refinement from distortions D_1 to D_2 ($D_1 > D_2$) is achievable if and only if

$$\left(1 + e^{R'(D_1)}\right)\left(1 + e^{R'(D_2)}\right) \le 2. \tag{42}$$

Notice that even if $\hat{\chi}_{\text{supp}} = \{1,2,3\}$ for all D (for example, if $p > 3 - 2\sqrt{2}$), we can choose D_1 and D_2 so that $(1 + e^{R'(D_1)})(1 + e^{R'(D_2)}) > 2$ and successive refinement is not achievable. This serves as a counterexample to a possible conjecture that successive refinement is always possible when $\chi = \hat{\chi}$ over the entire rate-distortion curve.

V. EXAMPLES OF SUCCESSIVE REFINEMENT

We now show that the following rate distortion problems are successively refinable.

1) X Gaussian, squared error distortion, $d(x,\hat{x}) = (x - \hat{x})^2$.
2) X arbitrary discrete, Hamming distortion, $d(x,\hat{x}) = 1 - \delta(x - \hat{x})$.

3) X Laplacian, absolute error distortion, $d(x,\hat{x}) = |x - \hat{x}|$.

The details are developed in [18].

A. Gaussian Distribution with Squared-Error Distortion

If X is Gaussian $N(0,\sigma^2)$, then $R(D)$ is achieved by $p(\hat{x}) = N(0, \sigma^2 - D)$, $p(x|\hat{x}) = N(\hat{x}, D)$. It follows from the work of Gray and Wyner [11] that this problem is successively refinable in our context. It is easy to show, for $D_1 > D_2$, that

$$p(\hat{x}_1) = N(0, \sigma^2 - D_1) \tag{43}$$

$$p(\hat{x}_2|\hat{x}_1) = N(\hat{x}_1, D_1 - D_2) \tag{44}$$

$$p(x|\hat{x}_2) = N(\hat{x}_2, D_2) \tag{45}$$

yields a joint density $p(x, \hat{x}_1, \hat{x}_2) = p(\hat{x}_1)p(\hat{x}_2|\hat{x}_1)p(x|\hat{x}_2)$ having the desired marginal $p(x) = N(0,\sigma^2)$ and satisfying the conditions of Theorem 2, thus guaranteeing the achievability of

$$(R_1, R_2) = (R(D_1), R(D_2)) = \left(\frac{1}{2}\log\frac{\sigma^2}{D_1}, \frac{1}{2}\log\frac{\sigma^2}{D_2}\right). \tag{46}$$

The code achieving these bounds has an especially nice tree structure. Let $2^{nR(D_1)} = (\sigma^2/D_1)^{n/2}$ $\hat{x}_1$'s be drawn i.i.d. $\sim N_n(0, (\sigma^2 - D_1)I_n)$. Label them $\hat{x}_1(1), \hat{x}_1(2), \cdots, \hat{x}_1(2^{nR(D_1)})$. Let $2^{n(R(D_2)-R(D_1))} = (D_1/D_2)^{n/2}$ u's be drawn i.i.d. $\sim N_n(0, (D_1 - D_2)I_n)$. Label them $u(1), u(2), \cdots, u(2^{n(R(D_2)-R(D_1))})$. Then, given $x \sim N_n(0, \sigma^2 I_n)$, let $i(x)$ denote the index i minimizing $\|\hat{x}_1(i) - x\|^2$, and let $j(x)$ denote the index j minimizing $\|\hat{x}_1(i(x)) + u(j) - x\|^2$. Then, the reconstruction $\hat{x}_1 = \hat{x}_1(i(x))$ and $\hat{x}_2 = \hat{x}_1(i(x)) + u(j(x))$ asymptotically achieves distortions D_1 and D_2 at rates R_1 and R_2, respectively. Note that it takes only $2^{nR(D_1)} + 2^{n(R(D_2)-R(D_1))}$ distance calculations to encode and decode x. This number of calculations is exponentially smaller than the $2^{nR(D_2)}$ calculations required to describe x at distortion D_2 in one step.

B. Arbitrary Discrete Distribution with Hamming Distortion

We now consider Example 2. Here $X \sim p(x)$, $x \in \{1,2,\cdots,m\}$, and $D = Ed(X, \hat{X}) = \Pr\{X \ne \hat{X}\}$. This is a probability of error distortion measure for an arbitrary discrete source. It has been shown by Erokhin [19] and Pinkston [20] that $R(D)$ is achieved by upside down waterfilling. Specifically,

$$p(\hat{x}) = \frac{(p(x) - \lambda)^+}{\sum_x (p(x) - \lambda)^+}, \tag{47}$$

and

$$p(x|\hat{x}) = \begin{cases} \overline{D}, & x = \hat{x} \\ \lambda, & x \ne \hat{x}, x \in \hat{X}_{\text{supp}}, \\ p_k, & x = k \notin \hat{X}_{\text{supp}} \end{cases} \tag{48}$$

where λ is chosen so that $\sum_{\hat{x}} p(\hat{x})p(x|\hat{x}) = p(x)$, and $\hat{X}_{\text{supp}} = \{x: p(x) - \lambda > 0\}$.

Let $D_1 > D_2$. The required transition $p(\hat{x}_2|\hat{x}_1)$ to establish $\hat{X}_1 \to \hat{X}_2 \to X$ (and thus $X \to \hat{X}_2 \to \hat{X}_1$) is

$$p(\hat{x}_2|\hat{x}_1) = \begin{cases} \dfrac{\overline{D}_1 - \lambda_2}{\overline{D}_2 - \lambda_2}, & \hat{x}_2 = \hat{x}_1 \\[2mm] \dfrac{\lambda_1 - \lambda_2}{\overline{D}_2 - \lambda_2}, & \hat{x}_2 \neq \hat{x}_1, \hat{x}_2 \in \hat{X}_{1\text{supp}}, \\[2mm] \dfrac{p_k - \lambda_2}{\overline{D}_2 - \lambda_2}, & \hat{x}_2 = k \notin \hat{X}_{1\text{supp}} \end{cases} \quad (49)$$

where $\hat{X}_{1\text{supp}}$ is the support set of $\hat{X}_1$, and λ_1, λ_2 are the "waterfilling" levels. Thus

$$p(\hat{x}_i) = \frac{(p(x) - \lambda_i)^+}{\sum_x (p(x) - \lambda_i)^+}, \quad (50)$$

for $i = 1, 2$, and $p(\hat{x}_2|\hat{x}_1)$, as given in (49), achieves $(R(D_1), R(D_2))$.

C. Laplacian Density with Absolute Error

We now show that random variables drawn from a Laplacian distribution can be successively refined when distortion is measured using the absolute distortion criterion. We say that a random variable X is drawn from a Laplacian distribution if it has a density f (parameterized by α) such that $f(x) = (\alpha/2)e^{-\alpha|x|}$. We assume the absolute distortion measure $d(x, \hat{x}) = |x - \hat{x}|$.

We first recall the rate-distortion solution. Here $R(D)$ and $f(\hat{x}|x)$ are given by

$$R(D) = -\log(\alpha D), \quad 0 < D \leq D_{\max} = 1/\alpha, \quad (51)$$

and

$$f_{X|\hat{X}}(x|\hat{x}) = g(x - \hat{x}) \quad (52)$$

$$= \frac{1}{2D}e^{-|x-\hat{x}|/D}. \quad (53)$$

We wish to show for $D_1 \geq D_2$ that $\hat{X}_1 \to \hat{X}_2 \to X$ can form a Markov chain by finding $f_{\hat{X}_2|\hat{X}_1}(\hat{x}_2|\hat{x}_1)$ such that

$$f_{X|\hat{X}}(x|\hat{x}_1) = \int_{\hat{x}_2} f_{X|\hat{X}}(x|\hat{x}_2) f_{\hat{X}_2|\hat{X}_1}(\hat{x}_2|\hat{x}_1), \quad (54)$$

or

$$g_1(x - \hat{x}_1) = \int_{\hat{x}_2} g_2(x - \hat{x}_2) f_{\hat{X}_2|\hat{X}_1}(\hat{x}_2|\hat{x}_1), \quad (55)$$

where

$$g_i(t) = \frac{1}{2D_i}e^{-|t|/D_i}. \quad (56)$$

The characteristic function Φ_{g_i} of g_i is given by

$$\Phi_{g_i} = \frac{(1/D_i)^2}{(1/D_i)^2 + \omega^2}. \quad (57)$$

The question is whether there is a distribution h with a characteristic function Φ_h satisfying $\Phi_{g_1}(\omega) = \Phi_{g_2}(\omega)\Phi_h(\omega)$.

Thus

$$\Phi_h(\omega) = \frac{\Phi_{g_1}(\omega)}{\Phi_{g_2}(\omega)} \quad (58)$$

$$= \epsilon^{j\omega\hat{x}_1}\left(\frac{D_2}{D_1}\right)^2\left(\frac{\omega^2 + (1/D_2)^2}{\omega^2 + (1/D_1)^2}\right) \quad (59)$$

$$= e^{j\omega\hat{x}_1}\left(\frac{D_2}{D_1}\right)^2\left[1 + \left(\left(\frac{D_1}{D_2}\right)^2 - 1\right)\left(\frac{(1/D_1)^2}{\omega^2 + (1/D_1)^2}\right)\right]. \quad (60)$$

Taking the inverse transform, we obtain

$$h_{\hat{X}_2|\hat{X}_1}(\hat{x}_2|\hat{x}_1) = \frac{1}{2D_1}e^{-|\hat{x}_2 - \hat{x}_1|/D_1}$$

$$+ \left(\frac{D_2}{D_1}\right)^2\left[\delta(\hat{x}_2 - \hat{x}_1) - \frac{1}{2D_1}e^{-|\hat{x}_2 - \hat{x}_1|/D_1}\right]. \quad (61)$$

This is nonnegative and integrates to one, so we have found the conditional density establishing that $\hat{X}_1 \to \hat{X}_2 \to X$ can be written as a Markov chain. Therefore successive refinement is achievable.

VI. Conclusion

Successively refinable source coding problems have simple descriptions that can be stopped at any point without loss of optimality. This is only possible if the conditional distributions $p(\hat{x}_i|x)$ can be written as a Markov chain.

References

[1] J. Max, "Quantizing for minimum distortion," *IRE Trans. Inform. Theory*, vol. IT-6, no. 1, pp. 7–12, Mar. 1960.
[2] H. S. Witsenhausen and A. D. Wyner, "Source coding for multiple descriptions II: A binary source. *Bell Syst. Tech. J.*, vol. 60, no. 10, pp. 2281–2292, Dec. 1981.
[3] J. K. Wolf, A. D. Wyner, and J. Ziv, "Source coding for multiple descriptions," *Bell Syst. Tech. J.*, vol. 59, no. 8, pp. 1417–1426, Oct. 1980.
[4] L. Ozarow, "On a source-coding problem with two channels and three receivers," *Bell Syst. Tech. J.*, vol. 59, no. 10, pp. 1909–1921, Dec. 1980.
[5] A. El Gamal and T. M. Cover, "Achievable rates for multiple descriptions," *IEEE Trans. Inform. Theory*, vol. IT-28, no. 6, pp. 851–857, Nov. 1982.
[6] Z. Zhang and T. Berger, "New results in binary multiple descriptions," *IEEE Trans. Inform. Theory*, vol. IT-33, no. 4, pp. 502–521, July 1987.
[7] R. Ahlswede, "The rate-distortion region for multiple descriptions without excess rate," *IEEE Trans. Inform. Theory*, vol. IT-31, no. 6, pp. 721–726, Nov. 1985.

[8] R. M. Gray, "A new class of lower bounds to information rates of stationary sources via conditional rate-distortion functions," *IEEE Trans. Inform. Theory*, vol. IT-19, no. 4, pp. 480–489, July 1973.

[9] ___, "Conditional rate-distortion theory," Technical Report 6502-2, Stanford Electronics Laboratory, Stanford Univ., Stanford, CA, Oct. 1972.

[10] ___, "Source coding a binary symmetric source over a simple network," in *Proc. Hawaii Int. Conf. Syst. Sci.*, no. 6, Jan. 1973, pp. 354–355.

[11] R. M. Gray and A. D. Wyner, "Source coding for a simple network," *Bell Syst. Tech. J.*, vol. 53, no. 9, pp. 1681–1721, Nov. 1974.

[12] H. Yamamoto, "Source coding theory for cascade and branching communication systems," *IEEE Trans. Inform. Theory*, vol. IT-27, no. 3, pp. 299–308, May 1981.

[13] A. H. Kaspi and T. Berger, "Rate-distortion for correlated sources with partially separated encoders," *IEEE Trans. Inform. Theory*, vol. IT-28, no. 6, pp. 828–840, Nov. 1982.

[14] P. A. Chou, T. Lookabaugh, and R. M. Gray, "Optimal prun-

ing with applications to tree-structured source coding and modeling," *IEEE Trans. Inform. Theory*, vol. 35, no. 2, pp. 299–315, 1989.

[15] E. A. Riskin, T. Lookabaugh, P. A. Chou, and R. M. Gray, "Variable rate vector quantization for medical image compression," *IEEE Trans. Med. Imaging*, vol. 9, no. 3, pp. 290–298, Sept. 1990.

[16] A. M. Gerrish, "Estimation of information rates," Ph.D. thesis, Yale Univ., New Haven, CT, Dept. of Elect. Eng., 1963.

[17] T. Berger, *Rate Distortion Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1971.

[18] W. H. Equitz, "Successive refinement of information," Ph.D. thesis, Stanford Univ., Dept. of Elect. Eng., 1989.

[19] V. Erokhin, "Epsilon-entropy of a discrete random variable," *Theory of Probability and Applications*, vol. 3, pp. 97–100, 1958, English translation.

[20] J. T. Pinkston, "An application of rate-distortion theory to a converse to the coding theorem," *IEEE Trans. Inform. Theory*, vol. IT-15, no. 1, pp. 66–71, January 1969.