On Reliability of Content Identification from Databases based on Noisy Queries

Gautam Dasarathy
Electrical and Computer Engineering
University of Wisconsin
Madison, WI 53706, USA
Email: dasarathy@wisc.edu

Stark C. Draper
Electrical and Computer Engineering
University of Wisconsin
Madison, WI 53706, USA
Email: sdraper@ece.wisc.edu

Abstract—**In this paper we quantify an achievable error-exponent for the problem of content identification from a large database based on noisy queries.**

I. Introduction

We consider the problem of content identification from a database wherein quantized representations of 2^{nR_I} length-n "enrollment" vectors are stored. Based upon a noisy observation of one of the (non-quantized) enrollment vectors and the stored data, the goal is to identify from which enrollment vector the noisy observation (or the "query") was generated.

The information-theoretic limits of this problem were found in [1] (see also [2] for related problem formulations) under the following model. The enrollment vectors, $\mathbf{x}_m$ for $m = 1, \ldots, 2^{nR_I}$ are chosen in an i.i.d. manner according to p_X. An index of a codeword in a pre-defined rate-R_C codebook $\mathcal{C}$ is used to represent each of these vectors in a database. As it takes nR_C bits to store the index of the representation of each enrollment, and 2^{nR_I} representations are stored, the entire database is of size $nR_C 2^{nR_I}$ bits. The query $\mathbf{Z}$ presented during the identification phase is a length-n observation of one of the 2^{nR_I} (unquantized) enrollment vectors via the discrete memoryless channel (DMC) $W(\cdot|\cdot)$. The decoder's objective is to identify reliably the codeword corresponding to the enrollment vector from which the query was generated.

In [1] the capacity region of this problem was shown to be parameterized by a (rate-distortion) test channel $p_{U|X}$. Given the joint distribution $p_{U|X}(u|x)p_X(x)W(z|x)$ the "compression/identification" rate pair (R_C, R_I) is achievable if

$$R_C > I(U;X) \qquad R_I < I(U;Z).$$

The achievable rate region is the convex hull of the union of achievable rate pairs over all test channels. The achievability is closely related the the Wynzer-Ziv problem. The codebook $\mathcal{C}$ needs to have good covering properties ($R_C > I(U;X)$) to ensure reliable encoding. The union of codewords stored in the database (which is a subset of $\mathcal{C}$) needs to form a code that has good packing properties ($R_I < I(U;Z)$) for reliable decoding (this set plays the role of the "random bin" of codewords in Wyner-Ziv).

G. Dasarathy and S. C. Draper were supported by the Air Force Office of Scientific Research under Grant FA9550-09-1-0140.

In this paper we look into convergence issues. We quantify an achievable error exponent tradeoff for a particular encoding and decoding strategy. The coding strategy and form of the results are reminiscent both of previous error exponent studies of the Wyner-Ziv and of the Gel'fand-Pinsker problems (see, e.g., [3], [4] and [5], respectively). However, while those settings are concerned with the statistical behavior of a *triplet* of vectors (source/side-information/codeword in Wyner-Ziv and state/input/channel-output in Gel'fand-Pinsker), in our problem we will only be concerned with the *pairwise* end-to-end statistical relationship between codeword and channel-output. This focus yields a novel form of error exponent.

The remainder of the paper is organized as follows. In Sec. II we introduce the problem formulation and coding scheme. In Sec. III we state our main result, which is proved in Sec. IV. The appendix contains the proof of one of the lemmas that we use to obtain our result. We follow [6] for our notational choices.

II. Problem formulation and coding scheme

In this section we formally state the problem setting and define the family of codes we analyze.

A. Problem Setup

Environment: We suppose that there are $M = 2^{nR_I}$ (ignoring integer effects) items to be represented in the database. To each is associated a length-n "feature vector" or "enrollment vector" $\mathbf{x}(m)$, $m = 1, 2, \ldots, M$ which is made up of letters from the alphabet $\mathcal{X} = \{b_1, b_2, \ldots, b_{|\mathcal{X}|}\}$. The $\mathbf{x}(m)$ are drawn independently and from p_X in an i.i.d. manner.

Enrollment Phase: Each enrollment vector is represented in the database by a codeword selected from a pre-defined rate-R_C codebook $\mathcal{C}$. Each codeword $\mathbf{u}$ in the codebook is made up of symbols from the alphabet $\mathcal{U} = \{a_1, a_2, \ldots, a_{|\mathcal{X}|}\}$

Identification Phase: An index J is selected uniformly at random from $\{1, 2, \ldots, M\}$. A noisy version of $X^n(J)$, Z^n is observed at the database as a "query" where $\Pr[\mathbf{Z} = \mathbf{z}|\mathbf{X} = \mathbf{x}] = \prod_{i=1}^n W(z_i|x_i)$, i.e., the noise is modeled by the DMC $W : \mathcal{X} \to \mathcal{Z}$, where $\mathcal{Z} = \{z_1, z_2, \ldots, z_{|\mathcal{Z}|}\}$

The goal is to design an encoding function $f(\cdot)$ and a decoding function $g(\cdot)$ such that, with high probability, $g(\cdot)$ returns the correct value of J.

B. Coding scheme

Our encoding scheme is based upon a source-type-dependent encoding rule. As there are only a polynomial number of types we can design a rate-R_C codebook for *each* possible type and concatenate them into a single "mother" code without the mother code being of higher rate. Further, this construction also indicates to the decoding algorithm the type of the encoded source sequence (though knowledge of the sequence itself is, of course, lost).

Codebook construction: Any observed enrollment vector $\mathbf{x}$ has some empirical distribution (type) $P_{\mathbf{x}}$. As we will have frequent occassion to index our encoding scheme and error events by the source type, we index each possible source type by i, where $1 \leq i \leq \binom{n+|\mathcal{X}|-1}{|x|-1}$. (Notice that the range of i comes from the standard type counting lemma.) Thus, every $\mathbf{x} \in \mathcal{X}^n$ satisfies $\mathbf{x} \in T_{P_i}$ for some i in this range.

To each source type $P_i, 1 \leq i \leq \binom{n+|\mathcal{X}|-1}{|x|-1}$ we associate a rate-R_c (vector quantization) codebook as follows. We pick a conditional type $S_i : \mathcal{X} \to \mathcal{U}$ and compute the marginal type Q_i induced by S_i and P_i on $\mathcal{U}$, i.e., $Q_i(a) = \sum_{b \in \mathcal{X}} P_i(b) S_i(b \mid a)$. Then, we generate a codebook $\mathcal{C}_i$ which consists of 2^{nR_c} length n vectors chosen independently and uniformly at random from T_{Q_i}. (Note that two codebooks $\mathcal{C}_i$ and $\mathcal{C}_j$ for $i \neq j$ may be generated from the same type class.)

The overall codebook $\mathcal{C}$ is the *concatenation* of the $\mathcal{C}_i$. Since the number of types is upper bounded by $(n+1)^{|\mathcal{X}|}$ the codebook $\mathcal{C}$ is still rate-R_c. Note that by concatenation we do not mean union. The index of each codeword in $\mathcal{C}$ indicates both the codeword sequence itself *and* the index i of the codebook from which it was generated (thus P_i and S_i as well). To denote the codebook, the induced type on $\mathcal{U}$ and the assigned conditional type that correspond to a particular vector $\mathbf{x}$, we sometimes write $\mathcal{C}_{\mathbf{x}}, Q_{\mathbf{x}}$ and $S_{\mathbf{x}}$ respectively.

Enrollment: Given enrollment vector $\mathbf{x} \in \mathcal{T}_{P_i}$ we find a $\mathbf{u} \in \mathcal{C}_i$ such that $\mathbf{u} \in \mathcal{T}_{S_i}(\mathbf{x})$, i.e., that is in the S_i-shell of $\mathbf{x}$. If there is more than one such element of $\mathcal{C}_i$ we selected one randomly and make note of its index. If there is no such codeword in $\mathcal{C}_i$ there is an encoding error. We call the set of all stored indices the *database* and we denote it by $\mathcal{D}$. We enumerate the set of stored codewords as $\mathbf{u}(m)$ for $1 \leq m \leq 2^{nR_I}$ and write $i(m)$ for their corresponding indices in the codebook. Henceforth, we use the function $f : \mathcal{X}^n \times \mathcal{C} \to \{1, 2, \ldots, |\mathcal{C}|\}$ to denote this operation.

Identification: Given observation of query $\mathbf{Z}$ we select the stored codeword $\mathbf{u}^*$ such that $\mathbf{u}^* = \arg\max\limits_{\mathbf{u} \in \mathcal{D}} I(\mathbf{u}; \mathbf{Z})$, i.e., we pick the codeword that has the maximum empirical mutual information with $\mathbf{Z}$. As above, we use the function $g : \mathcal{Z}^n \times \mathcal{D} \to \{1, 2, \ldots, 2^{nR_I}\}$ to denote the operation of the decoder.

III. Main results

The following theorem is the main result of our paper.

Theorem 1: The reliability function (error exponent) of the problem is lower bounded by

$$\inf_{P_X} \sup_{\substack{S: \\ I(P_X, S) < R_C}} \inf_{V: \mathcal{U} \to \mathcal{Z}} D(P_X \| p_X) + D\left(\tilde{J}^*_{V, P_X, S} \| J \mid Q_U\right) + |I(Q_{\mathbf{U}}, V) - R_I|^+$$

where $Q_{\mathbf{U}}(\cdot) = \sum_{b \in \mathcal{X}} P_X(b) S(\cdot \mid b)$ is the distribution induced on $\mathcal{U}$ by P_X and S, $J : \mathcal{U} \to \mathcal{X} \times \mathcal{Z}$ is a stochastic matrix (or conditional type) defined as $J(b, c \mid a) = \frac{S(a|b) P_X(b)}{Q_U(a)} W(c \mid b)$, for all $(a, b, c) \in \mathcal{U} \times \mathcal{X} \times \mathcal{Z}$ and $\tilde{J}^*_{V, P_X, S} : \mathcal{U} \to \mathcal{X} \times \mathcal{Z}$ is a stochastic matrix defined as

$$\tilde{J}^*_{V, P_X, S} = \arg \min_{\tilde{J} \in \mathcal{E}(V)} D(\tilde{J} \| J \mid Q_U), \tag{1}$$

where

$$\mathcal{E}(V) = \left\{ \tilde{J} : \sum_{b \in \mathcal{X}} \tilde{J}(b, c \mid a) = V(c \mid a), \forall (a, c) \in \mathcal{U} \times \mathcal{X} \right\} \tag{2}$$

It can be shown that this error exponent is positive in the capacity region indicated in [1]. To prove this theorem we need the following two lemmas. They are proved in Appendix A and Appendix B respectively.

Lemma 1: If the S_i in the encoding rule are chosen so that $R_C - I(P_i, S_i) > \epsilon$, then the probability of encoding error is upper bounded as

$$\Pr[\text{Encoding Error}] \leq \rho_1(n) 2^{-n[D(P_{i^*} \| p_{\mathbf{x}})]} 2^{-\rho_2(n) 2^{n\epsilon}}$$

where the $\rho_1(n)$ and $\rho_2(n)$ are strictly positive polynomials in n. $P_{i^*} = \arg\max_{P_i} 2^{-nD(P_i \| p_X)} 2^{-\rho_2(n) 2^{n|R_C - I(P_i, S_i)|}}$ where the maximization is over all possible types on $\mathcal{X}^n$.

Notice that this lemma guarantees that as long as our choice of S_i is such that $I(P_i, S_i)$ is less than R_C for every i, the probability that an encoding error occurs goes down doubly exponentially in the blocklength.

Lemma 2: Let $S : \mathcal{X} \to \mathcal{U}$ and $W : \mathcal{X} \to \mathcal{Z}$ be fixed conditional types. Further, suppose that $\mathbf{x} \in \mathcal{X}^n$ is of type $P_{\mathbf{x}}$ and $\mathbf{u} \in \mathcal{U}^n$ is such that $(\mathbf{x}, \mathbf{u})$ have the joint type $P_{\mathbf{x}} \times S$. Then for any conditional type $V : \mathcal{U} \to \mathcal{Z}$, the following holds

$$W^n\left(\mathcal{T}_V(\mathbf{u}) | \mathbf{x}\right) \leq \rho(n) 2^{-nD\left(\tilde{J}^*_{V, P_{\mathbf{x}}, S} \| J | Q_{\mathbf{u}}\right)} \tag{3}$$

where $\rho(n)$ is a strictly positive polynomial in n and $Q_{\mathbf{u}}$ is the type of $\mathbf{u}$. $J : \mathcal{U} \to \mathcal{X} \times \mathcal{Z}$ is a stochastic matrix defined as $J(b, c \mid a) = \frac{S(a|b) P_{\mathbf{x}}(b)}{Q_{\mathbf{u}}(a)} W(c \mid b)$, for all $(a, b, c) \in \mathcal{U} \times \mathcal{X} \times \mathcal{Z}$ and $\tilde{J}^*_{V, P_{\mathbf{x}}, S}$ is defined according to (1)

Observe that, given a Markov Chain $U^n - X^n - Z^n$, such that the relationship between X and Z is memoryless, this lemma bounds the probability that Z^n has a particular empirical relationship with a realization u^n of U^n, given that X^n takes on the value x^n.

IV. Error Analysis

In this section we bound the probability of error, i.e., the probability that the recovered codeword index does not match the index of the true enrollment vector (the one that generated the query). Let E_e denote the event that encoding fails (i.e., we are unable to encode one or more of the enrollment vectors) and let E_d denote the event that decoding fails. By symmetry we can assume, without any loss of generality, that the observed query is a corrupted version of $\mathbf{X}(1)$ and that we are interested in the probability that $g\left(\mathbf{Z}, \mathcal{D}\right) \neq 1$. We begin by writing

$$\Pr\left[\text{Error}\right] \leq \Pr\left[E_e\right] + \Pr\left[E_d \mid E_e^c\right]. \tag{4}$$

By Lemma 1, it is clear that $\Pr[E_e]$ is not the dominating term as long as we constrain each S_i to satisfy $R_C - I(P_i, S_i) > \epsilon$. So, we focus our attention on the second term which corresponds to the probability of a decoding error, given that encoding was successful. We proceed by conditioning on the value of the feature vector and then by conditioning on the particular realizations of the database $\mathcal{D}$.

$$\begin{aligned}\Pr\left[E_d \mid E_e^c\right] &= \sum_{\mathbf{x}} p_{\mathbf{X}}(\mathbf{x}) \Pr\left[E_d \mid E_e^c, \mathbf{X} = \mathbf{x}\right] \\ &= \sum_{\mathbf{x}} p_{\mathbf{X}}(\mathbf{x}) \sum_{\mathcal{D}} \Pr\left[\mathcal{D} \mid E_e^c, \mathbf{x}\right] \Pr\left[E_d \mid E_e^c, \mathbf{x}, \mathcal{D}\right]\end{aligned} \tag{5}$$

Let us first look at the term $\Pr\left[E_d \mid E_e^c, \mathbf{x}, \mathcal{D}\right]$ in (5). By definition, it is the probability that decoding fails ($g(\mathbf{Z}, \mathcal{D}) \neq 1$) given a particular enrollment vector ($\mathbf{x}$), the database ($\mathcal{D}$) and given that encoding succeeds. We proceed to expand this as follows

$$\Pr\left[E_d \mid E_e^c, \mathbf{x}, \mathcal{D}\right]$$

$$= \Pr\left[g(\mathbf{Z}, \mathcal{D}) \neq 1 \mid E_e^c, \mathbf{x}, \mathcal{D}\right]$$

$$= \sum_{\mathbf{z}} p_{\mathbf{Z}|\mathbf{X}}(\mathbf{z} \mid \mathbf{x}) \mathbb{1}\left\{g(\mathbf{z}, \mathcal{D}) \neq 1\right\} \tag{6}$$

$$= \sum_{\mathbf{z}} p_{\mathbf{Z}|\mathbf{X}}(\mathbf{z} \mid \mathbf{x}) \mathbb{1}\left\{\bigcup_{\tilde{\mathbf{u}} \in \mathcal{D} \setminus \{\mathbf{u}(1)\}} I(\mathbf{z}, \tilde{\mathbf{u}}) > I(\mathbf{z}, \mathbf{u}(1))\right\} \tag{7}$$

$$\leq \min\left\{\sum_{\mathbf{z}} p_{\mathbf{Z}|\mathbf{X}}(\mathbf{z} \mid \mathbf{x}) \sum_{\tilde{\mathbf{u}} \in \mathcal{D} \setminus \{\mathbf{u}(1)\}} \mathbb{1}\left\{I(\mathbf{z}, \tilde{\mathbf{u}}) > I(\mathbf{z}, \mathbf{u}(1))\right\}, 1\right\}, \tag{8}$$

where $\mathbf{u}(1)$ is the stored codeword corresponding to item 1. In (6), we condition on the values $\mathbf{z}$ can take. Equation (7) follows from the definition of our decoder; since a decoding error occurs if any one of the other stored codewords has a higher mutual information with $\mathbf{z}$ compared to $\mathbf{u}(1)$. Equation (8) is essentially the union bound and the $\min$ in (8) ensures that this bound is useful as an upper bound on a probability. For compactness, in the sequel we temporarily ignore the upper bound of unity and reintroduce it in the final steps of the analysis.

Next, let us look at the term $\Pr[\mathcal{D} \mid E_e^c, \mathbf{x}]$ in (5). Since the codewords are picked uniformly at random and since the only codeword in $\mathcal{D}$ that depends on E_e^c and $\mathbf{x}$ is the codeword corresponding to the first item, this can be factored as $\Pr[\mathbf{u}(1) \mid E_e^c, \mathbf{x}] \prod_{j=2}^{2^{nR_I}} \Pr[\mathbf{u}(j)]$. Substituting this and (8) in the inner sum of (5), we have the following.

$$\begin{aligned}&\sum_{\mathcal{D}} \Pr[\mathcal{D} \mid E_e^c, \mathbf{x}] \sum_{\mathbf{z}} p_{\mathbf{Z}|\mathbf{X}}\left(\mathbf{z} \mid \mathbf{x}\right) \sum_{\tilde{\mathbf{u}} \in \mathcal{D} \setminus \{\mathbf{u}(1)\}} \mathbb{1}\left\{I(\mathbf{z}, \tilde{\mathbf{u}}) > I(\mathbf{z}, \mathbf{u}(1))\right\} \\ &= \sum_{\mathbf{u}} \Pr\left[\mathbf{U}(1) = \mathbf{u} \mid E_e^c, \mathbf{x}\right] \sum_{\mathbf{z}} p_{\mathbf{Z}|\mathbf{X}}\left(\mathbf{z} \mid \mathbf{x}\right) \ldots \\ &\qquad \sum_{m=2}^{M} \sum_{\tilde{\mathbf{u}}} p_{\mathbf{U}(m)}(\tilde{\mathbf{u}}) \mathbb{1}\left\{I(\tilde{\mathbf{u}}, \mathbf{z}) > I(\mathbf{u}, \mathbf{z})\right\}\end{aligned} \tag{9}$$

Observe that in (9), we have distributed the product term over the subsequent summations. Using Baye's rule, we can analyze the term $\Pr\left[\mathbf{U}(1) = \mathbf{u} \mid E_e^c, \mathbf{x}\right]$

$$\Pr\left[\mathbf{U}(1) = \mathbf{u} \mid E_e^c, \mathbf{x}\right] = \frac{\Pr\left[E_e^c \mid \mathbf{x}, \mathbf{u}\right] \Pr[\mathbf{u} \mid \mathbf{x}]}{\Pr[E_e^c \mid \mathbf{x}]} \tag{10}$$

$$= \frac{\mathbb{1}\{\mathbf{u} \in \mathcal{T}_{S_{\mathbf{x}}}(\mathbf{x})\} \frac{1}{|\mathcal{T}_{Q_{\mathbf{x}}}|}}{\sum_{\mathbf{u}} \frac{1}{|\mathcal{T}_{Q_{\mathbf{x}}}|} \mathbb{1}\{\mathbf{u} \in \mathcal{T}_{S_{\mathbf{x}}}(\mathbf{x})\}} \tag{11}$$

$$= \mathbb{1}\{\mathbf{u} \in \mathcal{T}_{S_{\mathbf{x}}}(\mathbf{x})\} \frac{1}{|\mathcal{T}_{S_{\mathbf{x}}}(\mathbf{x})|} \tag{12}$$

Using equations (8), (9) and (12) in (5) gives us the inequality (13) below. In (14), we merely partition all summations based on corresponding marginal and conditional types.

$$\begin{aligned}&\Pr\left[E_d \mid E_e^c\right] \\ &\leq \sum_{\mathbf{x}} p_{\mathbf{X}}(\mathbf{x}) \sum_{\mathbf{u}} \frac{\mathbb{1}\{\mathbf{u} \in \mathcal{T}_{S_{\mathbf{x}}}(\mathbf{x})\}}{|\mathcal{T}_{S_{\mathbf{x}}}(\mathbf{x})|} \sum_{\mathbf{z}} p_{\mathbf{Z}|\mathbf{X}}\left(\mathbf{z} \mid \mathbf{x}\right) \\ &\quad \times \sum_{m=2}^{M} \sum_{\tilde{\mathbf{u}}} p_{\mathbf{U}(m)}(\tilde{\mathbf{u}}) \mathbb{1}\left\{I(\tilde{\mathbf{u}}, \mathbf{z}) > I(\mathbf{u}, \mathbf{z})\right\} \qquad (13) \\ &= \sum_{i=1}^{|\{P_i\}|} \sum_{\mathbf{x} \in \mathcal{T}_{P_i}} p_{\mathbf{X}}(\mathbf{x}) \sum_{\mathbf{u} \in \mathcal{T}_{S_i}(\mathbf{x})} \frac{1}{|\mathcal{T}_{S_i}(\mathbf{x})|} \sum_{V} \sum_{\mathbf{z} \in \mathcal{T}_V(\mathbf{u})} p_{\mathbf{Z}|\mathbf{X}}(\mathbf{z} \mid \mathbf{x}) \\ &\quad \times \sum_{m=2}^{M} \sum_{\tilde{\mathbf{u}}} p_{\mathbf{U}(m)}(\tilde{\mathbf{u}}) \mathbb{1}\left\{I(\tilde{\mathbf{u}}, \mathbf{z}) > I(Q_i, V)\right\} \qquad (14)\end{aligned}$$

To proceed, we bound the contribution from the final terms, starting with the sum over incorrect codewords. First we rewrite $Q_i(a)V(c|a) = Q_iV(c)\tilde{V}(a|c)$ where $Q_iV(c)$ is the induced marginal distribution on the $\mathcal{Z}$ alphabet.

$$\sum_{m=2}^{M} \sum_{\tilde{\mathbf{u}}} p_{\mathbf{U}(m)}(\tilde{\mathbf{u}}) \mathbb{1}\left\{I(\tilde{\mathbf{u}}, \mathbf{z}) > I(Q_i, V)\right\}$$

$$= \sum_{m=2}^{M} \sum_{\bar{V}} \sum_{\tilde{\mathbf{u}} \in \mathcal{T}_{\bar{V}}(\mathbf{z})} p_{\mathbf{U}(m)}(\tilde{\mathbf{u}}) \mathbb{1}\left\{I(\tilde{\mathbf{u}}, \mathbf{z}) > I(Q_i, V)\right\} \tag{15}$$

$$= \sum_{m=2}^{M} \sum_{\substack{\bar{V}: \\ I(Q_iV, \bar{V}) \geq I(Q_iV, \tilde{V})}} \Pr[\tilde{\mathbf{U}}(m) \in \mathcal{T}_{\bar{V}}(\mathbf{z})] \tag{16}$$

$$= \sum_{m=2}^{M} \sum_{\substack{\bar{V}: \\ I(Q_iV, \bar{V}) \geq I(Q_iV, \tilde{V})}} \frac{|\mathcal{T}_{\bar{V}(\mathbf{z})}|}{|\mathcal{T}_{\mathbf{u}(m)}|} \tag{17}$$

where in (15), we partition the summation over $\mathbf{u}$ based on the possible conditional types $\bar{V} : \mathcal{Z} \to \mathcal{U}$. In (16), we merely re-index the summation over $\bar{V}$ based on the indicator function and in (17), we use the fact that the codewords are drawn uniformly at random from their respective marginal types. For the next step, recall that when we write $Q_{\mathbf{x}(m)}$, we mean the marginal type induced on $\mathcal{U}$ corresponding to the m-th enrollment vector. Also, observe that the type class $\mathcal{T}_{\bar{V}(\mathbf{z})}$ is either empty (if $\sum_c Q_i V(c)\bar{V}(a|c) \neq Q_{\mathbf{x}(m)}(a)$ for all $a \in \mathcal{U}$) or, if non-empty, is upper bounded as $2^{nH(\bar{V}|Q_iV)}$. Continuing we get

$$\begin{aligned}
&\sum_{m=2}^{M} \sum_{\substack{\bar{V}: \\ I(Q_iV,\bar{V}) \geq I(Q_iV,\tilde{V})}} \frac{|\mathcal{T}_{\bar{V}(\mathbf{z})}|}{|\mathcal{T}_{\mathbf{u}(m)}|} \\
&\quad \leq \sum_{m=2}^{M} \sum_{\substack{\bar{V}: \\ I(Q_iV,\bar{V}) \geq I(Q_iV,\tilde{V})}} \frac{2^{nH(\bar{V}|Q_iV)}}{(n+1)^{|\mathcal{U}|} 2^{nH(Q_{\mathbf{x}(m)})}} \\
&\quad = \sum_{m=2}^{M} \sum_{\substack{\bar{V}: \\ I(Q_iV,\bar{V}) \geq I(Q_iV,\tilde{V})}} (n+1)^{-|\mathcal{U}|} 2^{-nI(Q_iV,\bar{V})} \quad (18) \\
&\quad \leq (n+1)^{|\mathcal{U}||\mathcal{Z}|-|\mathcal{U}|} 2^{-n|I(Q_i,V)-R_I|^+}, \quad (19)
\end{aligned}$$

where in the last step we recall that $I(Q_iV, \tilde{V}) = I(Q_i, V)$. Also, we use the notation $|\ x\ |^+$ to mean $\min\{x, 0\}$ and this is where we reintroduce the trivial upper bound of unity from (8). Substituting this bound into (14) we have

$$\begin{aligned}
\Pr[E_d \mid E_e^c] &\leq (n+1)^{|\mathcal{U}||\mathcal{Z}|-|\mathcal{U}|} \sum_{i=1}^{|\{P_i\}|} \sum_{\mathbf{x}\in\mathcal{T}_{\mathbf{P}_i}} p_{\mathbf{X}}(\mathbf{x}) \sum_{\mathbf{u}\in\mathcal{T}_{S_i}(\mathbf{x})} \frac{1}{\mathcal{T}_{S_i}(\mathbf{x})} \\
&\quad \times \sum_{V} \sum_{\mathbf{z}\in\mathcal{T}_V(\mathbf{u})} p_{\mathbf{Z}|\mathbf{X}}(\mathbf{z} \mid \mathbf{x}) 2^{-n|I(Q_i,V)-R_I|^+} \quad (20) \\
&\leq (n+1)^{|\mathcal{U}||\mathcal{Z}|-|\mathcal{U}|} \sum_{i=1}^{|\{P_i\}|} \sum_{\mathbf{x}\in\mathcal{T}_{\mathbf{P}_i}} p_{\mathbf{X}}(\mathbf{x}) \sum_{\mathbf{u}\in\mathcal{T}_{S_i}(\mathbf{x})} \frac{1}{|\mathcal{T}_{S_i}(\mathbf{x})|} \\
&\quad \times \sum_{V} 2^{-nD\left(\tilde{J}^*_{V,i}\|J|Q_i\right)} 2^{-n|I(Q_i,V)-R_I|^+}, \quad (21)
\end{aligned}$$

where in the second step we apply Lemma 2 and we use $\tilde{J}^*_{V,i}$ in the place of $\tilde{J}^*_{V,P_i\times S_i}$ since the type $P_i \times S_i$ is indexed by i. Now we further upper-bound via the worst-case end-to-end behavior V for any choice of S_i. Letting $V_i^* = \arg\min D\left(\tilde{J}^*_{V,i}\|J \mid Q_i\right) + |I(Q_i, V) - R_I|^+$ we have

$$\begin{aligned}
&\Pr[E_d \mid E_e^c] \\
&\quad \leq (n+1)^{2|\mathcal{U}||\mathcal{Z}|-|\mathcal{U}|} \sum_{i=1}^{|\{P_i\}|} \sum_{\mathbf{x}\in\mathcal{T}_{\mathbf{P}_i}} p_{\mathbf{X}}(\mathbf{x}) \sum_{\mathbf{u}\in\mathcal{T}_{S_i}(\mathbf{x})} \frac{1}{|\mathcal{T}_{S_\mathbf{x}}(\mathbf{x})|} \\
&\qquad \times 2^{-n\left[D\left(\tilde{J}^*_{V_i^*,i}\|J|Q_i\right)+|I(Q_i,V_i^*)-R_I|^+\right]},
\end{aligned}$$

This bound holds for all choices of S_i. Thus, we pick the $\{S_i\}$ to maximize the exponent (for each source type P_i), subject to the constraint that $I(P_i, S_i) < R_C$. Denote each such choice as S_i^* and let $Q_i^*(a) = \sum_b P_i(b)S_i^*(a|b)$. Using these choices we get equation (22) (which is shown on the top of the following page).

Finally, we minimize the exponent in (23) over the worst case source distribution P_{i^*} to get

$$\begin{aligned}
\Pr[E_d \mid E_e^c] \leq &\rho(n) 2^{-nD(P_{i^*}\|p_X)+D\left(\tilde{J}^*_{V^*_{i^*},i}\|J|Q_i^*\right)} \\
&\times 2^{-n|I(Q^*_{i^*},V^*_{i^*})-R_I|^+]}.
\end{aligned}$$

where $\rho(n)$ is a polynomial in n which is always positive. To conclude the proof, we relax the range of each of the above optimizations to distributions instead of types. That this can be done follows from the fact that collection of $n-$types is dense in the set of all distributions and we omit a rigorous proof of this due to space constraints.

Appendix

A. Proof of Lemma 1

The probability that an encoding error occurs is the probability that for a randomly chosen X^n, one is unable to find a codeword in $\mathcal{C}_\mathbf{x}$ which is in $\mathcal{T}_{S_\mathbf{x}}(\mathbf{x})$ Observe that this can be written as follows

$$\begin{aligned}
&\Pr[\text{Encoding Error}] \\
&\quad = \sum_{i=1}^{|\{P_i\}|} \sum_{\mathbf{x}\in\mathcal{T}_{\mathbf{P}_i}} p_{\mathbf{X}}(\mathbf{x}) \Pr[\nexists \mathbf{u} \in \mathcal{C}_i \text{ such that } \mathbf{u} \in \mathcal{T}_{S_i}(\mathbf{x})] \\
&\quad = \sum_{i=1}^{|\{P_i\}|} \sum_{\mathbf{x}\in\mathcal{T}_{\mathbf{P}_i}} p_{\mathbf{X}}(\mathbf{x}) \left[1 - \Pr\left\{\tilde{\mathbf{U}} \in \mathcal{T}_{S_i}(\mathbf{x})\right\}\right]^{2^{n\mathbf{R}_c}} \quad (24)
\end{aligned}$$

where the last step follows from the fact that all codewords are drawn independently and uniformly from $\mathcal{T}_{Q_i}$. Now, we rewrite $P_i(b)S_i(a \mid b) = Q_i(a)\bar{S}_i(b \mid a)$. Since $\tilde{\mathbf{U}}$ is drawn uniformly from $\mathcal{T}_{Q_i}$, we have

$$\begin{aligned}
\Pr\{\tilde{\mathbf{U}} \in \mathcal{T}_{S_i}(\mathbf{x})\} &= \frac{|\{\tilde{\mathbf{u}} \in \mathcal{T}_{Q_i}, \mathbf{x} \in \mathcal{T}_{\bar{S}_i}(\tilde{\mathbf{u}})\}|}{|\{\mathcal{T}_{Q_i}\}|} \\
&\geq \frac{(n+1)^{-|\mathcal{X}||\mathcal{U}|} 2^{nH(\bar{S}_i|Q_i))}}{2^{nH(Q_i)}} = (n+1)^{-|\mathcal{X}||\mathcal{U}|} 2^{-nI(P_i,S_i)} \quad (25)
\end{aligned}$$

where in the last step we have used the fact that $P_i(b)S_i(a \mid b) = Q_i(a)\bar{S}_i(b \mid a)$. Using (25) and the inequality $(1-x)^n < 2^{-nx\log_2 e}$ in (24) immediately gives us the result after we apply standard typicality bounds to $p_{\mathbf{X}}(\mathbf{x})$.

B. Proof of Lemma 2

We first introduce some notation. For a length - n vector $\mathbf{z}$, we use $\mathbf{z}(k)$ and $\mathbf{z}_k$ interchangeably to denote the k-th value. If $\mathcal{I} \subset \{1, 2, \ldots, n\}$, then we write $\mathbf{z}_\mathcal{I}$ to denote the concatenation of $\mathbf{z}_k$ for $k \in \mathcal{I}$. Finally, we define $U(a) := \{\ell : \mathbf{u}_\ell = a\}$. Observe that the $\mathcal{T}_V(\mathbf{u})$ can be written using the above notation as follows.

$$\begin{aligned}
\mathcal{T}_V(\mathbf{u}) = \Big\{\mathbf{z} \in \mathcal{Z}^n : &\frac{|\ U(a_\ell) \cap \{k : \mathbf{z}_k = c_j\}\ |}{nQ_\mathbf{u}(a_\ell)} = V(c_j \mid a_l), \\
&\forall \ell \in \{1, \ldots, |\mathcal{U}|\}, j \in \{1, \ldots, |\mathcal{Z}|\}\Big\}
\end{aligned}$$

$$\Pr[E_d \mid E_e^c] \leq (n+1)^{2|\mathcal{U}||\mathcal{Z}|-|\mathcal{U}|} \sum_{i=1}^{|\{P_i\}|} \sum_{\mathbf{x}\in\mathcal{T}_{P_i}} p_{\mathbf{X}}(\mathbf{x}) \sum_{\mathbf{u}\in\mathcal{T}_{S_i^*}(\mathbf{x})} \frac{1}{|\mathcal{T}_{S_{\mathbf{x}}}(\mathbf{x})|} 2^{-n[D\left(\tilde{J}^*_{V_i^*,i}\|J|Q_i^*\right)+|I(Q_i^*,V_i^*)-R_I|^+]}, \tag{22}$$

$$\leq (n+1)^{2|\mathcal{U}||\mathcal{Z}|-|\mathcal{U}|} \sum_{i=1}^{|\{P_i\}|} 2^{-nD(P_i\|p_X)} 2^{-n[D\left(\tilde{J}^*_{V_i^*,i}\|J|Q_i^*\right)+|I(Q_i^*,V_i^*)-R_I|^+]}, \tag{23}$$

Let $\mathcal{M}_\ell(\mathbf{u}), 1 \leq \ell \leq |\mathcal{U}|$ denote the following set

$$\left\{\mathbf{z} \in \mathcal{Z}^{|U(a_\ell)|} : \frac{|\{k : \mathbf{z}_k = c_j\}|}{nQ_{\mathbf{u}}(a_\ell)} = V(c_j \mid a_\ell) \quad \forall j \in \{1,\ldots,|\mathcal{Z}|\}\right\}$$

and observe that since the sets $U(a_\ell), 1 \leq \ell \leq |\mathcal{U}|$ partition $\{1,2,\ldots,n\}$, we have $\mathbb{1}\{\mathbf{z} \in \mathcal{T}_V(\mathbf{u})\} = \prod_{\ell=1}^{|\mathcal{U}|} \mathbb{1}\{\mathbf{z}_{U(a_\ell)} \in \mathcal{M}_\ell(\mathbf{u})\}$. In the sequel, we write $W(\mathbf{z} \mid \mathbf{x})$ instead of $W^n(\mathbf{z} \mid \mathbf{x})$ where the value of n is clear from context and we suppress the dependence of $\mathcal{M}_\ell$ on $\mathbf{u}$. Now, consider the quantity we wish to bound and expand as follows.
.

$$\begin{aligned}
&W[\mathcal{T}_V(\mathbf{u}) \mid \mathbf{x}] \\
&= \sum_{\mathbf{z}\in\mathcal{Z}^n} W(\mathbf{z} \mid \mathbf{x}) \mathbb{1}\{\mathbf{z} \in \mathcal{T}_V(\mathbf{u})\} \\
&= \sum_{\mathbf{z}\in\mathcal{Z}^n} \prod_{k=1}^{n} W(\mathbf{z}_k \mid \mathbf{x}_k) \prod_{\ell=1}^{|\mathcal{U}|} \mathbb{1}\{\mathbf{z}_{U(a_\ell)} \in \mathcal{M}_\ell\} \\
&= \sum_{\mathbf{z}\in\mathcal{Z}^n} \prod_{\ell=1}^{|\mathcal{U}|} W\left(\mathbf{z}_{U(a_\ell)} \mid \mathbf{x}_{U(a_\ell)}\right) \mathbb{1}\{\mathbf{z}_{U(a_\ell)} \in \mathcal{M}_\ell\} \\
&= \prod_{\ell=1}^{|\mathcal{U}|} \sum_{\mathbf{z}_{U(\bullet_\ell)}} W\left(\mathbf{z}_{U(a_\ell)} \mid \mathbf{x}_{U(a_\ell)}\right) \mathbb{1}\{\mathbf{z}_{U(a_\ell)} \in \mathcal{M}_\ell\} \\
&= \prod_{\ell=1}^{|\mathcal{U}|} \sum_{\tilde{W}} \sum_{\mathbf{z}_{U(\bullet_\ell)} \in \mathcal{T}_{\tilde{W}}(\mathbf{x}_{U(\bullet_\ell)})} W\left(\mathbf{z}_{U(a_\ell)} \mid \mathbf{x}_{U(a_\ell)}\right) \mathbb{1}\{\mathbf{z}_{U(a_\ell)} \in \mathcal{M}_l\} \\
&= \prod_{\ell=1}^{|\mathcal{U}|} \sum_{\tilde{W}\in\mathcal{E}_\ell} \sum_{\mathbf{z}_{U(\bullet_\ell)} \in \mathcal{T}_{\tilde{W}}(\mathbf{x}_{U(\bullet_\ell)})} W\left(\mathbf{z}_{U(a_\ell)} \mid \mathbf{x}_{U(a_\ell)}\right) && (26) \\
&= \prod_{\ell=1}^{|\mathcal{U}|} \sum_{\tilde{W}\in\mathcal{E}_\ell} W^{|U(a_\ell)|}\left(\mathcal{T}_{\tilde{W}}(\mathbf{x}_{U(a_\ell)}) \mid \mathbf{x}_{U(a_\ell)}\right) && (27) \\
&\leq \prod_{\ell=1}^{|\mathcal{U}|} \sum_{\tilde{W}\in\mathcal{E}_\ell} 2^{-|U(a_\ell)| D(\tilde{W}\|W|\bar{S}(\cdot|a_\ell))} && (28) \\
&\leq \rho(n) \prod_{\ell=1}^{|\mathcal{U}|} 2^{-|U(a_l)| D(\tilde{W}^*_{\ell,V,P_{\mathbf{x}}\times S}\|W|\bar{S}(\cdot|a_l))} && (29) \\
&= \rho(n) 2^{-n\sum_{\ell=1}^{|\mathcal{U}|} Q_{\mathbf{u}}(a_\ell) D(\tilde{W}^*_{\ell,V,P_{\mathbf{x}}\times S}\|W|\bar{S}(\cdot|a_l))} && (30)
\end{aligned}$$

In the third step we merely combine terms that correspond to the indices in $U(a_\ell)$ for each ℓ. The fact that each term in the product involves a sub-vector of $\mathbf{z}$ that is disjoint from other sub-vectors allows us to switch the sum and product in the fourth step. In (26), we define $\mathcal{E}_\ell$ as the set of all $\tilde{W}$ such that $\sum_{b\in\mathcal{X}} \bar{S}(b \mid a)\tilde{W}(c \mid b) = V(c \mid a), \forall(a,c) \in \mathcal{U}\times\mathcal{X}$ and $\bar{S}$ is such that $P_{\mathbf{x}} \times S = Q_{\mathbf{u}} \times \bar{S}$. Observe that the marginal type of $\mathbf{x}_{U(a_\ell)}$ is given by $\bar{S}(\cdot \mid a_\ell)$. Therefore, in (27), the term $\sum_{\mathbf{z}_{U(\bullet_l)}\in\mathcal{T}_{\tilde{W}}(\mathbf{x}_{U(\bullet_l)})} W\left(\mathbf{z}_{U(a_l)} \mid \mathbf{x}_{U(a_l)}\right)$ can be written as $W^{|U(a_l)|}\left(\mathcal{T}_{\tilde{W}}\left(\mathbf{x}_{U(a_l)}\right) \mid \mathbf{x}_{U(a_l)}\right)$ per the notation of [6]. From this, we get (28) using standard bounds. Finally defining $\tilde{W}^*_{\ell,V,P_{\mathbf{x}}\times S} := \arg\min_{\tilde{W}\in\mathcal{E}_\ell} D(\tilde{W}\|W \mid \bar{S}(\cdot \mid a_\ell))$ and $\rho(n)$ as some positive polynomial in n gives us (29) of which (30) is a direct consequence once we make the substitution $|U(a_\ell)| = nQ_{\mathbf{u}}(a_\ell)$.

To continue, we consider the exponent in (30)

$$\begin{aligned}
&\sum_{a\in\mathcal{U}} Q_{\mathbf{u}}(a) D(\tilde{W}^*_a\|W \mid \bar{S}(\cdot \mid a)) \\
&\qquad = \sum_{a\in\mathcal{U}} Q_{\mathbf{u}}(a) D(\tilde{W}^*_a \bar{S}(\cdot \mid a)\|W\bar{S}(\cdot \mid a)) && (31) \\
&\qquad = \sum_{a\in\mathcal{U}} Q_{\mathbf{u}}(a) D(\tilde{J}(\cdot \mid a)\|J(\cdot \mid a)) && (32) \\
&\qquad = D(\tilde{J}\|J \mid Q_{\mathbf{u}}) && (33)
\end{aligned}$$

where in (32), $\tilde{J} : \mathcal{U} \to \mathcal{X}\times\mathcal{Z}, J : \mathcal{U} \to \mathcal{X}\times\mathcal{Z}$ are stochastic matrices (or conditional types) defined as follows $\tilde{J}(b,c \mid a) = \tilde{W}^*_a(c \mid b)\bar{S}(b \mid a) and J(b,c \mid a) = W(c \mid b)\bar{S}(b \mid a)$. Finally, using (33) in (30) and rewriting the optimizations in terms of $\tilde{J}$ concludes the proof.

REFERENCES

[1] E. Tuncel. Capacity/storage tradeoff in high-dimensional identification systems. *IEEE Trans. Inform. Theory*, 55:2097–2106, November 2009.

[2] M. B. Westover and J. A. O'Sullivan. Achievable rates for pattern recognition. *IEEE Trans. Inform. Theory*, 54:299–320, Jan 2008.

[3] S. Jayaraman and T. Berger. An error exponent for lossy source coding with side information at the decoder. In *Proc. Int. Symp. Inform. Theory*, page 263, Whisler, BC, Canada, 1995.

[4] B. Kelly and A. B. Wagner. Error exponents and test channel optimization for the Wyner-Ziv problem. In *Proc. Allerton Conf. on Communication, Control and Computing*, pages 510–517, September 2007.

[5] A. Somekh-Baruch and N. Merhav. On the random coding error exponents of the single-user and the multiple-access Gel'fand-Pinsker channels. In *Proc. Int. Symp. Inform. Theory*, page 448, Chicago, IL, USA, June 2004.

[6] I. Csiszár and J. Körner. *Information Theory, Coding Theorems for Discrete Memoryless Systems*. Akadémiai Kiadó, 1981.