Because the last expression is independent of \( k \) or \( \tau^* \), its negative logarithm is a lower bound on \( I(\mu_m) \), that is
\[
I(\mu_m) \geq -E_{\tau^*} \log E e^{-n/4H^2(\tau^*)} = -E_{\tau^*} \log E e^{-n/4C_{\mu_m}W}
\]
\[
= -\log(2^{-m}(1 + e^{-m}))^m = m \log 2 - m \log (1 + e^{-m}) = m \log 2 - m \log (1 + e^{-m/2G_{\mu}}) \geq m \log 2 - m \log (1 + e^{-2B_{\mu}m^{-2s-1}}).
\]

Choosing \( m = An^{1/(2s+1)} \) \((A > 0)\) to maximize the rate in the above lower bound, we get the following theorem.

**Theorem:**
\[
\min_{q_n} \max_{f \in LIP(s, C)} E_{f^n} \log (f^n / q_n) \geq I(\mu_m) \geq A_{p, s} n^{1/(2s+1)}
\]
where
\[
A_{p, s} = A(\log 2 - \log (1 + e^{-2B_{\mu}(4s+1)})) = A(\log 2 - \log (1 + e^{-p/2G_{\mu}})) > 0.
\]

Taking \( k = 0, \nu = 1 \) therefore \( s = 1 \) in the theorem, we obtain the optimal rate lower bound in [13], as shown in the corollary below.

**Corollary:**
\[
\min_{q_n} \max_{f \in LIP(1, C)} E_{f^n} \log (f^n / q_n) \geq O(n^{1/3}).
\]

**Remarks:**
1. In general, we can consider the \( LIP(s, C) \) classes on \([0, 1]^d \) \((d \geq 1)\). Minimax lower bounds on redundancy of rates \( O(n^{d/(2s+5)}) \) can be obtained. These rates are believed to be optimal in the sense that universal codes can be constructed to achieve these rates. In the case of \( LIP(1, C) \) the rate \( n^{-2} \) has been shown to be optimal in [13].
2. The proof for the minimax lower bound \( \frac{1}{2} \log n \) in the parametric case follows from the asymptotic expansion of \( I(\mu) \) in [1] or [8] for smooth priors. Superficially, this approach has a continuous flavor since \( \mu \) needs to have nice smoothness properties on the whole parameter space, whereas the proof in the nonparametric case as we just saw has a discrete flavor because of the hypercube subclass over which \( I(\mu_m) \) is estimated. Heuristically, however, the continuous prior can be made discrete. Under regularity conditions, we believe that \( I(\mu) \) should give the same lower bound \( \frac{1}{2} \log n \) even for a discrete uniform prior \( \mu \) on a grid subset of the parametric space, as long as the grid size is of the order or smaller than \( n^{-1/2} \). Note that the nearest neighbors on the hypercube for the optimal choice \( m = n^{1/(2s+1)} \) also have Hellinger distances of order \( n^{-1/2} \), the rate at which \( n \) i.i.d. data points can possibly distinguish two probability densities. In other words, what seems essential to both the parametric and the nonparametric case is to find a subclass of densities whose closest elements are \( n^{-1/2} \) apart from each other in terms of Hellinger distance.

**Acknowledgment**

The author wishes to thank Dr. J. Rissanen, an anonymous referee, the Associate Editor, and J. Huang for their very helpful comments. Special thanks are due to Dr. N. Hengartner for spotting an inexcusable error in an earlier version of the correspondence and for pointing to the Haussler paper.

**Error Exponents for Successive Refinement by Partitioning**

Angelos Kanlis and Prakash Narayan

**References**

I. INTRODUCTION

The problem of successive refinement of information by partitioning, also referred to variously as “hierarchical lossy data compression” and “sequential approximations,” has received much attention over the years (Koshlev [9], [10], Equitz-Cover [4], Yamamoto [19], Rimoldi [15]). Related problems include those of “multiple descriptions” (cf. [1], [3], [14], [17], [18], [20]) and that of determining the achievable rate region for cascade communication systems [19]. Given a discrete memoryless source (DMS) with probability mass function (pmf) \( P \), and a suitable distortion measure, suppose that we first seek to describe the source with distortion not exceeding \( \Delta_1 \). The (asymptotically) minimum rate of coding is, of course, given by the rate distortion function \( R(P, \Delta_1) \). Subsequently, if a better (finer) description is required, say with distortion \( \Delta_2 < \Delta_1 \), additional information at rate \( \Delta R \) can be provided, so that the overall augmented rate is \( R(P, \Delta_1) + \Delta R \). Clearly, \( R(P, \Delta_1) + \Delta R \geq R(P, \Delta_2) \). It is of interest to determine the rate-distortion function \( R(P, \Delta_1, \Delta_2) \) for this two-step coding process, and find conditions under which it coincides with \( R(P, \Delta_2) \).

The condition under which these two rate-distortion functions coincide was determined independently by Koshlev [9] and Equitz-Cover [4], and subsequently given a geometrical interpretation by Rimoldi [15]. This condition requires that the source random variable (rv) and the two reproduction rv’s satisfy a Markov property. Rimoldi [15] also provided a complete characterization of the achievable rate region for two-step coding.

In this correspondence, we determine the error exponents for the two-step coding process. It is then shown that even under the Markov condition, when the two rate-distortion functions coincide, the performance of the two-step coding process—as measured by its error exponents—may be inferior to that of one-step coding.

II. PRELIMINARIES

Let \( \mathcal{X} \) be a finite set. Let \( \{ X_i \}_{i=1}^{\infty} \) be a \( \mathcal{X} \)-valued discrete memoryless source (DMS), i.e., an independent and identically distributed (i.i.d.) process, with (common) probability mass function (pmf) \( P \). Let \( \mathcal{Y}_i \) be a finite reproduction alphabet. Let \( d_1 : \mathcal{X} \times \mathcal{Y}_1 \to \mathbb{R}^+ \) be a nonnegative-valued mapping with \( \min_{x \in \mathcal{X}} d_1(x, y) = 0 \). This mapping induces a distortion measure on \( \mathcal{X}^n \times \mathcal{Y}_1^n \) according to

\[
    d_1(x, y) = \frac{1}{n} \sum_{i=1}^{n} d_1(x_i, y_i), \quad x \in \mathcal{X}^n, \ y \in \mathcal{Y}_1^n.
\]

An \( n \)-length block code consists of two mappings: An encoder

\[
    f_1^{(n)} : \mathcal{X}^n \to \mathcal{M}_1 = \{ 1, \ldots, M_1 \}
\]

and a decoder

\[
    \phi_1^{(n)} : \mathcal{M}_1 \to \mathcal{Y}_1^n.
\]

The rate of this code is \( R_1 = \frac{1}{n} \log M_1 \). All logarithms and exponentials are with respect to the base 2.

For \( R_1 > 0, \Delta_1 > 0 \), we say that the pair \( (R_1, \Delta_1) \) is achievable if for every \( \epsilon > 0, \delta > 0 \) and \( n \) sufficiently large, there exists an \( n \)-length block code \( (f_1^{(n)}, \phi_1^{(n)}) \) of rate not exceeding \( R_1 + \delta \) such that

\[
    \Pr \left\{ d_1(X^n, \phi_1^{(n)}(f_1^{(n)}(X^n))) \leq \Delta_1 \right\} \geq 1 - \epsilon.
\]

The corresponding rate-distortion function, \( R(P, \Delta_1) \), characterizing the minimum achievable rate for a distortion \( \Delta_1 \), is well known and given by

\[
    R(P, \Delta_1) = \inf_{P_X=P, \mathcal{Y}_1(X,Y_1) \leq \Delta_1} I(X \land Y_1)
\]

where \( E \) denotes expectation.

Let \( \mathcal{Y}_2 \) be a (refining) finite reproduction alphabet. A refined description of the source \( \{ X_i \}_{i=1}^{\infty} \) can be provided by means of a \( n \)-length refining block code \( (f_2^{(n)}, \phi_2^{(n)}) \), specified by an encoder

\[
    f_2^{(n)} : \mathcal{X}^n \to \mathcal{M}_2 = \{ 1, \ldots, M_2 \}
\]

and a decoder

\[
    \phi_2^{(n)} : \mathcal{M}_1 \times \mathcal{M}_2 \to \mathcal{Y}_2^n.
\]

The rate of the refining code \( (f_2^{(n)}, \phi_2^{(n)}) \) is defined as

\[
    R_2 = \frac{1}{n} \log M_1 M_2.
\]

Let \( d_2 : \mathcal{X} \times \mathcal{Y}_2 \to \mathbb{R}^+ \) be a nonnegative-valued mapping with \( \min_{x \in \mathcal{X}} d_2(x, y) = 0 \), which induces a distortion measure on \( \mathcal{X}^n \times \mathcal{Y}_2^n \) according to

\[
    d_2(x, y) = \frac{1}{n} \sum_{i=1}^{n} d_2(x_i, y_i), \quad x \in \mathcal{X}^n, \ y \in \mathcal{Y}_2^n.
\]

Definition 1 (Rimoldi [15]): For the DMS \( \{ X_i \}_{i=1}^{\infty} \) with pmf \( P \) and distortion measures \( d_1, d_2 \), the quadruple \( (R_1, R_2, \Delta_1, \Delta_2) \), \( R_1 > 0, R_2 > 0, \Delta_1 > 0, \Delta_2 > 0 \), is achievable if for every \( \epsilon > 0, \delta > 0 \) and \( n \) sufficiently large, there exist

- an \( n \)-length code \( (f_1^{(n)}, \phi_1^{(n)}) \) such that

\[
    \frac{1}{n} \log M_1 \leq R_1 + \delta.
\]

and

\[
    \Pr \left\{ d_1(X^n, \phi_1^{(n)}(f_1^{(n)}(X^n))) \leq \Delta_1 \right\} \geq 1 - \epsilon
\]

- an \( n \)-length refining code \( (f_2^{(n)}, \phi_2^{(n)}) \) such that

\[
    \frac{1}{n} \log M_1 M_2 \leq R_2 + \delta
\]

and

\[
    \Pr \left\{ d_2(X^n, \phi_2^{(n)}(f_1^{(n)}(X^n), f_2^{(n)}(X^n))) \leq \Delta_2 \right\} \geq 1 - \epsilon.
\]

A characterization of the set of achievable quadruples \( (R_1, R_2, \Delta_1, \Delta_2) \) has been provided independently by Koshlev [10] and Rimoldi [15] and also follows as a combination of the results of El Gamal and Cover [3] and Yamamoto [19].

Theorem 1 ([3], [10], [15], [19]): Consider a DMS \( \{ X_i \}_{i=1}^{\infty} \) with pmf \( P \) and distortion measures \( d_1, d_2 \). The quadruple \( (R_1, R_2, \Delta_1, \Delta_2) \) is achievable iff there exists a pmf \( P_{X,Y_1,Y_2} \) on \( X \times \mathcal{Y}_1 \times \mathcal{Y}_2 \) with marginal \( P \) on \( X \) satisfying the following inequalities:

\[
    I(X \land Y_1) \leq R_1, \quad I(X \land Y_1 Y_2) \leq R_2 \quad \text{Ed}_{d_1}(X, Y_1) \leq \Delta_1, \quad \text{Ed}_{d_2}(X, Y_1 Y_2) \leq \Delta_2
\]

where \( (X, Y_1, Y_2) \) is a \( X \times \mathcal{Y}_1 \times \mathcal{Y}_2 \)-valued rv with pmf \( P_{X,Y_1,Y_2} \).

Suppose that the coding in the first step were done at rate \( R_1 > R(P, \Delta_1) \). If the DMS is now to be described in a second step at distortion \( \Delta_2 < \Delta_1 \), let \( R(P, R_1, \Delta_1, \Delta_2) \) be the minimum rate \( R_2 \) of the refining code such that \( (R_1, R_2, \Delta_1, \Delta_2) \) is achievable. This minimal rate \( R(P, R_1, \Delta_1, \Delta_2) \) is characterized by the following Corollary to Theorem 1.
Corollary 1: For $\Delta_1 > 0$, $\Delta_2 > 0$, and $R_1 > R(P, \Delta_1)$, we have
\[
R(P, R_1, \Delta_1, \Delta_2) = \inf_{P_{XY} \in P} I(X \land Y_1 Y_2),
\]
where the conditional rate-distortion function $R_{X|Y_1}(\Delta_2)$ [7] is given by
\[
R_{X|Y_1}(\Delta_2) = \inf_{P_{XY} \in P_{X|Y_1}} \sum_{x \in \mathcal{X}} P_{X|Y_1}(x) R_{X|Y_1}(\Delta_2) P_{Y_1}(y_1) P_{Y_2}(y_2 | \Delta_2)
\]
\[
I(X \land Y_2 | Y).
\]
Suppose next that the coding in the first step at distortion $\Delta_1$ were done optimally, i.e., $R_1 \approx R(P, \Delta_1)$. If the DMS is now to be described in a second step at distortion $\Delta_2 < \Delta_1$, what is the smallest possible amount of additional information required for this purpose? The following definition is a special case of Definition 1 for the situation wherein the first step is optimal.

Definition 2: For the DMS $\{X_t\}_{t=1}^{\infty}$ with pmf $P$, the rate $R_2$ is $(\Delta_1, \Delta_2)$-refinement-achievable if for every $\epsilon, \delta > 0$ and $n$ sufficiently large, there exist
\begin{itemize}
  \item an $n$-length code $(f_1^{(n)}, \phi_1^{(n)})$ such that
    \[
    \frac{1}{n} \log M_1 \leq R(P, \Delta_1) + \delta
    \]
    and
    \[
    \Pr \left\{ d_1(X^n, \phi_1^{(n)}(f_1^{(n)}(X^n))) \leq \Delta_1 \right\} \geq 1 - \epsilon
    \]
  \item an $n$-length refining code $(f_2^{(n)}, \phi_2^{(n)})$ such that
    \[
    \frac{1}{n} \log M_2 \leq R_2 + \delta
    \]
    and
    \[
    \Pr \left\{ d_2(X^n, \phi_2^{(n)}(f_2^{(n)}(X^n))) \leq \Delta_1, d_2(X^n, \phi_2^{(n)}(f_2^{(n)}(X^n))) \leq \Delta_2 \right\} \geq 1 - \epsilon
    \]
\end{itemize}

Let $R(P, \Delta_1, \Delta_2)$ denote the infimum of the set of $(\Delta_1, \Delta_2)$-refinement-achievable rates. It constitutes the rate-distortion function for the refining code and is given by the following Corollary to Theorem 1.

Corollary 2: For $\Delta_1 > 0$, $\Delta_2 > 0$, we have
\[
R(P, \Delta_1, \Delta_2) = R(P, R(P, \Delta_1), \Delta_1, \Delta_2) = \inf_{P_{XY} \in P} I(X \land Y_1 Y_2),
\]
where $P_{XY}$ is a, i.i.d.

Remarks:
1) For $d_1 = d_2$, $\Delta_1 = \Delta_2$, and $Y_1 = Y_2$, we have that $R(P, \Delta_1, \Delta_2) = R(P, \Delta_1)$, the minimum achievable rate for one-step coding.
2) For two-step coding with $\Delta_2 < \Delta_1$, and $R(P, \Delta_2) > R_1 > R(P, \Delta_1)$, clearly
\[
R(P, \Delta_2) \leq R(P, R_1, \Delta_1, \Delta_2) \leq R(P, \Delta_1, \Delta_2).
\]
Koshkev [9] has provided a sufficient condition for the inequalities above to hold as equalities. Cover and Equitz [4] have independently shown this condition to be both necessary and sufficient (see Theorem 2 below).

3) It follows from the observation of Equitz and Cover [4, p. 271] in the context of Gray's work [7] on conditional rate-distortion function that if $\{Y_t\}_{t=1}^{\infty}$ is an i.i.d.

III. THE ERROR EXPONENTS

In this section, we shall characterize the error exponents for successive refinement by partitioning. Corollaries 2 and 1 imply that for $\Delta_2 \leq \Delta_1$ and for numbers $R_1, R_2$ with $R_1 < R_2$ and $R_1 > R(P, \Delta_1, \Delta_2)$, there exists a sequence of $n$-length block codes $(f_1^{(n)}, \phi_1^{(n)})$, $(f_2^{(n)}, \phi_2^{(n)})$, such that
\[
\lim \frac{1}{n} \log \|f_1^{(n)}\| = R_1
\]
\[
\lim \frac{1}{n} \log \|f_2^{(n)}\| = R_2
\]
and
\[
\lim \Pr (d_1(X^n, \phi_1(f_1(X^n))) > \Delta_1)
\]
\[
d_2(X^n, \phi_2(f_1(X^n), f_2(X^n))) > \Delta_2 = 0
\]
where $\|f_1^{(n)}\|$ (resp., $\|f_2^{(n)}\|$) denotes the cardinality of the domain of $f_1^{(n)}$ (resp., $f_2^{(n)}$). Our objective is to characterize the rate of convergence to zero of the previous probability.

Definition 3: For given distortion measures $d_1, d_2$, positive numbers $\Delta_1 > \Delta_2 > 0$, $n$-length block code $(f_1, \phi_1)$, and $n$-length refining block code $(f_2, \phi_2)$, we define the error function $e(P, (f_1, \phi_1), (f_2, \phi_2), \Delta_1, \Delta_2)$ as the probability that the source sequence $X^n$ in $\mathcal{X}^n$ of the DMS with (common) pmf $P$ is not reproduced within distortion $\Delta_1$ in the first step of coding, or within distortion $\Delta_2$ by the refining code. Thus the error function for two-step coding is defined as
\[
e(P, (f_1, \phi_1), (f_2, \phi_2), \Delta_1, \Delta_2)
\]
\[
\lim \Pr (d_1(X^n, \phi_1(f_1(X^n))) > \Delta_1)
\]
\[
d_2(X^n, \phi_2(f_1(X^n), f_2(X^n))) > \Delta_2 = 0
\]
Definition 4: The conditional error function $e(P, (f_1, f_2), (\phi_2, \phi_3), \Delta_2 \mid R_1, \Delta_1)$ is the probability that the source sequence $X^n$ in $\mathcal{X}^n$ is not reproduced within distortion $\Delta_2$ by the refining code, given that $(f_1, f_2)$, i.e., the code for the first step, has rate that does not exceed $R_1$ and distortion that does not exceed $\Delta_1$. Thus

$$e(P, (f_1, f_2), (\phi_2, \phi_3), \Delta_2 \mid R_1, \Delta_1) = \Pr(d(X^n, \phi_2 f_1(X^n), \phi_3 f_2(X^n)) > \Delta_2). \quad (7)$$

We show below for suitable two-step $n$-length block codes with rates converging to $R_1$ and $R_2$, that $e(P, (f_1^{(n)}, \phi_2^{(n)}), (f_2^{(n)}, \phi_3^{(n)}), \Delta_1, \Delta_2)$ converges to zero exponentially rapidly with rate given by the error exponent

$$F(P, R_1, R_2, \Delta_1, \Delta_2) \triangleq \inf_{Q \in \mathcal{Q}(Q,R_1,\Delta_1,R_2,\Delta_2)} D(Q \| P) \quad (8)$$

provided $R_1 > R(P, \Delta_1)$ and $R_2 > R(P, R_1, \Delta_1, \Delta_2)$. This provides an extension of the result of Marton [11] (cf. also Csiszár and Körner [2]) on the error exponent for one-step coding. Our approach is along the lines of Csiszár and Körner [2, sec. 2.4].

It follows from [6] that upon setting $d_1 = d_2, \Delta_1 = \Delta_2, Y_1 = Y_2$ and $\phi_2(m_1, m_2) = \phi_1(m_1), m_1 \in \mathcal{X}_1, m_2 \in \mathcal{M}_2$ that

$$e(P, (f_1, \phi_2), (f_2, \phi_3), \Delta_1, \Delta_2) = e(P, (f_1, \phi_1), \Delta_1)$$

where the term on the right side above corresponds to the error function for one-step coding which converges to zero exponentially rapidly with rate given by the error exponent

$$F(P, R_1, \Delta_1) \triangleq \inf_{Q \in \mathcal{Q}(Q,R_1,\Delta_1)} D(Q \| P) \quad (9)$$

provided $R_1 > R(P, \Delta_1)$ (cf. Marton [11], Csiszár and Körner [2]).

We further show that for suitable two-step $n$-length block codes with rates converging to $R_1$, $R_2$, that $e(P, (f_1, f_2), (\phi_2, \phi_3), \Delta_1, \Delta_2)$, i.e., the conditional error function converges to zero exponentially rapidly with rate given by the error exponent

$$F_c(P, R_1, R_2, \Delta_1, \Delta_2) \triangleq \inf_{Q \in \mathcal{Q}(Q,R_1,R_2,\Delta_1,\Delta_2)} D(Q \| P). \quad (10)$$

Theorem 3 (Two-Step Coding Error Exponent): Let $(X^n, Y^n)_n$ be a DMS with pmf $P$. For every $R_1 < R_2 < \log |\mathcal{Y}|$ and distortion measures $d_1$ on $\mathcal{X} \times \mathcal{Y}_1$ and $d_2$ on $\mathcal{X} \times \mathcal{Y}_2$, there exists a sequence of $n$-length block codes for two-step coding such that

- $\lim_{n \to \infty} \frac{1}{n} \log \| f_1^{(n)} \| = R_1$
- $\lim_{n \to \infty} \frac{1}{n} \log \| f_2^{(n)} \| = R_2$

for every pmf $P$ on $\mathcal{X}$, $\Delta_1 > \Delta_2 \geq 0$, and $\delta_1 > 0, \delta_2 > 0$.

$$\frac{1}{n} \log e(P, (f_1^{(n)}, \phi_1^{(n)}), \Delta_1) \leq -F(P, R_1, \Delta_1) + \delta_1$$

and

$$\frac{1}{n} \log e(P, (f_2^{(n)}, \phi_2^{(n)}), \Delta_2 \mid R_1, \Delta_1) \leq -F_c(P, R_2, \Delta_2 \mid R_1, \Delta_1) + \delta_2$$

whenever $n \geq N(|\mathcal{X}|, d_1, d_2, \delta_1, \delta_2, \delta_3)$.

Further, for every sequence of codes satisfying (10) and every distribution $P$ on $\mathcal{X}$, the following hold:

$$\lim_{n \to \infty} \frac{1}{n} \log e(P, (f_1^{(n)}, \phi_1^{(n)}), \Delta_1) \geq -F(P, R_1, \Delta_1)$$

$$\lim_{n \to \infty} \frac{1}{n} \log e(P, (f_2^{(n)}, \phi_2^{(n)}), (f_2^{(n)}, \phi_2^{(n)}), \Delta_1, \Delta_2) \geq -F(P, R_1, R_2, \Delta_1, \Delta_2)$$

and

$$\lim_{n \to \infty} \frac{1}{n} \log e(P, (f_1^{(n)}, \phi_1^{(n)}), (f_2^{(n)}, \phi_2^{(n)}), \Delta_2 \mid R_1, \Delta_1) \geq -F_c(P, R_2, \Delta_2 \mid R_1, \Delta_1).$$

Remark: Note that the two-step code in the statement of the forward part of Theorem 3 does not rely on a knowledge of the pmf $P$ of the DMS. Hence, this code is universal in that it is applicable to any $\mathcal{X}$-valued DMS.

As is to be expected, the error exponent for two-step coding cannot exceed that for one-step coding. This is obvious from (8) by observing for $d_1 = d_2 = d$ and $Y_1 = Y_2 = Y$, that

$$F(P, R_1, R_2, \Delta_1, \Delta_2) = \min \{ F(P, R_1, \Delta_1), F_c(P, R_2, \Delta_2 \mid R_1, \Delta_1) \}. \quad (11)$$

Even with the Markov condition in effect so that the rate-distortion functions for one- and two-step coding coincide, i.e.,

$$R(P, R_1, \Delta_1, \Delta_2) = R(P, \Delta_2), \quad \Delta_2 < \Delta_1$$

note that if

$$F(P, R_1, \Delta_1) < F_c(P, R_2, \Delta_2 \mid R_1, \Delta_1) \quad (12)$$

then

$$F(P, R_1, R_2, \Delta_1, \Delta_2) < F(P, R_2, \Delta_2). \quad (13)$$

This is illustrated in the following example.

Example: Let $(X^n, Y^n)_n$ be a DMS with pmf $P$, where $P(x) \geq 0$, $x \in \mathcal{X}$, $Y_1 = Y_2 = Y$, and $d_1 = d_2 = d$ where $d$ denotes Hamming distortion measure. Thus

$$d(x, y) = \frac{1}{n} \sum_{t=1}^{n} 1(x_t \neq y_t), \quad x \in \mathcal{X}^n, y \in \mathcal{Y}^n.$$
Now, the necessary and sufficient condition for (12) and (13) to hold can be expressed as $Q_{c(R_1, \Delta_1)} \supset Q_{c(R_2, \Delta_2)}$, or, equivalently
\[ c(R_1, \Delta_1) < c(R_2, \Delta_2) \]
which is the same as
\[ R(P, \Delta_1) < R_1 - R(P, \Delta_2) - R(P, \Delta_1) \].

The condition above says that, with the Markov condition in effect, the error exponent (15) for two-step coding is worse than that for one-step coding iff
\[ R_2 - R(P, \Delta_2) > R_1 - R(P, \Delta_1). \tag{16} \]
Note that this condition is given by (14) and (15) as $D(Q_2 \parallel P) < D(Q_1 \parallel P)$, where
\[ D(Q_i \parallel P) \triangleq \min_{Q_i \in \mathcal{Q}(\mathcal{X} \parallel \mathcal{A}_i)} D(Q_i \parallel P), \quad i = 1, 2. \]

This means that in the case of the Hamming distortion measures (divergence) distances in the space of pmf's on $\mathcal{X}$ correspond to distances (differences) between the actual rates and the corresponding values of the rate-distortion functions.

Clearly, if $R_1, R_2$ are chosen to violate inequality (16), e.g., with $R_1$ large enough, the two-step coding process will no longer suffer from the disadvantage of a smaller error exponent.

The proof of Theorem 3 relies on the following Covering Lemma for two-step coding, which is a straightforward extension of the corresponding lemma for one-step coding [2, Lemma 2.4.1].

Let $P = P^{(n)}$ be a type on $\mathcal{X}$ (cf. e.g., [2]), i.e., a pmf with rational probabilities with (common) denominator $n$. Let $T_0^n$ denote the set of sequences in $\mathcal{X}^n$ of (common) type $P$.

**Lemma 1:** For distortion measures $d_1$ on $\mathcal{X} \times \mathcal{Y}_1$ and $d_2$ on $\mathcal{X} \times \mathcal{Y}_2$, type $P = P^{(n)}$ on $\mathcal{X}$ and numbers $\Delta_1 > \Delta_2 > 0$, $R_1 > R(P, \Delta_1)$, $b_1 > b_2 > 0$, there exist
- a set $B_1 \subset \mathcal{Y}_1^n$, such that
\[ \frac{1}{n} \log |B_1| \leq R_1 + \delta_1 \] \tag{17}
and $B_1$ covers $T_0^n$
\[ \bigcup_{y_1 \in B_1} \mathcal{N}_1(y_1) = T_0^n \]
where
\[ \mathcal{N}_1(y_1) = \{ x \in T_0^n : d_1(x, y_1) \leq \Delta_1 \}, \quad y_1 \in \mathcal{Y}_1^n \]
- sets $B_2(y_1) \subset \mathcal{Y}_2^n$, $y_1 \in B_1$, such that
\[ \frac{1}{n} \log \left( \sum_{y_1 \in B_1} |B_2(y_1)| \right) \leq R(P, R_1, \Delta_1, \Delta_2) + \delta_2 \] \tag{18}
and $B_2(y_1)$ covers $\mathcal{N}_1(y_1)$, i.e.,
\[ \bigcup_{y_2 \in B_2(y_1)} \mathcal{N}_2(y_2) \supset \mathcal{N}_1(y_1), \quad y_1 \in B_1, \]
where
\[ \mathcal{N}_2(y_2) = \{ x \in T_0^n : d_2(x, y_2) \leq \Delta_2 \}, \quad y_2 \in \mathcal{Y}_2^n \]
provided that $n \geq N(d_1, d_2, \delta_1, \delta_2)$.

**Proof of Theorem 3:** We commence with the existence part of the proof. It is convenient to define the following quantities: For a pmf $Q$ on $\mathcal{X}$ and $R_2 > R_1 > 0$, let
\[ \Delta(Q, R_1) \triangleq \inf_{P_X = Q_{X | (X,Y_1) \leq R_1}} E_{d_1}(X, Y_1) \]
and
\[ \Delta(Q, R_1, R_2) \triangleq \inf_{P_X = Q_{X | (X,Y_1,Y_2) \leq R_1}} E_{d_1}(X, Y_1, Y_2). \]
Consider the sets
\[ U_1^{(n)} \triangleq \bigcup_{Q : E(Q, R_1, \Delta_1, \Delta_2) > R_2} T_0^n \]
and
\[ U_2^{(n)} \triangleq \bigcup_{Q : E(Q, R_1, \Delta_1) > R_1} T_0^n. \]

Obviously
\[ P^n \left[ U_1^{(n)} \right] \leq (n + 1)^{|X|} \exp \left\{ -n F(P, R_1, \Delta_1) \right\} \]
\[ \leq \exp \left\{ -n F(P, R_1, \Delta_1) - \delta_1 \right\} \]
for all $n$ large (cf. [2, Lemma 1.2.6]), and
\[ P^n \left[ U_2^{(n)} \right] \leq (n + 1)^{|X|} \exp \left\{ -n F(P, R_2, \Delta_2 | R_1, \Delta_1) \right\} \]
\[ \leq \exp \left\{ -n F(P, R_2, \Delta_2 | R_1, \Delta_1) - \delta_2 \right\} \]
for all $n$ large. Next, by Lemma 1, there exist sequences $e_1^{(n)}$ and $e_2^{(n)}$ with $\lim_n e_1^{(n)} = 0$ and $\lim_n e_2^{(n)} = 0$, such that for every type $Q$ of sequences in $\mathcal{X}^n$, there exist sets $B_{Q,1} \subset \mathcal{Y}_1^n$ and $B_{Q,2}(y_1) \subset \mathcal{Y}_2^n$, $y_1 \in B_{Q,1}$, satisfying
\[ \frac{1}{n} \log |B_{Q,1}| \leq R_1 + e_1^{(n)} \]
\[ \frac{1}{n} \log \left( \sum_{y_1 \in B_{Q,1}} |B_{Q,2}(y_1)| \right) \leq R_2 + e_2^{(n)} \]
for every $x \in T_0^n$
\[ d_1(x, B_{Q,1}) \triangleq \min_{y_1 \in B_{Q,1}} d_1(x, y_1) \leq \Delta(Q, R_1) \]
and
\[ d_2(x, B_{Q,2}(y_1)) \triangleq \min_{y_2 \in B_{Q,2}(y_1)} d_2(x, y_2) \leq \Delta(Q, R_1, R_2), \quad y_1 \in B_{Q,1}, \]
for $x$ such that $d_1(x, y_1) \leq \Delta_1$. Next, we set
\[ B_1 \triangleq \bigcup_{Q_{Y_1}} B_{Q,1}, \]
\[ B_2(y_1) \triangleq B_{Q_{Y_1,2}(y_1)}, \quad y_1 \in B_1, \]
where $Q_{Y_1}$ is the type of $y_1$. Then, the Type Counting Lemma [2, Lemma 1.2.2] yields that
\[ \frac{1}{n} \log |B_1| \leq R_1 + e_1^{(n)} \]
\[ \frac{1}{n} \log \left( \sum_{y_1 \in B_1} |B_2(y_1)| \right) \leq R_2 + e_2^{(n)} \]
where $\lim_n e_1^{(n)} = \lim_n e_2^{(n)} = 0$. Furthermore, $R(Q, \Delta_1) \leq R_1$ implies in a standard manner that
\[ \Delta(Q, R_1) \leq \Delta_1 \] \tag{19}
and \( R(Q, R_1, \Delta_1, \Delta_2) \leq R_2 \) implies

\[
\Delta(Q, R_1, \Delta_1, R_2) \leq \Delta_2
\]

(20)

which is seen as follows.

Let \( F_{X_1,Y_2} \) achieve \( R(Q, R_1, \Delta_1, \Delta_2) = I(X \geben X_1, Y_2) \leq R_2 \), where the rv \((X, Y_1, Y_2)\) is distributed according to \( F_{X_1,Y_2} \). Then

\[
E_{d_1}(X, Y_1) \leq \Delta_1, \quad E_{d_2}(X, Y_2) \leq \Delta_2, \quad I(X \geben X_1, Y_1, Y_2) \leq R_2
\]

Hence, \( \Delta(Q, R_1, \Delta_1, \Delta_2) \leq E_{d_1}(X, Y_1) \leq \Delta_1 \), since

\[
E_{d_1}(X, Y_1) \leq \Delta_1, \quad E_{d_2}(X, Y_2) \leq \Delta_2, \quad I(X \geben X_1, Y_1, Y_2) \leq R_2
\]

As a consequence of (19) and (20), we have

\[
d_1(x, S_1) \leq \Delta_1, \quad x \in \mathcal{X}^n \setminus U_1^{(n)}
\]

\[
d_2(x, R_2(y_1)) \leq \Delta_2, \quad y_1 \in R_1, x \in \mathcal{X}^n \setminus U_2^{(n)}. \]

These inequalities establish the direct (existence) part of the proof.

Turning to the converse part, consider any pmf \( Q \) on \( \mathcal{X} \) with the property that

\[
R(Q, R_1, \Delta_1, \Delta_2) > R_2 + \delta_2.
\]

(21)

The remainder of the proof relies on the following Claim which constitutes a strong converse to Theorem 1.

**Claim 1:** For any sequence of codes \((f_1^{(n)}, \phi_1^{(n)}), (f_2^{(n)}, \phi_2^{(n)})\) with

\[
\frac{1}{n} \log \|f_1^{(n)}\| \leq R_1 + \delta_1
\]

\[
\frac{1}{n} \log \|f_2^{(n)}\| \leq R_2 + \delta_2
\]

for all \( n \) large, (21) implies

\[
e(Q, (f_1^{(n)}, \phi_1^{(n)}), (f_2^{(n)}, \phi_2^{(n)}), \Delta_2 \mid R_1, \Delta_1) \geq \frac{1}{2}
\]

whenever \( n \geq N(d_1, d_2, \delta_1, \delta_2) \).

The proof of the Claim is obtained by mimicking the proof of [2, Theorem 2.2.3], and is, therefore, omitted. Then, by [2, Corollary 1.1.2] it holds for all \( n \) large enough that

\[
e(P, (f_1^{(n)}, \phi_1^{(n)}), (f_2^{(n)}, \phi_2^{(n)}), \Delta_2 \mid R_1, \Delta_1)
\]

\[
\geq \exp \{-n[D(Q \parallel P) + \delta_1]\}.
\]

Note that the case \( R(Q, \Delta_1) > R_1 + \delta_1 \) alone was shown ([2, Theorem 2.4.5]) to imply that

\[
e(P, (f_1^{(n)}, \phi_1^{(n)}), \Delta_1) \geq \exp \{-n[D(Q \parallel P) + \delta_1]\}.
\]

Since \( \delta_1, \delta_2 \) were arbitrary as was \( Q \) subject to (21), the desired converse follows.

It is obvious that, since

\[
\max \{e(P, (f_1^{(n)}, \phi_1^{(n)}), \Delta_1), e(P, (f_2^{(n)}, \phi_2^{(n)}), (f_2^{(n)}, \phi_2^{(n)}), \Delta_2 \mid R_1, \Delta_1)\}
\]

\[
\leq e(P, (f_1^{(n)}, \phi_1^{(n)}), (f_2^{(n)}, \phi_2^{(n)}), \Delta_2 \mid R_1, \Delta_1)
\]

\[
\leq e(P, (f_1^{(n)}, \phi_1^{(n)}), \Delta_1) + e(P, (f_2^{(n)}, \phi_2^{(n)}), (f_2^{(n)}, \phi_2^{(n)}), \Delta_2 \mid R_1, \Delta_1),
\]

the overall error exponent will be the greater of the two. Hence, (8) follows.

\[\square\]

## IV. Discussion

We have determined the error exponent for the problem of successive refinement by partitioning for a DMS. As expected, it is generally smaller than the error exponent for one-step coding. It is interesting to note that even when the Markov condition (3) holds so that the rate-distortion functions coincide for one-step and two-step coding, it may hold that the error exponents for the latter are strictly smaller than that for the former.

For a DMS with Hamming distance distortion measure, a simple necessary and sufficient condition for the error exponents to differ in the presence of Markov condition, can be expressed in terms of the coding rates \( R_1, R_2 \) and the one-step rate-distortion function. An extension of this result to arbitrary distortion measures remains unresolved.

Finally, it can be shown that the error function for two-step coding (cf. (6)) goes to one exponentially fast, whenever \( R_2 < R(P, \Delta_1, \Delta_2) \). To this end, we conclude from Rimoldi ([15, Theorem 1, converse part]) that if \( R_2 < R(P, \Delta_1, \Delta_2) \), then for every sequence of \( n \)-length block codes satisfying (4) and (5), we have

\[
\lim_n e(P, (f_1^{(n)}, \phi_1^{(n)}), (f_2^{(n)}, \phi_2^{(n)}), \Delta_1, \Delta_2) = 1.
\]

Indeed, in analogy with one-step coding (cf. Csiszár and Körner [2, sec. 2.4]), this convergence occurs exponentially fast. To see this, define

\[
e^{(n)}(P, R_1, R_2, \Delta_1, \Delta_2) = \min e(P, (f_1^{(n)}, \phi_1^{(n)}), (f_2^{(n)}, \phi_2^{(n)}), \Delta_1, \Delta_2)
\]

where the minimum is taken over all codes satisfying (4) with \( R_1 > R(P, \Delta_1) \) and

\[
\frac{1}{n} \log \|f_1^{(n)}\| + \|f_2^{(n)}\| \leq R_2
\]

where \( 0 < R_2 < R(P, \Delta_1, \Delta_2) \). Then, it holds that

\[
\lim_n \left\{ -\frac{1}{n} \log \left[ 1 - e^{(n)}(P, R_1, R_2, \Delta_1, \Delta_2) \right] \right\} = G(R_1, R_2, \Delta_1, \Delta_2)
\]

where

\[
G(R_1, R_2, \Delta_1, \Delta_2) = \inf_Q \{ D(Q \parallel P) + |R(Q, R_1, \Delta_1, \Delta_2) - R_2| \}
\]

with \( |x| \) denoting \( \max \{x, 0\} \). The proof is identical to that for one-step coding (cf. Csiszár and Körner [2, problem 2.4.6, pp. 158–159]) and relies on the strong converse (Rimoldi, [15, Theorem 1, converse part]).

## Appendix

**Proof of Lemma 1**

Let \( P = P^{(n)} \) be a fixed but arbitrary type of sequences in \( \mathcal{X}^n \).

Let \( U_1, B_1 \) and \( U_2, B_2 \), respectively, denote the subsets of \( T^n \) not covered by \( B_1 \) and \( B_2 \), i.e.

\[
U_1 = T^n \setminus \bigcup_{y_1 \in B_1} N_1(y_1) = T^n \cap \left( \bigcap_{y_1 \in B_1} N_1(y_1)^c \right)
\]

and

\[
U_2 = \bigcup_{y_1 \in B_1} N_1(y_1) \setminus \bigcup_{y_2 \in B_2(y_1)} N_2(y_2) = \bigcup_{y_2 \in B_2(y_1)} N_2(y_2^c), \quad y_2 \in B_2(y_1).
\]
Consider a $X \times Y_1 \times Y_2$-valued RV $(X, Y_1, Y_2)$ with $P_X = P$ and
\[
E\|t_1\langle X, Y_1 \rangle \leq \| \Delta_1 - \eta_1 \|^+ \\
E\|t_2\langle X, Y_2 \rangle \leq \| \Delta_2 - \eta_2 \|^+ \tag{22}
\]
where $\eta_1, \eta_2 > 0$ will be specified later.

Let $T_{\gamma_1}^{(n)}$ (resp., $T_{\gamma_2,Y_2|Y_1}^{(n)}(y_1)$) denote the set of $Y_1$-typical sequences $y_1 \in \mathcal{Y}_1^n$ (resp., $Y_2$ $Y_1$-typical sequences $y_2 \in \mathcal{Y}_2^n$ with respect to $y_1 \in \mathcal{Y}_1^n$) [2, Definitions 1.2.8, 1.2.9, pp. 33–34]. Let $T_m$ denote the set of all (not necessarily distinct) collections of $m$ elements of $T_{\gamma_1}^{(n)}$. Similarly, for $i = 1, \ldots, m$, let $\mathcal{G}_m(y_i)$ denote the set of all (not necessarily distinct) collections of elements of $T_{\gamma_2,Y_2|Y_1}^{(n)}(y_1)$.

Let $Z^n = (Z_1, \ldots, Z_m)$ be an RV uniformly distributed on $T_m$. Next, for $i = 1, \ldots, m$, given that $Z_i = z_i$, let $W_{ni} = (W_{ni1}, \ldots, W_{ni,n})$, be an RV distributed (conditionally) uniformly on $\mathcal{G}_m(z_i)$. In other words, $Z_i$'s are i.i.d. with
\[
\Pr\{Z_i = z_i\} = \frac{1}{|T_{\gamma_1}^{(n)}|}, \quad i = 1, \ldots, m
\]
and the $W_{ni}$'s are conditionally i.i.d., with
\[
\Pr\{W_{ni} = w_{ni} \mid Z_i = z_i\} = \frac{1(|w_{ni} \in T_{\gamma_2,Y_2|Y_1}^{(n)}(z_i)|)}{|T_{\gamma_2,Y_2|Y_1}^{(n)}(z_i)|}, \quad j = 1, \ldots, n_i, i = 1, \ldots, m.
\]
We must show the existence of sets $B_1 \subseteq \mathcal{Y}_1^n$ and $B_2(y_1) \subseteq \mathcal{Y}_2^n$ $y_1 \in B_1$, such that
\[
\left\| U(B_1) \cup \bigcup_{y_1 \in B_1} U(B_2(y_1)) \right\| = 0.
\]
To this end, it suffices to show that
\[
E\left| U(Z^n) \cup \bigcup_{i=1}^m U(W_{ni}) \right| < 1.
\]
Now
\[
E\left| U(Z^n) \cup \bigcup_{i=1}^m U(W_{ni}) \right| = \sum_{x \in T_m} \Pr\{x \in U(Z^n) \cup \bigcup_{i=1}^m U(W_{ni})\} \leq \sum_{x \in T_m} \Pr\{x \in U(Z^n)\} + \sum_{x \in T_m} \sum_{i=1}^m \Pr\{x \in U(W_{ni})\}. \tag{23}
\]
We consider first the first term on the right side of the inequality above. By choosing
\[
\exp \left[ n \left( I(X \land Y_1) + \frac{2}{3} \right) \right] \leq m \leq \exp \left[ n \left( I(X \land Y_1) + \frac{2}{3} \delta_1 \right) \right]
\]
we get that
\[
\sum_{x \in T_m} \Pr\{x \in U(Z^n)\} \leq \exp \left[ n \log |X| - \exp \frac{n \delta_1}{6} \right] \leq \frac{1}{2} \tag{24}
\]
for all $n \geq N_0(\delta_1, d_1, (\eta_1)$ (cf., e.g., [2, p. 152]). Next, to bound the second term on the right side of (23), for each $x \in T_m$, let
\[
\mathcal{A}(x) = \{ z \in T_{\gamma_1}^{(n)} : x \in N_1(z) \} \subseteq \{ z \in T_{\gamma_1}^{(n)} : d_1(x, z) \leq \Delta_1 \}.
\]
Assume for the time being, the following

**Claim 2:** For any $\delta' > 0$, we have
\[
|\mathcal{A}(x)| \leq \exp \left[ n \left( I(Y_1 \land Y_2) - R(P, \Delta_1) + \delta' \right) \right]
\]
for all $n \geq N_2(d_1, \eta_1, \delta')$.

Then, the remainder of the proof of the Lemma is straightforward. The second term on the right side of (23) is
\[
\Pr\{x \in \mathcal{A}(x)\} = \Pr\{x \in N_1(z) \cap \bigcap_{i=1}^m N_2(W_{ni})^c\}
\]

Further
\[
\Pr\{x \in \mathcal{A}(x)\} = \prod_{j=1}^{n_1} \Pr\{x \in N_2(W_{nj})^c \mid Z_i = z_i\}
\]

where the previous inequality follows as in [2, p. 151]. We continue the bounding according to
\[
\prod_{j=1}^{n_1} \left( 1 - \Pr\{x \in N_2(W_{nj}) \mid Z_i = z_i\} \right)
\]

for all $n \geq N_2(d_2, \eta_2)$ (cf., [2, p. 151]). Furthermore, the right side above is bounded above by
\[
\left( 1 - \exp \left[ -\frac{n \delta_2}{6} \right] \right)^{n_1}
\]

provided that
\[
\exp \left[ n \left( I(Y_1 \land Y_2) + \frac{2}{3} \delta_2 \right) \right] \leq \exp \left[ n \left( I(Y_1 \land Y_2) + \frac{2}{3} \delta_2 \right) \right]
\]

Hence, we can finally bound the second term on the right side of (23) as
\[
\sum_{x \in T_m} \sum_{i=1}^m \Pr\{x \in U(W_{ni})\} \leq \exp \left[ n \log |X| - \exp \frac{n \delta_2}{6} \right] \frac{1}{|T_{\gamma_1}^{(n)}|} |\mathcal{A}(x)|.
\]
Using the facts
\[
m \leq \exp \left\{ n \left[ I(X \land Y_1) + \frac{3}{4} \delta_1 \right] \right\}
\]
\[
|\mathcal{T}_{\mathcal{Y}_1}| \geq \exp \left\{ n \left[ H(Y_1) - \frac{\delta_1}{4} \right] \right\}
\]
and Claim 2, the right side above is easily seen to be bounded by
\[
\exp \left\{ n \left[ 2 \log |\mathcal{X}| + 2 \delta_1 - \frac{n \delta_2}{6} \right] \right\} < \frac{1}{2}
\]

(25)
for all \( n \geq N_3(\delta_1, \delta_2) \) (because \( R_1 - R(P, \Delta_1) \leq \log |\mathcal{X}| \)). Finally, we obtain from (24) and (25) that
\[
E \left[ |\mathcal{U}(Z^n) \cup \left( \bigcup_{i=1}^{m} \mathcal{U}(W_i^n) \right) \right] < 1
\]
for all
\[
n \geq N(d_1, d_2, \eta_1, \eta_2, \delta_1, \delta_2) = \max \left\{ N_1, \ldots, N_3 \right\}
\]
It then follows for all \( n \) suitably large that there exist sets \( B_i \subset \mathcal{Y}_n, B_i(Y_i) \subset \mathcal{Y}_i^2, y_i \in B_i \), such that
\[
|B_i| \leq m, \quad |B_i(Y_i)| \leq n_i, \quad i = 1, \ldots, m
\]
so that
\[
\sum_{i=1}^{m} |B_i(Y_i)| \leq \sum_{i=1}^{m} n_i \leq m \exp \left\{ n \left[ I(X \land Y_1 | Y_1) + \frac{3}{4} \delta_1 \right] \right\}
\]
\[
\leq \exp \left\{ n \left[ I(X \land Y_1) + \frac{3}{4} \delta_2 \right] \right\}
\]
Equivalently
\[
\frac{1}{n} \log |B_1| \leq I(X \land Y_1 | Y_1) + \frac{3}{4} \delta_1
\]
\[
\frac{1}{n} \log |B_2| \leq I(X \land Y_1 | Y_2) + \frac{3}{4} \delta_2
\]

(26)
Assume that \( I(X \land Y_1) \leq R_1 \). Then by the uniform continuity of \( R(P, R_1, \Delta_1, \Delta_2) \) (which follows analogously as in [2, Lemma 2.2.2, pp. 124-125]), the desired inequalities (17), (18) follow for \( \eta_1, \eta_2 \) sufficiently small (cf. (22)).

It remains to establish the Claim. To this end, observe that \( z \in A(x) \) iff
\[
z \in T_{\mathcal{Y}_1}^x \quad \text{and} \quad d_1(x, z) = E_{P_{X,Y}}[d_1(X, \tilde{Y})] \leq \Delta_1
\]
where \( (X, \tilde{Y}) \) is a \( \mathcal{X} \times \mathcal{Y}_1 \)-valued r.v. with joint pmf equal to the joint type \( P_{X,Y} \) of \( (x, z) \). Further,
\[
H(V | P_z) = H(P_z) - I(X \land \tilde{Y})
\]
\[
\leq H(Y_1) + \frac{\delta'}{3} - \min I(X \land \tilde{Y})
\]
\[
= H(Y_1) + \frac{\delta'}{3} - R(P, \Delta_1)
\]
(27)
where the inequality follows for all \( n \) large (depending on \( \delta' \)) from [2, Lemma 1.2.7] and the minimum is taken over all r.v.'s \( (X, \tilde{Y}) \) such that \( P_X = P_z \) and \( Ed_1(X, \tilde{Y}) \leq \Delta_1 \).

Hence, finally
\[
A(x) = \bigcup_{\forall v \in E_{P_{X,Y}}[d_1(X, \tilde{Y})] \leq \Delta_1} T_{\mathcal{Y}_1}^v (z) \cap T_{\mathcal{Y}_1}^x
\]
so that
\[
|A(x)| \leq (1 + n)^{|\mathcal{X}|} \exp \left\{ n \left[ H(Y_1 | X) + \frac{2\delta'}{3} \right] \right\}
\]
for all \( n \geq N_3(d_1, \eta_1, \delta') \), where the inequality above is a consequence of (27).

ACKNOWLEDGMENT

The authors wish to thank the anonymous reviewers for their comments which resulted in a significant modification of our results.

REFERENCES