Source Coding, Lists, and Rényi Entropy

Christoph Bunte and Amos Lapidoth
ETH Zurich
Switzerland
Email: {bunte,lapidoth}@isi.ee.ethz.ch

***Abstract*—A sequence produced by a memoryless source is to be described using a fixed number of bits that is proportional to its length. Based on the description, a list that is guaranteed to contain the sequence must be produced. The trade-off between the description length and the moments of the listsize is studied when the sequence's length tends to infinity. It is characterized by the source's Rényi entropy. Extensions to scenarios with side information are also studied, where the key is conditional Rényi entropy. The lossy case where at least one of the elements of the list must be within a specified distortion from the source sequence is also solved.**

I. Introduction

Consider a discrete memoryless source (DMS) emitting IID symbols $X_1, X_2, \dots$ from a finite alphabet $\mathcal{X}$ with distribution P. A *rate-R blocklength-n list code* for this source is a pair of encoder/decoder mappings

$$f\colon \mathcal{X}^n \to \{1,\dots,e^{nR}\}, \quad \varphi\colon \{1,\dots,e^{nR}\} \to 2^{\mathcal{X}^n}, \tag{1}$$

where $2^{\mathcal{X}^n}$ denotes the power set of $\mathcal{X}^n$, i.e., the collection of all subsets of $\mathcal{X}^n$. We refer to $\varphi(f(\mathbf{x}))$ as the list produced by the decoder when the source emits the sequence $\mathbf{x}$. We say that the code is *lossless* if the list produced by the decoder always contains the sequence emitted by the source, i.e., if

$$\mathbf{x} \in \varphi(f(\mathbf{x})), \quad \mathbf{x} \in \mathcal{X}^n. \tag{2}$$

The "listsize", i.e., the cardinality of the list produced by the decoder, is a nonnegative integer-valued random variable (RV). For a lossless code, this RV is at least one, and if the rate R is below the entropy $H(P)$, then the probability that it is one tends to zero as the blocklength tends to infinity. We refer to its ρ-th moment as the ρ-th moment of the list.

In this paper we study the trade-off between the rate and the ρ-th moment of the list. Specifically, for a given $\rho > 0$ we find the smallest rate of lossless list codes for which the ρ-th moment of the list tends to one as the blocklength tends to infinity. In Section IV we show that this rate equals the *Rényi entropy of the source of order* $1/(1+\rho)$

$$H_{\frac{1}{1+\rho}}(P) \triangleq \frac{1}{\rho} \log \left[\sum_{x\in\mathcal{X}} P(x)^{\frac{1}{1+\rho}} \right]^{1+\rho}. \tag{3}$$

This gives a new operational characterization of Rényi entropy of all orders smaller than one.[1]

In Section V we generalize our result to a setting where side-information is available to the encoder and decoder. In this case the answer is given by a conditional version of Rényi entropy, a version that was proposed by Arimoto [2].

We also consider a lossy version of the problem: Suppose we are given a finite reconstruction alphabet $\hat{\mathcal{X}}$ and a distortion function $d\colon \mathcal{X} \times \hat{\mathcal{X}} \to [0,\infty)$ that extends to length-n sequences in the usual way:

$$d(\mathbf{x}, \hat{\mathbf{x}}) = \frac{1}{n} \sum_{j=1}^{n} d(x_j, \hat{x}_j), \quad (\mathbf{x}, \hat{\mathbf{x}}) \in \mathcal{X}^n \times \hat{\mathcal{X}}^n. \tag{4}$$

We assume that for every $x \in \mathcal{X}$ there is at least one $\hat{x} \in \hat{\mathcal{X}}$ such that $d(x,\hat{x}) = 0$. A *rate-R blocklength-n distortion-D list code* is a pair of encoder/decoder mappings

$$f\colon \mathcal{X}^n \to \{1,\dots,e^{nR}\}, \quad \varphi\colon \{1,\dots,e^{nR}\} \to 2^{\hat{\mathcal{X}}^n}, \tag{5}$$

having the property that for every $\mathbf{x} \in \mathcal{X}^n$ there is at least one $\hat{\mathbf{x}} \in \varphi(f(\mathbf{x}))$ satisfying $d(\mathbf{x},\hat{\mathbf{x}}) \le D$. We characterize the smallest rate of distortion-D codes for which the ρ-th moment of the list approaches one as the blocklength tends to infinity. In Section VI we show that this rate equals $R_\rho(D)$, where

$$R_\rho(D) \triangleq \max_{Q} \bigl\{ R(Q,D) - \rho^{-1} D(Q||P) \bigr\}, \quad \rho > 0, \tag{6}$$

and where $R(Q,D)$ is the rate-distortion function (see, e.g., [3, Chapter 7]) evaluated at the distortion level D for the source Q and the distortion function d. The function $R_\rho(D)$ has previously appeared in [4] in the context of guessing.

The rest of this paper is organized as follows. Section II introduces some notation; Section III contains a lemma that is key to all the converse proofs in this paper; and Sections IV through VI contain the main results and their proofs. We conclude with some remarks in Section VII.

II. Preliminaries and Notation

The cardinality of a finite set $\mathcal{X}$ is denoted by $|\mathcal{X}|$. If P is a PMF on $\mathcal{X}$, then P^n denotes the product PMF on $\mathcal{X}^n$

$$P^n(\mathbf{x}) = \prod_{j=1}^{n} P(x_j), \quad \mathbf{x} \in \mathcal{X}^n. \tag{7}$$

The support of P is denoted by supp(P), so

$$\text{supp}(P) = \bigl\{ x \in \mathcal{X} : P(x) > 0 \bigr\}. \tag{8}$$

If $V(\cdot|x)$ is a PMF on $\mathcal{Y}$ for every $x \in \mathcal{X}$, then $P \circ V$ denotes the induced joint PMF

$$(P \circ V)(x,y) = P(x)V(y|x), \quad (x,y) \in \mathcal{X} \times \mathcal{Y}. \tag{9}$$

The ceiling of a real number ξ, i.e., the smallest integer no smaller than ξ, is denoted by $\lceil \xi \rceil$. The collection of all PMFs on $\mathcal{X}$ is denoted by $\mathcal{P}(\mathcal{X})$. The set of types of sequences in $\mathcal{X}^n$, i.e., the set of rational PMFs with denominator n, is

[1] For other operational characterizations of Rényi entropy see [1] and references therein.

denoted by $\mathcal{P}_n(\mathcal{X})$. For information theoretic quantities, we adopt the notation in [3]. We frequently use basic results from the Method of Types [3, Chapter 2]. All logarithms are natural logarithms.

III. A Key Lemma

The following lemma is the key to all converse proofs in this paper. It is inspired by [5, Theorem 1].

Lemma III.1. *Let P be a PMF on a nonempty finite set $\mathcal{X}$, and let $\mathcal{L}_1, \ldots, \mathcal{L}_M$ be a partition of $\mathcal{X}$ into M lists, i.e., $\mathcal{L}_m \cap \mathcal{L}_{m'} = \emptyset$ if $m \neq m'$ and $\bigcup_{m=1}^{M} \mathcal{L}_m = \mathcal{X}$. For every $x \in \mathcal{X}$ let $L(x)$ denote the cardinality of the list containing x. Then*

$$\sum_{x \in \mathcal{X}} P(x) L^{\rho}(x) \geq \frac{1}{M^{\rho}} \left[\sum_{x \in \mathcal{X}} P(x)^{\frac{1}{1+\rho}} \right]^{1+\rho}, \quad \rho \geq 0. \quad (10)$$

Proof: The crucial observation is that

$$\sum_{x \in \mathcal{X}} \frac{1}{L(x)} = M. \quad (11)$$

Otherwise the proof follows similar steps as the proof of [5, Theorem 1] with $L(x)$ taking the role of the "guessing function" $G(x)$. ■

IV. Lossless Source Coding with Lists

For any $\rho > 0$, we say that a rate R is ρ-achievable with lossless list source coding if there exists a sequence $(f_n, \varphi_n)_{n \geq 1}$ of lossless rate-R blocklength-n list codes (see Section I) such that

$$\lim_{n \to \infty} \sum_{\mathbf{x} \in \mathcal{X}^n} P^n(\mathbf{x}) \big| \varphi_n(f_n(\mathbf{x})) \big|^{\rho} = 1. \quad (12)$$

Theorem IV.1. *For a DMS P, the infimum of all rates that are ρ-achievable with lossless list source coding is the Rényi entropy of P of order $1/(1+\rho)$ (see* (3)*).*

Theorem IV.1 reveals many of the known properties of Rényi entropy. For example, $H_{\frac{1}{1+\rho}}(P)$ is nondecreasing in ρ because any ρ-achievable rate is also ρ'-achievable for all $\rho' < \rho$. Also, as we next show,

$$H(P) \leq H_{\frac{1}{1+\rho}}(P) \leq \log|\text{supp}(P)|. \quad (13)$$

Indeed, if $R < H(P)$, then the probability that the listsize is at least 2 tends to one as $n \to \infty$, and any $R > \log|\text{supp}(P)|$ is ρ-achievable for all $\rho > 0$. The limit

$$\lim_{\rho \to \infty} H_{\frac{1}{1+\rho}}(P) = \log|\text{supp}(P)| \quad (14)$$

can be explained as follows. If $R < \log|\text{supp}(P)|$, then there must exist some $\mathbf{x}_0 \in \text{supp}(P)^n$ for which $|\varphi_n(f_n(\mathbf{x}_0))| \geq e^{n(\log|\text{supp}(P)| - R)}$. Since $P^n(\mathbf{x}_0) \geq p_{\min}^n$, where $p_{\min}$ denotes the smallest nonzero probability of any source symbol,

$$\sum_{\mathbf{x}} P^n(\mathbf{x}) |\varphi_n(f_n(\mathbf{x}))|^{\rho} \geq e^{n\rho(\log|\text{supp}(P)| - R - \frac{1}{\rho} \log \frac{1}{p_{\min}})}.$$

The RHS tends to infinity when ρ is sufficiently large, and thus for such ρ the rate R cannot be ρ-achievable. As to the limit when ρ approaches zero, note that if $R > H(P)$, then the probability of the listsize exceeding one can be driven to zero exponentially fast. And since $|\varphi_n(f_n(\mathbf{x}))|^{\rho}$ is upper-bounded by $e^{n\rho \log|\mathcal{X}|}$, the ρ-th moment of the list will tend to one if ρ is sufficiently small so as to guarantee that the product of the exponents decay to zero. Thus,

$$\lim_{\rho \to 0} H_{\frac{1}{1+\rho}}(P) = H(P). \quad (15)$$

Proof of Theorem IV.1: Direct Part. The encoder first describes the type Q of the sequence emitted by the source. Since the number of types $|\mathcal{P}_n(\mathcal{X})|$ is a polynomial in n, this requires an asymptotically negligible amount of rate. Since the set of sequences of type Q has cardinality at most $e^{nH(Q)}$, we may partition it into e^{nR} lists of lengths at most

$$\lceil e^{n(H(Q) - R)} \rceil. \quad (16)$$

Using Lemma A.1 (in the Appendix) to upper-bound the ρ-th power of (16) (to account for the ceiling), and using the fact that the probability of the source emitting a sequence of type Q is at most $e^{-nD(Q||P)}$, we can upper-bound the ρ-th moment of the length of the list containing the source sequence by

$$1 + 2^{\rho} \sum_{Q \in \mathcal{P}_n(\mathcal{X})} \exp\Big(-n\rho\big(R + \rho^{-1} D(Q||P) - H(Q)\big)\Big). \quad (17)$$

Upper-bounding the summand in (17) by its maximum over all $Q \in \mathcal{P}(\mathcal{X})$ and using the identity [5]

$$H_{\frac{1}{1+\rho}}(P) = \max_{Q \in \mathcal{P}(\mathcal{X})} \big\{ H(Q) - \rho^{-1} D(Q||P) \big\}, \quad (18)$$

we obtain that (17) is upper-bounded by

$$1 + 2^{\rho} \sum_{Q \in \mathcal{P}_n(\mathcal{X})} \exp\big(-n\rho(R - H_{\frac{1}{1+\rho}}(P))\big). \quad (19)$$

Since $|\mathcal{P}_n(\mathcal{X})|$ is a polynomial in n, we can rewrite (19) as

$$1 + \exp\big(-n\rho(R - H_{\frac{1}{1+\rho}}(P) - \delta_n)\big), \quad (20)$$

where $\delta_n \to 0$ as $n \to \infty$. This completes the direct part because (20) tends to one as $n \to \infty$ provided that

$$R > H_{\frac{1}{1+\rho}}(P). \quad (21)$$

Converse. Fix a sequence $(f_n, \varphi_n)_{n \geq 1}$ of rate-R blocklength-n lossless list codes. We may assume that $\varphi_n(m) = f_n^{-1}(\{m\})$ for every $m \in \{1, \ldots, e^{nR}\}$, i.e., that the list $\varphi_n(m)$ comprises the source sequences that are mapped to m by the encoder f_n. Indeed, the lossless property implies that $f_n^{-1}(\{m\}) \subseteq \varphi_n(m)$, and replacing $\varphi_n(m)$ with $f_n^{-1}(\{m\})$ can only reduce the ρ-th moment of the list while preserving the lossless property of the code. With this assumption the lists $\{\varphi_n(m)\}_{1 \leq m \leq e^{nR}}$ form a partition of $\mathcal{X}^n$, and we may invoke Lemma III.1 to obtain

$$\begin{aligned} \sum_{\mathbf{x}} P^n(\mathbf{x}) |\varphi_n(f_n(\mathbf{x}))|^{\rho} &\geq e^{-n\rho R} \left[\sum_{\mathbf{x}} P^n(\mathbf{x})^{\frac{1}{1+\rho}} \right]^{1+\rho} \\ &= \exp\big(-n\rho(R - H_{\frac{1}{1+\rho}}(P)\big). \quad (22) \end{aligned}$$

This completes the converse because the RHS of (22) tends to infinity as $n \to \infty$ unless $R \geq H_{\frac{1}{1+\rho}}(P)$. ■

V. Lossless Source Coding With Lists and Side-Information

Suppose that a DMS emits a sequence of pairs of chance variables $(X_1, Y_1), (X_2, Y_2), \ldots$ drawn IID according to a PMF $P_{X,Y}$ on $\mathcal{X} \times \mathcal{Y}$, where $\mathcal{X}$ and $\mathcal{Y}$ are both finite. We wish to describe the sequence $(X_1, \ldots, X_n)$ using the information provided by the sequence $(Y_1, \ldots, Y_n)$. In this setting, a rate-R blocklength-n list code comprises an encoder mapping of the form

$$f \colon \mathcal{X}^n \times \mathcal{Y}^n \to \{1, \ldots, e^{nR}\}, \tag{23}$$

and a decoder mapping of the form

$$\varphi \colon \{1, \ldots, e^{nR}\} \times \mathcal{Y}^n \to 2^{\mathcal{X}^n}. \tag{24}$$

The code is lossless if $\mathbf{x} \in \varphi(f(\mathbf{x}, \mathbf{y}), \mathbf{y})$ for all $(\mathbf{x}, \mathbf{y}) \in \mathcal{X}^n \times \mathcal{Y}^n$. For any $\rho > 0$, we say that a rate R is ρ-achievable with lossless list source coding if there exists a sequence $(f_n, \varphi_n)_{n \geq 1}$ of rate-R blocklength-n lossless list codes such that

$$\lim_{n \to \infty} \sum_{\mathbf{x} \in \mathcal{X}^n} \sum_{\mathbf{y} \in \mathcal{Y}^n} P^n_{X,Y}(\mathbf{x}, \mathbf{y}) \bigl|\varphi_n(f_n(\mathbf{x}, \mathbf{y}), \mathbf{y})\bigr|^\rho = 1. \tag{25}$$

Theorem V.1. *In the presence of side-information, the infimum of all rates that are ρ-achievable with lossless list source coding is the conditional Rényi entropy of order $1/(1+\rho)$ of X given Y*

$$H_{\frac{1}{1+\rho}}(X|Y) \triangleq \frac{1}{\rho} \log \sum_{y \in \mathcal{Y}} \biggl[\sum_{x \in \mathcal{X}} P_{X,Y}(x, y)^{\frac{1}{1+\rho}} \biggr]^{1+\rho}. \tag{26}$$

Theorem V.1 provides an information-theoretic proof that $H_{\frac{1}{1+\rho}}(X|Y) \leq H_{\frac{1}{1+\rho}}(X)$ for all $\rho > 0$. In fact, equality holds if, and only if, X and Y are independent [2]. We also note that $H_{\frac{1}{1+\rho}}(X|Y)$ is nondecreasing in ρ, that

$$\lim_{\rho \to 0} H_{\frac{1}{1+\rho}}(X|Y) = H(X|Y), \tag{27}$$

and that

$$\lim_{\rho \to \infty} H_{\frac{1}{1+\rho}}(X|Y) = \max_{y \in \mathcal{Y}} \log |\mathrm{supp}(P_{X|Y=y})|. \tag{28}$$

Proof of Theorem V.1: Direct Part. The encoder first describes the conditional type V of $\mathbf{x}$ given $\mathbf{y}$, which requires an asymptotically negligible amount of rate. Since the V-shell of $\mathbf{y}$ is of cardinality at most $e^{nH(V|P_{\mathbf{y}})}$, where $P_{\mathbf{y}}$ denotes the type of $\mathbf{y}$, it may be partitioned into e^{nR} lists of lengths at most

$$\bigl\lceil e^{n(H(V|P_{\mathbf{y}}) - R)} \bigr\rceil. \tag{29}$$

Conditional on $\mathbf{Y} = \mathbf{y}$, the probability that $\mathbf{X}$ is in the V-shell of $\mathbf{y}$ is at most $e^{-nD(V||P_{X|Y}|P_{\mathbf{y}})}$. Consequently, we can use Lemma A.1 to upper-bound the conditional ρ-th moment of the length of the list containing $\mathbf{X}$ by

$$1 + 2^\rho \sum_V e^{-nD(V||P_{X|Y}|P_{\mathbf{y}})} e^{n\rho(H(V|P_{\mathbf{y}}) - R)}, \tag{30}$$

where the sum extends over all V such that the V-shell of $\mathbf{y}$ is nonempty. Since the cardinality of the V-shell of $\mathbf{y}$ depends on $\mathbf{y}$ only via its type, it follows that (30) depends on $\mathbf{y}$ only via $P_{\mathbf{y}}$. Noting that the probability that $\mathbf{Y}$ is of type Q is at most $e^{-nD(Q||P_Y)}$, we can thus upper-bound the unconditional ρ-th moment of the length of the list containing $\mathbf{X}$ by

$$1 + 2^\rho \sum_{Q,V} e^{-nD(Q||P_Y)} e^{-nD(V||P_{X|Y}|Q)} e^{n\rho(H(V|Q) - R)}, \tag{31}$$

where the sum extends over all types $Q \in \mathcal{P}_n(\mathcal{Y})$ and all V such that the V-shell of a sequence of type Q is nonempty. In the Appendix we prove the identity

$$\begin{aligned} &H_{\frac{1}{1+\rho}}(X|Y) \\ &\quad = \max_{Q_{X,Y}} \bigl\{ H(Q_{X|Y}|Q_Y) - \rho^{-1} D(Q_{X,Y}||P_{X,Y}) \bigr\}, \end{aligned} \tag{32}$$

where $H(Q_{X|Y}|Q_Y)$ denotes the conditional entropy of X given Y when the pair (X, Y) has distribution $Q_{X,Y}$. Using (32), the identity

$$D(Q \circ V||P_{X,Y}) = D(Q||P_Y) + D(V||P_{X|Y}|Q), \tag{33}$$

and the fact that the number of types and conditional types is at most a polynomial in n, we can upper-bound (31) by

$$1 + \exp\Bigl(-n\rho\bigl(R - H_{\frac{1}{1+\rho}}(X|Y) - \delta_n\bigr)\Bigr), \tag{34}$$

where $\delta_n \to 0$ as $n \to \infty$. This completes the direct part because (34) tends to one if $R > H_{\frac{1}{1+\rho}}(X|Y)$.

Converse. Fix a sequence $(f_n, \varphi_n)_{n \geq 1}$ of rate-R blocklength-n lossless list codes. As in the proof of Theorem IV.1, we may assume that for every m and $\mathbf{y}$ the list $\varphi_n(m, \mathbf{y})$ is the inverse image of $\{m\}$ under $\mathbf{x} \mapsto f_n(\mathbf{x}, \mathbf{y})$. For every $\mathbf{y}$ these inverse images partition $\mathcal{X}^n$ so Lemma III.1 implies that

$$\begin{aligned} &\sum_{\mathbf{x} \in \mathcal{X}^n} P^n_{X|Y}(\mathbf{x}|\mathbf{y}) \bigl|\varphi_n(f_n(\mathbf{x}, \mathbf{y}), \mathbf{y})\bigr|^\rho \\ &\qquad \geq e^{-n\rho R} \biggl[\sum_{\mathbf{x} \in \mathcal{X}^n} P^n_{X|Y}(\mathbf{x}|\mathbf{y})^{\frac{1}{1+\rho}} \biggr]^{1+\rho}, \quad \mathbf{y} \in \mathcal{Y}^n. \end{aligned} \tag{35}$$

Multiplying both sides of (35) by $P^n_Y(\mathbf{y})$ and summing over all $\mathbf{y} \in \mathcal{Y}^n$ yields

$$\begin{aligned} &\sum_{\mathbf{x} \in \mathcal{X}^n} \sum_{\mathbf{y} \in \mathcal{Y}^n} P^n_{X,Y}(\mathbf{x}, \mathbf{y}) \bigl|\varphi_n(f_n(\mathbf{x}, \mathbf{y}), \mathbf{y})\bigr|^\rho \\ &\qquad \geq e^{-n\rho R} \sum_{\mathbf{y} \in \mathcal{Y}^n} \biggl[\sum_{\mathbf{x} \in \mathcal{X}^n} P^n_{X,Y}(\mathbf{x}, \mathbf{y})^{\frac{1}{1+\rho}} \biggr]^{1+\rho} \\ &\qquad = \exp\Bigl(-n\rho\bigl(R - H_{\frac{1}{1+\rho}}(X|Y)\bigr)\Bigr). \end{aligned} \tag{36}$$

This completes the converse because the RHS of (36) tends to infinity as $n \to \infty$ unless $R \geq H_{\frac{1}{1+\rho}}(X|Y)$. ■

VI. Source Coding with Lists Under a Fidelity Criterion

For any positive ρ, we say that a rate R is ρ-achievable with list source coding and maximal distortion D if there exists a sequence $(f_n, \varphi_n)_{n \geq 1}$ of rate-R blocklength-n distortion-D list codes (see Section I) such that

$$\lim_{n \to \infty} \sum_{\mathbf{x} \in \mathcal{X}^n} P^n(\mathbf{x}) \bigl|\varphi_n(f_n(\mathbf{x}))\bigr|^\rho = 1. \tag{37}$$

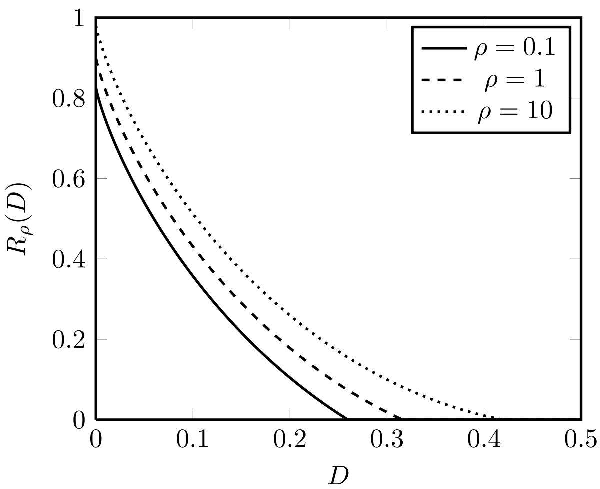

Fig. 1. $R_\rho(D)$ in bits for Bernoulli-(1/4) source and Hamming distortion

Theorem VI.1. *The infimum of all rates that are ρ-achievable with list source coding and maximal distortion D is $R_\rho(D)$ as defined in* (6).

Listed below are some properties of $R_\rho(D)$. See [4] for a more comprehensive list and proofs.

i) $R_\rho(D)$ is nonnegative.
ii) $R_\rho(D)$ is nondecreasing in $\rho > 0$.
iii) $R_\rho(D)$ is nonincreasing, continuous and convex in $D \geq 0$.
iv) $R_\rho(0) = H_{\frac{1}{1+\rho}}(P)$.
v) $\lim_{\rho \to 0} R_\rho(D) = R(P, D)$.
vi) $\lim_{\rho \to \infty} R_\rho(D) = \max_Q R(Q, D)$.

In view of iv, Theorem IV.1 can be recovered from Theorem VI.1 by considering the case where $\hat{\mathcal{X}} = \mathcal{X}$ and d is the Hamming distortion (which is zero if $\hat{x} = x$ and is one otherwise).

It was noted in [4] that $R_\rho(D)$ can be expressed in closed form for binary sources and Hamming distortion:

Proposition VI.2. *If $\mathcal{X} = \hat{\mathcal{X}} = \{0, 1\}$, d is the Hamming distortion function, and $P(0) = p$, then*

$$R_\rho(D) = \begin{cases} H_{\frac{1}{1+\rho}}(p) - h(D) & \text{if } 0 \leq D < h^{-1}(H_{\frac{1}{1+\rho}}(p)), \\ 0 & \text{if } D \geq h^{-1}(H_{\frac{1}{1+\rho}}(p)), \end{cases}$$

where $h^{-1}(\cdot)$ denotes the inverse of the binary entropy function $h(\cdot)$ on the interval $[0, 1/2]$.

For a proof of Proposition VI.2 see [4, Thereom 3] and subsequent remarks. A plot of $R_\rho(D)$ for $p = 1/4$ and different values of ρ is shown in Figure 1.

Proof of Theorem VI.1: Direct Part. The encoder first describes the type Q of the source sequence $\mathbf{x}$. This requires an asymptotically negligible amount of rate. Fix some $\delta > 0$. According to the Type Covering Lemma [3, Lemma 9.1], if $n \geq n_0(\delta)$, then for every type $Q \in \mathcal{P}_n(\mathcal{X})$ we can find a set $B_Q \subset \hat{\mathcal{X}}^n$ of cardinality at most $e^{n(R(Q,D)+\delta)}$ that covers all source sequences of type Q in the sense that for every $\mathbf{x}$ of type Q there is at least one $\hat{\mathbf{x}} \in B_Q$ for which $d(\mathbf{x}, \hat{\mathbf{x}}) \leq D$. Assume therefore that $n \geq n_0(\delta)$. We partition B_Q into e^{nR} lists of lengths at most

$$\left\lceil e^{n(R(Q,D)+\delta-R)} \right\rceil. \tag{38}$$

Using Lemma A.1 and the fact that the probability of the source emitting a sequence of type Q is at most $e^{-nD(Q||P)}$, we can upper-bound the ρ-th moment of the list by

$$1 + 2^\rho \sum_{Q \in \mathcal{P}_n(\mathcal{X})} e^{-nD(Q||P)} e^{n\rho(R(Q,D)+\delta-R)}. \tag{39}$$

Using the definition of $R_\rho(D)$ (6) and the fact that the number of different types is a polynomial in n, we can upper-bound (39) by

$$1 + e^{-n\rho(R - R_\rho(D) - \delta - \delta_n)}, \tag{40}$$

where $\delta_n \to 0$ as $n \to \infty$. The direct part is completed by noting that if $R > R_\rho(D)$, then we can choose δ small enough so that (40) will tend to one as $n \to \infty$.

Converse. Fix a sequence $(f_n, \varphi_n)_{n \geq 1}$ of rate-R blocklength-n distortion-D list codes. We may assume that

$$\varphi_n(m) \cap \varphi_n(m') = \emptyset \quad \text{whenever } m \neq m'. \tag{41}$$

Indeed, if $m \neq m'$ and $\hat{\mathbf{x}} \in \varphi_n(m) \cap \varphi_n(m')$, then we can delete $\hat{\mathbf{x}}$ from the larger of the two lists, say $\varphi_n(m)$, and map to m' all the source sequences $\mathbf{x}$ that where mapped to m by f_n and that satisfy $d(\mathbf{x}, \hat{\mathbf{x}}) \leq D$. This could only reduce the ρ-th moment of the list while preserving the distortion-D property of the code.

Define the set

$$\mathcal{Z}_n = \bigcup_{1 \leq m \leq e^{nR}} \varphi_n(m). \tag{42}$$

Assumption (41) implies that the union on the RHS of (42) is disjoint. Consequently, we may define $m(\hat{\mathbf{x}})$ for every $\hat{\mathbf{x}} \in \mathcal{Z}_n$ as the unique element of $\{1, \ldots, e^{nR}\}$ for which $\hat{\mathbf{x}} \in \varphi_n(m(\hat{\mathbf{x}}))$. Moreover, the distortion-D property guarantees the existence of a mapping $g_n \colon \mathcal{X}^n \to \mathcal{Z}_n$ (not necessarily unique) such that, for all $\mathbf{x} \in \mathcal{X}^n$,

$$g_n(\mathbf{x}) \in \varphi_n(f_n(\mathbf{x})) \quad \text{and} \quad d(\mathbf{x}, g_n(\mathbf{x})) \leq D. \tag{43}$$

With these definition of $m(\hat{\mathbf{x}})$ and g_n, we have

$$\begin{aligned} \sum_{\mathbf{x}} P^n(\mathbf{x}) \left|\varphi_n(f_n(\mathbf{x}))\right|^\rho &= \sum_{\hat{\mathbf{x}} \in \mathcal{Z}_n} P^n\big(g_n^{-1}(\{\hat{\mathbf{x}}\})\big) \left|\varphi_n(m(\hat{\mathbf{x}}))\right|^\rho \\ &= \sum_{\hat{\mathbf{x}} \in \mathcal{Z}_n} \tilde{P}_n(\hat{\mathbf{x}}) \left|\varphi_n(m(\hat{\mathbf{x}}))\right|^\rho, \end{aligned} \tag{44}$$

where we defined the PMF

$$\tilde{P}_n(\hat{\mathbf{x}}) = P^n\big(g_n^{-1}(\{\hat{\mathbf{x}}\})\big), \quad \hat{\mathbf{x}} \in \mathcal{Z}_n. \tag{45}$$

Since the lists $\{\varphi_n(m)\}_{1 \leq m \leq e^{nR}}$ partition the set $\mathcal{Z}_n$, we may apply Lemma III.1 to the RHS of (44) to obtain

$$\begin{aligned} \sum_{\mathbf{x}} P^n(\mathbf{x}) \left|\varphi_n(f_n(\mathbf{x}))\right|^\rho &\geq e^{-n\rho R} \left[\sum_{\hat{\mathbf{x}} \in \mathcal{Z}_n} \tilde{P}_n(\hat{\mathbf{x}})^{\frac{1}{1+\rho}} \right]^{1+\rho} \\ &= \exp\Big(\rho\big(H_{\frac{1}{1+\rho}}(\tilde{P}_n) - nR\big)\Big). \end{aligned} \tag{46}$$

The converse will follow from (46) once we show that

$$H_{\frac{1}{1+\rho}}(\tilde{P}_n) \geq nR_\rho(D). \tag{47}$$

It follows from (18) that, for every PMF Q on $\mathcal{Z}_n$,

$$H_{\frac{1}{1+\rho}}(\tilde{P}_n) \geq H(Q) - \rho^{-1} D(Q||\tilde{P}_n). \tag{48}$$

The PMF $\tilde{P}_n$ can be written as

$$\tilde{P}_n = P^n \widetilde{W}_n, \tag{49}$$

where $\widetilde{W}_n$ is the deterministic channel from $\mathcal{X}^n$ to $\hat{\mathcal{X}}^n$ induced by g_n:

$$\widetilde{W}_n(\hat{\mathbf{x}}|\mathbf{x}) = \begin{cases} 1 & \text{if } \hat{\mathbf{x}} = g_n(\mathbf{x}), \\ 0 & \text{otherwise.} \end{cases} \tag{50}$$

Let $Q_\star$ be a PMF on $\mathcal{X}$ that achieves the maximum in the definition of $R_\rho(D)$, i.e.,

$$R_\rho(D) = R(Q_\star, D) - \rho^{-1} D(Q_\star||P). \tag{51}$$

Substituting $Q_\star^n \widetilde{W}_n$ for Q in (48) and using (49),

$$\begin{aligned} H_{\frac{1}{1+\rho}}(\tilde{P}_n) &\geq H(Q_\star^n \widetilde{W}_n) - \rho^{-1} D(Q_\star^n \widetilde{W}_n || P^n \widetilde{W}_n) \\ &\geq H(Q_\star^n \widetilde{W}_n) - \rho^{-1} D(Q_\star^n || P^n) \\ &= H(Q_\star^n \widetilde{W}_n) - n\rho^{-1} D(Q_\star || P), \end{aligned} \tag{52}$$

where the inequality in the second line follows from the Data Processing Inequality [3, Lemma 3.11]. Let $\tilde{\mathbf{X}} = (\tilde{X}_1, \ldots, \tilde{X}_n)$ be produced by the DMS $Q_\star$, and set $\hat{\mathbf{X}} = g_n(\tilde{\mathbf{X}})$. Then

$$\begin{aligned} H(Q_\star^n \widetilde{W}_n) &= H(\hat{\mathbf{X}}) \\ &= I(\tilde{\mathbf{X}} \wedge \hat{\mathbf{X}}). \end{aligned} \tag{53}$$

In view of (43), we have

$$\mathrm{E}[d(\tilde{\mathbf{X}}, \hat{\mathbf{X}})] \leq D, \tag{54}$$

so applying [6, Theorem 9.2.1] (which is the main ingredient in the classical Rate-Distortion converse) to the pair $(\tilde{\mathbf{X}}, \hat{\mathbf{X}})$ yields

$$\begin{aligned} I(\tilde{\mathbf{X}} \wedge \hat{\mathbf{X}}) &\geq nR\big(Q_\star, \mathrm{E}[d(\tilde{\mathbf{X}}, \hat{\mathbf{X}})]\big) \\ &\geq nR(Q_\star, D), \end{aligned} \tag{55}$$

where we used the monotonicity of the rate-distortion function in the second line. Combining (55), (53), (52), and (51) establishes (47). ■

VII. Final Remarks

1) The converse proofs in this paper show that for all rates strictly below $H_{\frac{1}{1+\rho}}(P)$, $H_{\frac{1}{1+\rho}}(X|Y)$, and $R_\rho(D)$, respectively, the ρ-th moment of the list must grow to infinity exponentially fast in the blocklength. In fact, since the exponents in the achievability and converse parts of the proofs match, we can characterize the best exponents possible: they are given by $\rho(H_{\frac{1}{1+\rho}}(P) - R)$, $\rho(H_{\frac{1}{1+\rho}}(X|Y) - R)$, and $\rho(R_\rho(D) - R)$, respectively.
2) The fact that $R_\rho(D)$ is a continuous function of D [4] allows us to strengthen the converse statement in Theorem VI.1 as follows. If $(f_n, \varphi_n)_{n\geq 1}$ is a sequence of rate-R blocklength-n distortion-D_n codes, where $\overline{\lim}_{n\to\infty} D_n \leq D$ and $R < R_\rho(D)$, then the ρ-th moment of the list grows to infinity exponentially fast in n. Indeed, continuity implies that $R < R_\rho(D+\epsilon)$ for some $\epsilon > 0$, and $\overline{\lim}_{n\to\infty} D_n \leq D$ implies that $D_n \leq D + \epsilon$ for all sufficiently large n. The claim thus follows from the converse part of the proof of Theorem VI.1.

Appendix

Lemma A.1. *For every $\alpha \in \mathbb{R}$ and $\rho > 0$,*

$$\lceil e^{n\alpha} \rceil^\rho < 1 + 2^\rho e^{n\rho\alpha}, \quad n = 1, 2, \ldots \tag{56}$$

Proof: If $\alpha > 0$, then $\lceil e^{n\alpha} \rceil < 1 + e^{n\alpha} < 2e^{n\alpha}$. If $\alpha \leq 0$, then $\lceil e^{n\alpha} \rceil = 1$. Thus, $\lceil e^{n\alpha} \rceil^\rho \leq \max\{1, 2^\rho e^{n\rho\alpha}\} < 1 + 2^\rho e^{n\rho\alpha}$. ■

Proof of (32)*:* Observe that for any PMF $Q_{X,Y}$ on $\mathcal{X} \times \mathcal{Y}$,

$$\begin{aligned} &H(Q_{X|Y}|Q_Y) - \rho^{-1} D(Q_{X,Y}||P_{X,Y}) \\ &= \frac{1+\rho}{\rho} \sum_y Q_Y(y) \sum_x Q_{X|Y}(x|y) \log \frac{P_{X|Y}(x|y)^{\frac{1}{1+\rho}}}{Q_{X|Y}(x|y)} \\ &\qquad - \frac{1}{\rho} \sum_y Q_Y(y) \log \frac{Q_Y(y)}{P_Y(y)} \\ &\leq \frac{1+\rho}{\rho} \sum_y Q_Y(y) \log \sum_x P_{X|Y}(x|y)^{\frac{1}{1+\rho}} \\ &\qquad - \frac{1}{\rho} \sum_y Q_Y(y) \log \frac{Q_Y(y)}{P_Y(y)} \\ &= \frac{1}{\rho} \sum_y Q_Y(y) \log \frac{P_Y(y) \big[\sum_x P_{X|Y}(x|y)^{\frac{1}{1+\rho}}\big]^{1+\rho}}{Q_Y(y)} \\ &\leq \frac{1}{\rho} \log \sum_y P_Y(y) \Bigg[\sum_x P_{X|Y}(x|y)^{\frac{1}{1+\rho}}\Bigg]^{1+\rho} \\ &= H_{\frac{1}{1+\rho}}(X|Y), \end{aligned}$$

where the inequalities follow from Jensen's Inequality. Equality is attained in both inequalities by the choice

$$Q_Y(y) = \frac{P_Y(y) \big[\sum_x P_{X|Y}(x|y)^{\frac{1}{1+\rho}}\big]^{1+\rho}}{\sum_{y'} P_Y(y') \big[\sum_x P_{X|Y}(x|y')^{\frac{1}{1+\rho}}\big]^{1+\rho}},$$

and $Q_{X|Y}(x|y) = P_{X|Y}(x|y)^{\frac{1}{1+\rho}} / \sum_{x'} P_{X|Y}(x'|y)^{\frac{1}{1+\rho}}$. ■

References

[1] O. Shayevitz, "A note on a characterization of Rényi measures and its relation to composite hypothesis testing," *arXiv preprint arXiv:1012.4401*, 2010.

[2] S. Arimoto, "Information measures and capacity of order α for discrete memoryless channels," in *Topics in Information Theory*, I. Csiszár and P. Elias, Eds. Amsterdam, The Netherlands: North Holland, 1977, vol. 17, pp. 41–52.

[3] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, 2nd ed. New York: Cambridge University Press, 2011.

[4] E. Arikan and N. Merhav, "Guessing subject to distortion," *IEEE Trans. Inf. Theory*, vol. 44, no. 3, pp. 1041–1056, 1998.

[5] E. Arikan, "An inequality on guessing and its application to sequential decoding," *IEEE Trans. Inf. Theory*, vol. 42, no. 1, pp. 99–105, 1996.

[6] R. Gallager, *Information Theory and Reliable Communication*. New York: John Wiley & Sons, 1968.