ON SPLIT QUANTIZATION OF LSF PARAMETERS

Fredrik Nordén

Department of Communication Technology, Aalborg University, DK-9220 Aalborg Ø, Denmark

ABSTRACT

In this paper we investigate split quantization of the LSF source from an information theoretic perspective. It is a well known fact that split quantization is inferior to unconstrained quantization, due to the independent treatment of sub vectors. Here, we quantify lower bounds for the split loss, and suggest conditional quantization of splits as a method to reduce the losses. The investigation is based on information theoretic estimates of performance, rather than on the performance of actual quantization schemes.

Simulations for a 10-dimensional LSF source point at a loss of around 8 bits if scalar quantization is performed, and a loss of approximately 2 bits for a 2-split quantization, both compared to 10 dimensional vector quantization. Furthermore, the results suggest that these losses can be considerably reduced using conditional quantization, down to as low as 0 to 2 bits.

1. INTRODUCTION

In current low to medium rate speech coders the speech signal is often separated into a spectral envelope and a residual signal before coding. The spectral envelope is typically represented with linear prediction coefficients (LPC) [1]. Commonly, these coefficients are separately quantized using the line spectral frequency (LSF) representation [2]. Quantization of LSF vectors typically requires a substantial part of the total bit rate. In order to reduce the required rate, vector quantization (VQ) of LSF vectors is often employed. The potential of LSF VQ has been much studied, and a substantial bit rate reduction, in the order of 10 bits, compared to scalar quantization can be achieved, but complexity is a problem [3, 4]. In order to reduce complexity some constrained VQ structure is often employed [5]. One such structure is split VQ [6].

In split VQ the vector to be coded is split into sub-vectors before quantization. It is a well known fact that split VQ is inferior to unconstrained VQ, due to the independent treatment of sub-vectors. In this paper we use information theoretic measures to quantify performance aspects of split quantization [7]. We quantify the information theoretic split loss, and we study schemes to reduce the losses of split quantization; conditional quantization of the splits.

This paper is organized as follows. In Section 2, we discuss information theoretic measures that constitute the base of this paper. In Section 3, this is followed by a definition of split VQ, and a discussion of how to apply the information theoretic measures to split quantization. In Section 4, we discuss how to interpret the proposed measures for practical split quantization, and in Section 5 we apply the proposed performance analysis to the LSF source. Finally, we give some conclusions in Section 6. Thomas Eriksson

Department of Signals and Systems, Chalmers University of Technology, S-412 96 Göteborg, Sweden

2. INFORMATION MEASURES

Information theory is a valuable tool box for analyzing compression properties of random variables [7]. The information content, or the uncertainty, of a discrete random variable, A, is given by its entropy, H(A). For continuous random variables, there is a corresponding quantity denoted differential entropy.

The differential entropy of a continuous random variable, \mathbf{X} , is defined as

$$h(\mathbf{X}) = -\int_{\Omega_{\mathbf{X}}} f_{\mathbf{X}}(\mathbf{x}) \log_2\left(f_{\mathbf{X}}(\mathbf{x})\right) d\mathbf{x},\tag{1}$$

where Ω_x is the support region of **X**. Just as the entropy for the discrete case, the differential entropy is related to the shortest description length.

Another measure useful for the purposes of this paper is conditional entropy. The differential entropy for the random variable \mathbf{X} given the random variable \mathbf{Y} is referred to as conditional entropy, and defined as

$$h(\mathbf{X}|\mathbf{Y}) = -\iint_{\Omega_{\mathbf{x}\mathbf{y}}} f_{\mathbf{X}\mathbf{Y}}(\mathbf{x}, \mathbf{y}) \log_2\left(f_{\mathbf{X}|\mathbf{Y}}(\mathbf{x}|\mathbf{y})\right) d\mathbf{x} d\mathbf{y}$$
$$= h(\mathbf{X}, \mathbf{Y}) - h(\mathbf{Y}).$$
(2)

When studying pairs of random variables we can express the information that one variable contains about another using the notion of mutual information. The mutual information, $I(\mathbf{X}; \mathbf{Y})$, between two random variables, \mathbf{X} and \mathbf{Y} , is defined as

$$I(\mathbf{X}; \mathbf{Y}) = \iint_{\Omega_{\mathbf{X}\mathbf{y}}} f_{\mathbf{X}\mathbf{Y}}(\mathbf{x}, \mathbf{y}) \log_2\left(\frac{f_{\mathbf{X}\mathbf{Y}}(\mathbf{x}, \mathbf{y})}{f_{\mathbf{X}}(\mathbf{x}) f_{\mathbf{Y}}(\mathbf{y})}\right) d\mathbf{x} d\mathbf{y}$$
$$= h(\mathbf{X}) + h(\mathbf{Y}) - h(\mathbf{X}, \mathbf{Y}). \tag{3}$$

Mutual information possess the property $I(\mathbf{X}; \mathbf{Y}) \ge 0$, with equality if \mathbf{X} and \mathbf{Y} are independent.

A more general concept than mutual information is the Kullback Leibler distance [7], also referred to as the relative entropy. The Kullback Leibler distance, $\mathcal{D}(f_{\mathbf{X}}(\mathbf{x})||g_{\mathbf{X}}(\mathbf{x}))$, is a description of the difference between the pdfs $f_{\mathbf{X}}(\mathbf{x})$ and $g_{\mathbf{X}}(\mathbf{x})$, defined as

$$\mathcal{D}\left(f_{\mathbf{X}}(\mathbf{x}) \| g_{\mathbf{X}}(\mathbf{x})\right) = \int_{\Omega_{\mathbf{x}}} f_{\mathbf{X}}(\mathbf{x}) \log_2\left(\frac{f_{\mathbf{X}}(\mathbf{x})}{g_{\mathbf{X}}(\mathbf{x})}\right) d\mathbf{x}.$$
 (4)

3. INFORMATION IN SPLIT QUANTIZATION

Here, focus is on a complexity reduction technique referred to as split VQ, where a high dimensional vector, \mathbf{z} , is split into a number

of lower dimensional vectors before quantization,

$$\mathbf{z} = [\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_N]. \tag{5}$$

The complexity reduction comes at the cost of an increased distortion. A split VQ is suboptimal, due to the independent treatment of sub-vectors.

Below, we discuss methods to evaluate the information theoretics *losses* for split VQ, and the gains of some potential *remedies*, all based on the information theoretic measures discussed in Section 2. In Section 5 we present results from applying the methods discussed below to the LSF source.

3.1. Losses - Independent Treatment

The information theoretic loss for a split representation is due to the independent treatment of sub-vectors. One way to quantify this loss is to evaluate the amount of information that is treated a multiple of times. For a two split representation, $\mathbf{z} = [\mathbf{x}, \mathbf{y}]$, this can be evaluated by the mutual information, $I(\mathbf{X}; \mathbf{Y})$, c.f. Eq. (3). In order to evaluate the cost for any number of splits we need to generalize.

From Eqs. (3) and (4) it is obvious that the mutual information corresponds to the Kullback Leibler distance between the joint pdf, $f_{\mathbf{XY}}(\mathbf{x}, \mathbf{y})$, and the product of the marginals, $f_{\mathbf{X}}(\mathbf{x})f_{\mathbf{Y}}(\mathbf{y})$. Generalizing to more than two random variables, we can define the loss for an N split, $L_N^{(\text{split})}$, as the Kullback Leibler distance between the joint pdf, $f_{\mathbf{Z}}(\mathbf{z})$, and the product of the split marginals, $\prod_{i=1}^N f_{\mathbf{Z}_i}(\mathbf{z}_i)$,

$$L_N^{(\text{split})} = \mathcal{D} \quad f_{\mathbf{Z}}(\mathbf{z}) \| \prod_{i=1}^N f_{\mathbf{Z}_i}(\mathbf{z}_i) \right)$$
$$= \sum_{i=1}^N h(\mathbf{Z}_i) - h(\mathbf{Z}), \tag{6}$$

where $\mathbf{Z} = [\mathbf{Z}_1, \mathbf{Z}_2, \dots, \mathbf{Z}_N]$. Using this measure we can evaluate the split loss for an arbitrary number of splits, N.

As is obvious from Eq. (6), we can interpret $L_N^{\text{(split)}}$ as the difference in differential entropy between the sum of the split sources, \mathbf{Z}_i , and the joint source, \mathbf{Z} .

3.2. Remedies - Conditioning

The split loss, as measured with Kullback Leibler distance, can be recovered using conditional representations. If split *i*, \mathbf{Z}_i , is conditioned on splits $\mathbf{Z}_1, \mathbf{Z}_2, \ldots, \mathbf{Z}_{i-1}$, it can be shown, using the chain rule for conditional entropy [7], that the Kullback Leibler distance between the joint pdf, $f_{\mathbf{Z}}(\mathbf{z})$, and the conditional product pdf,

$$f_{\mathbf{Z}}^{(\text{cond})}(\mathbf{z}) = \prod_{i=1}^{N} f_{\mathbf{Z}_{i}|[\mathbf{Z}_{1},\mathbf{Z}_{2},...,\mathbf{Z}_{i-1}]}(\mathbf{z}_{i}|[\mathbf{z}_{1},\mathbf{z}_{2},...,\mathbf{z}_{i-1}]),$$
(7)

is equal to zero, i.e.

$$L_N^{(\text{cond})} = \mathcal{D}\left(f_{\mathbf{Z}}(\mathbf{z}) \| f_{\mathbf{Z}}^{(\text{cond})}(\mathbf{z})\right)$$
$$= \sum_{i=1}^N h(\mathbf{Z}_i | [\mathbf{Z}_1, \mathbf{Z}_2, \dots, \mathbf{Z}_{i-1}]) - h(\mathbf{Z}) = 0.$$
(8)

A more practical approach is to condition split *i*, \mathbf{Z}_i , on split i - 1, \mathbf{Z}_{i-1} . We can define the loss for an N split 1-step conditioning, $L_N^{(\text{cond}-1)}$, compared to the joint representation, as the Kullback Leibler distance between the joint pdf, $f_{\mathbf{Z}}(\mathbf{z})$, and the 1-step conditional product pdf,

$$f_{\mathbf{Z}}^{(\text{cond}-1)}(\mathbf{z}) = \prod_{i=1}^{N} f_{\mathbf{Z}_{i}|\mathbf{Z}_{i-1}}(\mathbf{z}_{i}|\mathbf{z}_{i-1}), \qquad (9)$$

as

$$L_N^{(\text{cond}-1)} = \mathcal{D}\left(f_{\mathbf{Z}}(\mathbf{z}) \| f_{\mathbf{Z}}^{(\text{cond}-1)}(\mathbf{z})\right)$$
$$= h(\mathbf{Z}_1) + \sum_{i=2}^N h(\mathbf{Z}_i | \mathbf{Z}_{i-1}) - h(\mathbf{Z}).$$
(10)

Another possible approach for conditioning is to condition split *i*, \mathbf{Z}_i , on the right most element of split i - 1, $\mathbf{Z}_{i-1}(M_{i-1})$, where M_{i-1} is the number of elements in split i-1. We can define the loss for an N split 1-step, 1-element conditioning, $L_N^{(\text{cond}-1,1)}$, compared to the joint representation, as the Kullback Leibler distance between the joint pdf, $f_{\mathbf{Z}}(\mathbf{z})$, and the conditional product pdf,

$$f_{\mathbf{Z}}^{(\text{cond}-1,1)}(\mathbf{z}) = \prod_{i=1}^{N} f_{\mathbf{Z}_{i}|\mathbf{Z}_{i-1}(M_{i-1})}(\mathbf{z}_{i}|\mathbf{z}_{i-1}(M_{i-1})), \quad (11)$$

as

$$L_{N}^{(\text{cond}-1,1)} = \mathcal{D}\left(f_{\mathbf{Z}}(\mathbf{z}) \| f_{\mathbf{Z}}^{(\text{cond}-1,1)}(\mathbf{z})\right)$$

= $h(\mathbf{Z}_{1}) + \sum_{i=2}^{N} h(\mathbf{Z}_{i} | \mathbf{Z}_{i-1}(M_{i-1})) - h(\mathbf{Z}).$ (12)

4. PRACTICAL SPLIT QUANTIZATION

The simultaneous treatment of blocks of samples in VQ gives a higher degree of freedom for choosing the reconstruction points compared to split quantization, and thus better performance in terms of incurred distortion. This advantage comes from the ability of exploiting statistical dependencies among samples in the treated vector, and the geometrical fact that operation in a high dimension enables more efficient decision regions.

Above, we discussed information theoretic tools for quantifying this advantage. Interpretation of these measures for practical split quantization should be done with some care. Firstly, the loss measures $(L_N^{\text{(split)}}, L_N^{\text{(cond)}} \text{ etc})$ are based on entropies, and are thus based on the indirect assumption of a simultaneous treatment of an infinite number of vectors. Secondly, the conditioning remedies do not consider the practical need of conditioning on quantized data.

4.1. Finite Dimension

A practical VQ quantizes a single vector at each quantization instance, while the proposed measures indirectly assume the simultaneous treatment of an infinite number of vectors. This fact means that the proposed measures will differ from the loss in practical split quantization. The proposed measures will act as *lower bounds* for the split loss of practical quantization, as argued below.

According to Eq. (6) split quantization of a vector with independent components would not incur any loss at all. This is for two reasons not true in practical split quantization; *local* and *global* sub optimalities. Locally, or *intra split*, the full dimensional VQ^1 enables quantization cell shapes, Voronoi shapes, that more efficiently fills space, compared to the cell shapes of the lower dimensional split VQs. Globally, or *inter split*, the combination of independent quantizers forces a rectangular structure of the layout of the quantization regions. These VQ advantages are often referred to as the space filling advantage and the shape advantage, respectively [8]. Neither of these two advantages come into play for the proposed measures, due to the indirect assumption of a simultaneous treatment of an infinite number of vectors, an infinite delay, for each split.

4.2. Conditioning on Quantized Data

Another aspect which should be considered is that in a real application, the conditional coding need to be performed on quantized data. Quantization reduces the information content, and thus

$$h(Y|X) \ge h(Y|X),\tag{13}$$

where \hat{X} represents a quantized version of X. Further, this means that the losses for practical conditional quantization are larger than suggested by the measures in Eqs. (8), (10) and (12). Thus, these measures should be considered as *lower bounds*.

4.3. Conditional Coding in Practice

In practice, utilization of the conditional information can be performed in a number of ways. Obviously, any method that can be employed to exploit interframe dependencies [9], can also be used to exploit inter split dependencies. The linear dependence among splits can for example be extracted by utilization of vector linear prediction [10]. To achieve a higher performance some non-linear coding approach can be employed, e.g. [11, 12]. Other practical approaches for exploiting memory in between sub-vectors, including delayed-decision approaches, have been proposed in e.g. [13, 14, 15]. We do not study practical schemes for conditional coding further in this paper, but it is an interesting topic for further research.

5. LSF SOURCE MEASUREMENTS

Here, the information measures regarding split quantization, discussed in Section 3, are applied to a 10-dimensional LSF representation of the speech spectral envelope. Since an analytic expression for the 10-dimensional LSF pdf, $f_{\mathbf{Z}}(\mathbf{z})$, is not available, we have chosen to base the simulations on a Gaussian mixture model (GMM) of the source pdf, $f_{\mathbf{Z}}^{\mathcal{M}}(\mathbf{z})$.

Below, we start with discussions on the simulation setup, the source modeling, and the calculation of the information measures, before presenting the actual results for the LSF source.

5.1. Setup

All modeling is based on the TIMIT speech database (clean speech), down sampled to 8 kHz. A 10th order LPC analysis using the autocorrelation method is performed every 20 ms, using a 25 ms Hamming window. A fixed 10 Hz bandwidth expansion is applied



Fig. 1: Modeling performance. Evaluation of loglikelihood, J, as function of the size of the training database, for a 128 mixture GMM, and as a function of the number of mixtures, for a training database of 700 000 vectors. Both the results from closed evaluations, evaluations using the same database as the GMM training, and open evaluations are shown (based on 250 000 vectors).

to each pole of the LPC coefficient vector, and finally the LPC vectors are transformed to the LSF representation.

All evaluation are based on a 128 mixture GMM. The training was conducted using a database of 700 000 LSF vectors, extracted from the *train* set in TIMIT. The modeling performance was evaluated using a database of 250 000 LSF vectors from the *test* set in TIMIT. The evaluation of information measures was performed using stochastic integration, based on $10 \cdot 10^6$ synthetically generated samples, c.f. Section 5.3.

5.2. Source Modeling

Evaluation of the information measures discussed above, requires an expression for the source pdf, $f_{\mathbf{Z}}(\mathbf{z})$. In this paper the joint source pdf is estimated using a GMM[16],

$$f_{\mathbf{Z}}^{\mathcal{M}}(\mathbf{z}) = \sum_{i=1}^{M} \rho_i f_{i,\mathbf{Z}}(\mathbf{z}), \qquad (14)$$

where M is the number of mixture components, and $f_{i,\mathbf{Z}}(\mathbf{z})$ are multivariate Gaussian densities.

Commonly the model parameters are found using the EMalgorithm [17]. The EM-algorithm guarantees a monotonic increase in log-likelihood,

$$J = \frac{1}{N_{\rm DB}} \ln \prod_{n=1}^{N_{\rm DB}} f_{\mathbf{Z}}^{\mathcal{M}}(\mathbf{z}_n), \tag{15}$$

for each EM-iteration, where $N_{\rm DB}$ represents the size of the training database. To get an appropriate model we need to perform enough EM-iterations, choose a model of appropriate order, and finally use a large enough database. We have found that for our purposes, a 128 mixture model trained on 700 000 vectors is satisfactory. This choice is partly based on the modeling performance shown in Fig. 1, where log-likelihood convergence as a function of the number of mixtures and the training database size are shown.

5.3. Evaluation of Information Measures

Direct evaluation of the integrals in the expressions for information theoretic measures are non-trivial for most source pdfs, including GMM representations. One possible approach is to resort to stochastic integration.

¹Here, we are comparing the dimension of the entire vector to be quantized and the dimensions of each split.



Fig. 2: Information theoretic split quantization loss, compared to unconstrained VQ, for a 10-dimensional LSF source. The results are based on evaluation of information theoretic measures, and should be considered as lower bounds. The crosses (×) show the loss for standard split quantization, c.f. Eq. (6). The circles (\circ) show the loss for 1-step conditional split quantization, c.f. Eq. (10). The triangles (\triangle) show the loss for 1-step, 1-element conditional split quantization, c.f. Eq. (12). Splitting has been performed as: (5, 5), (3, 3, 4), (2, 2, 3, 3), (2, 2, ..., 2), and (1, 1, ..., 1).

For example, in order to evaluate the differential entropy for a random vector \mathbf{X} using stochastic integration, we first note that the integral expression in Eq. (1) can be rewritten as an expectation,

$$h(\mathbf{X}) \approx -\mathbb{E}\left[\log_2\left(f_{\mathbf{X}}^{\mathcal{M}}(\mathbf{x})\right)\right],$$
 (16)

where the approximate equality is due to the usage of a model of the source pdf. Based on the law of large numbers we can now approximate the expectation with a sample mean,

$$h(\mathbf{X}) \approx -\frac{1}{N} \sum_{i=1}^{N} \left(\log_2 \left(f_{\mathbf{X}}^{\mathcal{M}}(\mathbf{x}_i) \right) \right), \tag{17}$$

where N is the number of samples used for evaluation. The samples, \mathbf{x}_i , are synthetically generated using the GMM. Alternatively, the samples can be drawn from a database (open or closed). In our case the discrepancy between these approaches are small, a fact that can be interpreted as a confirmation of the modeling.

5.4. Results

In Fig. 2, we present results giving the information theoretic loss for split quantization of a 10 dimensional LSF source, compared to unconstrained coding, and potential gains of conditional coding. All results are based on the information theoretic methods discussed in the previous section, and should be considered as lower bounds, c.f. Section 4.

The results show a split loss, $L_N^{(\text{split})}$, c.f. Eq. (6), ranging from 2 bits loss for a 2-split representation to an 8 bits loss for scalar representation. We can also see how conditioning on the split to the "left" $(L_N^{(\text{cond}-1)})$, c.f. Eq. (10), reduces the split loss to a maximum of approximately two bits. Naturally, a 2-split conditional representation performs without loss, c.f. Eq. (8). The results for the less complex approach of conditioning on one element of the split to the "left" $(L_N^{(\text{cond}-1,1)})$, c.f. Eq. (12), show a minimum loss of approximately 1 bit. For a scalar representation the performance coincide with the results according to, $L_N^{(\text{cond}-1)}$.

When interpreting these results one should remember the discussion in Section 4. Another source for "errors" is the pdf modeling. Remembering this, we can notice that the estimated loss for scalar quantization compared to unconstrained VQ, $L_{10}^{(\text{split})}$, of 8 bits, agree reasonably well with the 10-12 bit loss that can be found studying the literature [4, 3].

6. CONCLUSIONS

The main contribution of this paper is the use of information theoretic tools to bound losses regarding split quantization and conditioned split quantization compared to unconstrained VQ.

The results suggest that for a 10 dimensional LSF source, a large part of the loss due to split quantization can be recovered using conditional quantization. If an effective conditional coding can be implemented, a scalar quantization can essentially suffice.

7. REFERENCES

- M. R. Schroeder and B. S. Atal, "Code-excited linear prediction (CELP): High-quality speech at very low bit rates," in *Proc. ICASSP*, 1985, vol. 2, pp. 937–940.
- [2] F. Itakura, "Line spectrum representation of linear predictive coefficients," *Jour. of the Acous. Soc. of Am.*, vol. 57 Supplement, no. 1, pp. S35, 1975.
- [3] K. K. Paliwal and W. B. Kleijn, "Quantization of LPC parameters," in *Speech Coding and Synthesis*, pp. 433–466. Elsevier Science Publishers, Amsterdam, The Netherlands, 1995.
- [4] P. Hedelin and J. Skoglund, "Vector quantization based on Gaussian mixture models," *IEEE Trans. Speech and Audio Proc.*, vol. 8, no. 4, pp. 385–401, 2000.
- [5] A. Gersho and R. M. Gray, Vector Quantization and Signal Compression, Kluwer Academic Publishers, Boston, 1992.
- [6] K. K. Paliwal and B. S. Atal, "Efficient vector quantization of LPC parameters at 24 bits/frame," *IEEE Trans. Speech and Audio Proc.*, vol. 1, no. 1, pp. 3–14, 1993.
- [7] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, John Wiley & Sons, New York, NY, 1991.
- [8] T. D. Lookabaugh and R. M. Gray, "High-resolution quantization theory and the vector quantizer advantage," *IEEE Trans. Inf. Theory*, vol. 35, no. 5, pp. 1020–1033, 1989.
- [9] T. Eriksson, J. Lindén, and J. Skoglund, "Interframe LSF quantization for noisy channels," *IEEE Trans. Speech and Audio Proc.*, vol. 7, no. 5, pp. 495–509, 1999.
- [10] V. Cuperman and A. Gersho, "Vector predictive coding of speech at 16 kbits/s," *IEEE Trans. Comm.*, vol. 33, no. 7, pp. 685–696, 1985.
- [11] A. Gersho, "Optimal nonlinear interpolative vector quantization," *IEEE Trans. Comm.*, vol. 38, no. 9, pp. 1285–1287, 1990.
- [12] T. Eriksson and F. Nordén, "Memory vector quantization by power series expansion," in *Proc. IEEE Workshop on Speech Coding*, (Tsukuba, Japan), 2002, pp. 141–143.
- [13] F. K. Soong and B-H. Juang, "Optimal quantization of LSP parameters," *IEEE Trans. Speech and Audio Proc.*, vol. 1, no. 1, pp. 15–24, 1993.
- [14] M. Xie and J-P. Adoul, "Fast and low-complexity LSF quantization using algebraic vector quantizer," in *Proc. ICASSP*, 1995, vol. 1, pp. 716–719.
- [15] M. Y. Kim, N. K. Ha, and S. R. Kim, "Linked split-vector quantizer of LPC parameters," in *Proc. ICASSP*, 1996, vol. 2, pp. 741–744.
- [16] R. A. Redner and H. F. Walker, "Mixture densities, maximum likelihood and the EM algorithm," *SIAM Review*, vol. 26, no. 2, pp. 195– 239, 1984.
- [17] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society*, vol. 39, pp. 1–38, 1977.