Variable Length Compression of Codeword Indices for Lossy Compression*

Dror Baron and Theju Jacob

October 3, 2012

Abstract

Many problems in information theory feature an index into a random codebook being encoded with a fixed length scheme. We propose to purposefully select the index in a manner that skews its distribution, thus making variable length entropy coding of the index more attractive. In an application to lossy compression of a Bernoulli source, we illustrate that variable length coding yields a reduction in the rate over fixed length coding, and allows to reach a requisite rate distortion performance level using a smaller codebook.

Keywords: entropy coding, lossy compression, rate distortion, variable length coding.

1 Introduction

A ubiquitous feature of information theory is the encoding of an index. In lossy compression, an index into a random codebook selects the codeword with minimal distortion with respect to (w.r.t.) an input block [1]; in the Slepian-Wolf problem [2], an index into a coset allows side information to reconstruct an input block; and in Ziv-Lempel compression [3], an index into past data is used to encode a block of future data. To the best of our knowledge, most coding methods in the prior art have encoded the index using a fixed length scheme. That is, if there were M possible indices, then the index was encoded using $\log_2(M)$ bits.

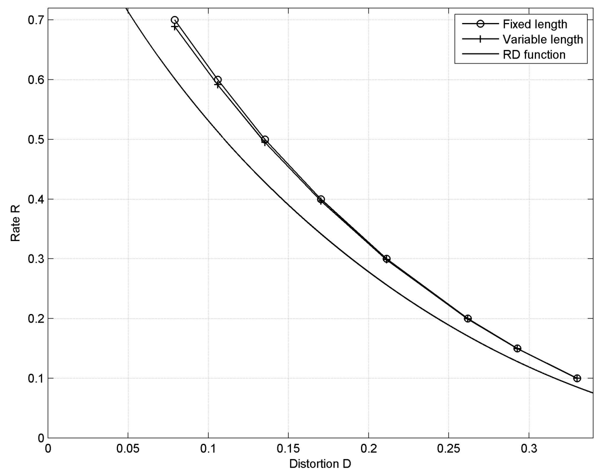

Figure 1: *Variable length coding of the codeword index offers a reduction in the rate over fixed length coding in lossy compression. (Random codebook of size $M = 2^{19}$; the Bernoulli input parameter was $p = 0.5$.)*

In this letter, we propose to select the index in a manner that skews its distribution, thus making variable length coding more attractive. Similar to the approach by Kontoyiannis and Zamir [4], we then perform entropy coding of the index.

We evaluate our idea by focusing on an application to lossy compression of a Bernoulli source, where the best match to an input block of length n is found in a random codebook of size $M = 2^{nR}$, where R is the rate. When the input alphabet is finite, there are often multiple codewords that have the same minimal distortion w.r.t. an input block. If this is the case, then any of these indices can be used with no impact on the distortion. We select the minimum such index; its distribution is skewed toward lower indices, and variable length entropy coding of the index via arithmetic encoding [5] reduces the overall transmission cost. Figure 1 illustrates a reduction in the rate over fixed length coding. The rate distortion (RD) performance using variable length coding is closer to the RD function than the performance using fixed lengths.

Variable length entropy coding of the index also allows to reach a requisite RD performance level using a smaller codebook (Section 4). Therefore, our contribution can also be seen along the lines of Gioran and Kontoyiannis [6], who show that the RD performance of a large random codebook can be mimicked with a smaller database. On the other hand, in order to eek the savings out of an entropy code, we would need to concatenate a large number of input blocks.

The sequel of the letter is organized as follows. Section 2 describes the setting of a lossy compressor based on a random codebook, and the rate distortion performance of this compressor is analyzed in Section 3. Section 4 describes numerical results, and we conclude in Section 5.

2 Background

Problem setting: Consider lossy compression over a binary alphabet, where the input is comprised of binary symbols, each input symbol is Bernoulli independent and identically distributed (iid), and is equal to one with probability p. The goal of the encoder is to communicate the input to the decoder with modest distortion using a low coding rate.

To do so, the input is partitioned into blocks of length n, and we maintain a codebook that consists of $M = 2^{nR}$ codewords, each of length n. Both the encoder and decoder know the codebook. For each input block, the encoder computes the Hamming distortion of the block with respect to (w.r.t.) each codeword. The encoder finds d_{min}, the minimal distortion achieved among all codewords. Seeing that there could be multiple codewords that achieve d_{min}, the encoder selects as the representative codeword that with minimal index. The encoder transmits this minimal index to the decoder. The distribution of the minimal index is skewed toward lower indices, and variable length entropy coding is used to reduce the coding length. The actual transmission can be performed using arithmetic coding [5].

The decoder is straightforward, it reconstructs each input block by the codeword whose index was transmitted by the encoder.

RD performance: It is well known that in order to optimize the rate distortion (RD) performance, the optimal codebook is comprised of codewords that are generated iid with probability

$$q = \frac{p - D}{1 - 2D},$$

resulting in the following RD function,

$$R(D) = H_2(p) - H_2(D), \tag{1}$$

where $H_2(p) = -p\log_2(p) - (1-p)\log_2(1-p)$ is the binary entropy.

Related work: Similar lossy compression schemes were proposed by Gupta et al. [7] and Gioran and Kontoyiannis [6]. Our contribution is to select the minimal index in order to skew its distribution and thus make variable length coding more attractive.

Entropy coding of the minimal index was proposed by Kontoyiannis and Zamir [4]. Our approach differs as follows, (*i*) in our work the input distribution is known, whereas they emphasize a setting with mismatch between the codebook statistics and true input statistics; and (*ii*) we choose the minimal index among those that offer smallest distortion, whereas they choose among indices that offer sufficiently small distortion.

3 Analysis

To analyze the RD performance of our approach, we compute the distribution of distortions achieved by our scheme and the distribution of indices. Our computation is taken with respect to an ensemble of random codebooks. RD performance for individual codebooks will vary, and the results here prove the existence of codebooks with the requisite RD performance.

Let $\mathcal{C}$ be a random codebook, where the codewords are generated iid according to $\hat{X}^n(j) \sim \prod_{i=1}^n p(\hat{x}_{ij})$ for $j = 1, \ldots, M$, where $\hat{X}^n(j)$ is codeword number j, $\hat{x}_{ij}$ is symbol i in codeword j, and $\hat{X}_{ij} \sim \text{Bernoulli}(q)$.

For a given source sequence $x^n \in \{0,1\}^n$, define the functions

$$D_{\text{min}}(x^n) = \min_{1\le j\le M} d_H(x^n, \hat{X}^n(j)),$$
$$N(x^n, d) = \left|\left\{ j : d_H(x^n, \hat{X}^n(j)) = d, 1 \le j \le M \right\}\right|,$$

where $d_H(\cdot,\cdot)$ denotes Hamming distance. Observe that $D_{\text{min}}(x^n)$ and $N(x^n,d)$ are both random variables since they are functions of $\mathcal{C}$.

Let x^n be any sequence in $\{0,1\}^n$, and let $w(x^n) \triangleq d_H(x^n, 0)$ be the Hamming weight of the binary sequence x^n. For all $j = 1, \dots, M$:

$$\begin{aligned}&\Pr\left[d_H(x^n, \hat{X}^n(j)) = d\right] \\ &= \sum_{\substack{0 \le d_1 \le w(x^n) \\ 0 \le d_0 \le n - w(x^n) \\ d = w(x^n) - d_1 + d_0}} \left[\binom{w(x^n)}{d_1}\binom{n - w(x^n)}{d_0} q^{d_0 + d_1}(1-q)^{n - d_0 - d_1}\right].\end{aligned}$$

The codewords in $\mathcal{C}$ are generated independently from one another, hence:

$$\begin{aligned}&\Pr\left[D_{\min}(x^n) = d, N(x^n, d) = m\right] \\ &= \binom{M}{m}\Pr\left[d_H(x^n, \hat{X}^n(j)) = d\right]^m \Pr\left[d_H(x^n, \hat{X}^n(j)) > d\right]^{M-m} \\ &= \binom{M}{m}\Pr\left[d_H(x^n, \hat{X}^n(j)) = d\right]^m \left(\sum_{i=d+1}^{n} \Pr\left[d_H(x^n, \hat{X}^n(j)) = i\right]\right)^{M-m}.\end{aligned}$$

Since the codebook $\mathcal{C}$ is chosen independently of the realized source sequence, and $\Pr[D_{\min}(x^n) = d, N(x^n, d) = m]$ is equal for all x^n sequences with the same Hamming weight, it follows that

$$\begin{aligned}&\Pr\left[D_{\min}(X^n) = d, N(X^n, d) = m | w(X^n) = w(x^n)\right] \\ &= \Pr\left[D_{\min}(X^n) = d, N(X^n, d) = m | X^n = x^n\right] \\ &= \Pr\left[D_{\min}(x^n) = d, N(x^n, d) = m\right].\end{aligned}$$

Therefore, the joint distribution of $D_{\min}(X^n)$ and $N(X^n, D_{\min}(X^n))$ is given by

$$\begin{aligned}&\Pr\left[D_{\min}(X^n) = d, N(X^n, d) = m\right] \\ &= \sum_{x^n \in \{0,1\}^n} \Pr\left[X^n = x^n\right] \Pr\left[D_{\min}(X^n) = d, N(X^n, d) = m | X^n = x^n\right] \\ &= \sum_{t=0}^{n} \binom{n}{t} p^t (1-p)^{n-t} \Pr\left[D_{\min}(X^n) = d, N(X^n, d) = m | w(X^n) = t\right].\end{aligned}$$

From this, we can immediately compute the marginal distributions $\Pr\left[D_{\min}(X^n) = d\right]$ and $\Pr[N(X^n,$

$D_{\min}(X^n)) = m]$. This allows us to compute the expected distortion:

$$\mathbb{E}\left[D_{\min}(X^n)\right] = \sum_{d=0}^{n} d \cdot \Pr\left[D_{\min}(X^n) = d\right]. \tag{2}$$

We can now compute the probability that index i is the smallest index for which $d_H(X^n, \hat{X}^n(i)) = D_{\min}(X^n)$ over the ensemble of random codebooks:

$$\Pr\left[I_{\min} = i\right] = \sum_{m=1}^{M-i+1} \Pr\left[N(X^n, D_{\min}(X^n)) = m\right] \frac{\binom{M-i}{m-1}}{\binom{M}{m}}.$$

It might seem that we can now compute the rate as the entropy of the index distribution $\Pr\left[I_{\min} = i\right]$. However, such a calculation would be imprecise, because entropy is the expected value of a function that refers explicitly to the index distribution for a specific codebook. However, because the entropy is concave, we can upper bound the expected rate over the ensemble of random codebooks as follows,

$$\begin{aligned} \mathbb{E}[R] &\leq \frac{1}{n} H(\Pr\left[I_{\min=i}\right]) \\ &= -\frac{1}{n} \sum_{i=1}^{M} \Pr\left[I_{\min} = i\right] \log_2\left(\Pr\left[I_{\min} = i\right]\right). \end{aligned} \tag{3}$$

Recall that RD performance for individual codebooks will vary, and so we have proved the existence of codebooks whose expected distortion D satisfies (2) and whose rate R is upper bounded by (3).

4 Numerical examples

Using the expressions in Section 3, we computed the bound on the expected RD performance over the ensemble of random codebooks for several sets of parameters: (*i*) the Bernoulli input parameter was $p \in \{0.2, 0.4, 0.5\}$; (*ii*) the codebook size was $M = 2^j$, where $j \in \{10, 11, \ldots, 19\}$; and (*iii*) the target rate was $R_{target} \in \{0.1, 0.15, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7\}$. To make the meaning of target rate precise, the input blocks were set to have length $n = \lceil j/R_{target} \rceil$, where $\lceil \cdot \rceil$ denotes rounding up, resulting in a rate $R = j/n \approx R_{target}$ for fixed length coding. The expected distortion and rate were then computed using the results of Section 3, yielding a rate $R < R_{target}$ for the variable length regime.

Two sets of results are included for illustration. In the first set, the Bernoulli input parameter was $p = 0.5$, and codebooks of size $M = 2^{19}$ were used. We see in Figure 1 that variable length coding offers a reduction in rate over fixed length coding. It can be computed from this figure that the proposed technique would need approximately 18.5 bits on average to encode an input block, instead of the 19 bits required by fixed length coding.

In the second set, the Bernoulli input parameter was $p = 0.2$, and the target rate was $R_{target} = 0.5$. We see in Figure 2 that as M is increased, both variable and fixed length coding improve. In particular, the ratio between their rates and the theoretically optimal $R(D)$ (1) declines as $j = \log_2(M)$ is increased. More importantly, the variable length approach achieves the same RD performance using a codebook that is smaller; seeing that variable length coding with a codebook of size $M = 2^{14}$ has better RD performance than fixed length coding with $M = 2^{16}$, a reduction by a factor of 5 in the size of the codebook seems attainable. Therefore, variable length entropy coding enables to achieve a desired RD performance level using less storage and computational resources.

5 Conclusions

In this letter, we proposed to select the index into a random codebook in a manner that skews its distribution, thus making variable length entropy coding of the index more attractive. We illustrated for the application of lossy compression of iid Bernoulli binary input blocks that our scheme yields a reduction in the rate, and achieves the same RD performance level using a smaller codebook.

The concept of encoding the index into a random codebook using variable length coding can be used in other information theory problems that involve transmitting indices into codebooks. For example, in Ziv-Lempel coding [3] the index into past data used to encode a block of future data can be encoded with variable length, thus improving further the overall compression.

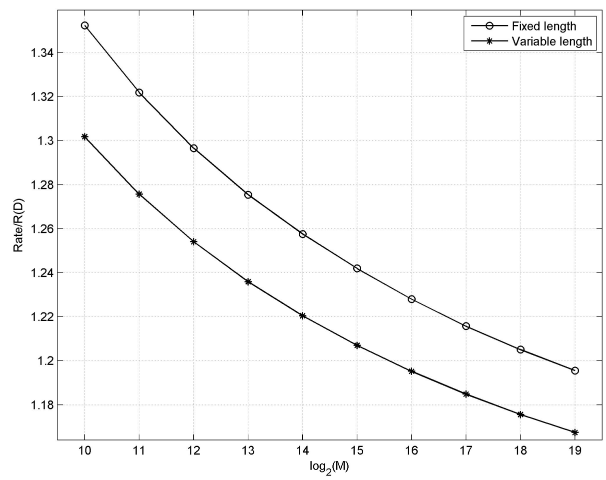

Figure 2: *The best match to a binary Bernoulli input block of length n is found in a random codebook of size $M = 2^{Rn}$, where the fixed length target rate is $R_{target} = 0.5$. With variable length coding, the rate approaches the RD function $R(D)$ more quickly as n is increased. (The Bernoulli input parameter was $p = 0.2$)*

Acknowledgments

We acknowledge preliminary conversations with Tsachy Weissman, which framed our thinking about the problems discussed here, and thank Ioannis Kontoyiannis for helping to tie our work into the prior art. We also thank the two anonymous reviewers whose constructive comments helped to greatly improve the paper.

References

[1] T. Berger, *Rate distortion theory; a mathematical basis for data compression*, Prentice-Hall Englewood Cliffs, NJ, 1971.

[2] D. Slepian and J. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inf. Theory*, vol. 19, no. 4, pp. 471–480, Jul. 1973.

[3] J. Ziv and A. Lempel, "A universal algorithm for sequential data compression," *IEEE Trans. Inf. Theory*, vol. 23, no. 3, pp. 337–343, 1977.

[4] I. Kontoyiannis and R. Zamir, "Mismatched codebooks and the role of entropy coding in lossy data compression," *IEEE Trans. Inf. Theory*, vol. 52, no. 5, pp. 1922–1938, May 2006.

[5] J. Rissanen and G. G. Langdon, "Arithmetic coding," *IBM J. Res. Dev.*, vol. 23, no. 2, pp. 149–162, 1979.

[6] C. Gioran and I. Kontoyiannis, "Lossy compression in near-linear time via efficient random codebooks and databases," *CoRR*, vol. abs/0904.3340, 2009.

[7] A. Gupta, S. Verdú, and T. Weissman, "Rate-distortion in near-linear time," in *Proc. Int. Symp. Inf. Theory (ISIT2008)*, Jul. 2008.