RATE DISTRIBUTION BETWEEN MODEL AND SIGNAL FOR MULTIPLE DESCRIPTIONS

Janusz Klejsa and W. Bastiaan Kleijn

ACCESS Linnaeus Center, Electrical Engineering, KTH - Royal Institute of Technology, 10044 Stockholm, Sweden
{ janusz.klejsa, bastiaan.kleijn } @ee.kth.se

ABSTRACT

We consider the rate allocation problem for multiple-description quantization of the signal described by an adaptive model with a fixed structure. The source modeling in coding generally results in a two-stage description of the data, where one of the stages describes the model parameters, and the other describes the signal. Such a setup implies the existence of a trade-off between the rate spent on the parameters and the rate spent on the signal. We optimize this trade-off analytically for the multiple-description case using a method inspired by Minimum Description Length principle. We also provide an algorithm for optimizing the rate allocation between the components of the model-based multiple description coder. Finally we experimentally confirm our results. Our method facilitates the rate-adaptive multiple-description coding.

Index Terms— source modeling, multiple description coding (MDC), audio coding.

1. INTRODUCTION

We address a problem of finding an optimal rate allocation between model parameters and multiple descriptions of the signal in the context of model-based multiple description coding (MDC). Previously the problem of finding the optimal rate allocation was solved for a single description case [1] and we extend the solution to the multiple description case. Our reasoning resembles that used in the Minimum Description Length (MDL) principle [2, 3]. We aim at minimizing the total average rate required to transmit the signal subject to an average distortion constraint for model-based multiple-description vector quantization of the signal. We propose a method optimizing rate allocation between components of the system that yields the globally optimal performance.

It is known that direct vector quantization leads to the highest coding efficiency due to its memory, space-filling and shape advantages [4]. A drawback of vector quantization is that it requires a large storage memory especially in the case of multiple-description quantization. The problem becomes particularly severe for adaptive applications, where the coder should adapt to any rate constraint. Usage of source modeling provides an alternative that is relevant for adaptive applications [5].

MDC aims at combating the effect of packet losses by introducing redundancy and exploiting diversity provided by the network. Among many multiple-description techniques the quantization-based methods have proven to perform well [6]. In our work we focus on minimizing an average rate (entropy) required to achieve a constrained average distortion. It is known that in the single-description entropy-constrained case lattice quantizers are optimal [7]. In the multiple-description case the lattice quantizers are used because they simplify the design of an entropy-constrained system. One of the first multiple-description lattice quantizers was presented in [8]. In this paper for the purposes of explaining the theory and implementing the experiments we use the lattice-based MDC of [6].

The usage of statistical models is common in the coding of speech and audio signals. Such an approach improves the coding efficiency and facilitates obtaining scalability of the coders with respect to the bit-rate constraints. A common practice is to use a fixed structure of the model that is parameterized. Then the signal is quantized assuming a certain signal statistics conditioned on the model parameters. In practice the models describe short-term statistics of the data. They are adaptive and their parameters are estimated from the data and quantized. This implies a two-stage description of the data where the first stage specifies the model parameters and the second stage describes the quantized data conditioned on that model. The two-stage description introduces a trade-off between rate that is spent for encoding the parameters of the model and rate that is required to encode the quantized signal. Contrary to the single-description case, the multiple description case allows for usage of both single-stage (describing only the signal or only the model) and two-stage descriptions of the data. The usage of both kinds of descriptions affects the mentioned rate-allocation trade-off.

The combination of coding based on the usage of the source models with multiple description coding introduces a non-trivial problem of embedding the model information into the descriptions of the signal. Particularly the transmission of the model parameters should be robust to possible description-losses, since the full information about the model is required to perform the successful decoding. An example of a system combining MDC with GMM-based source modeling may be found in [9]. In this case optimization of the rate allocation becomes trivial. Our method provides a general treatment of finding the optimal rate distribution between elements of the model-based MDC system.

The paper is organized as follows. A presentation of the multiple-description quantization setup with the usage of a statistical model of the data is described in Section 2. For the considered setup, an optimal rate-allocation rule is derived in Section 3 and implemented for the case of coding with autoregressive (AR) model. In Section 4 we present an experimental evaluation of the proposed solution.

2. MULTIPLE DESCRIPTION QUANTIZATION WITH A SIGNAL MODEL

We consider encoding of a zero-mean stochastic scalar process X_i. To facilitate the encoding, the scalar samples are blocked into non-overlapping L-dimensional vectors yielding X^L. Let x^L denote a realization of X^L. Each realization x^L is encoded independently using knowledge about the pdf of the data $f_{X|\Theta}(x^L|\theta)$, which is parameterized by the vector of parameters $\theta(x^L)$. We consider the entropy-constrained multiple-description quantization of x^L that creates k

signal descriptions for each x^L.

We use a lattice-based multiple-description quantizer that is a construction consisting of one central lattice Λ_c and k identical side lattices Λ_S [6]. The side lattices are sublattices of Λ_C related geometrically to Λ_C by means of a so-called sublattice index $N = |\Lambda_C/\Lambda_S|$. Each central lattice cell has volume v and each side lattice cell has volume Nv. The index N allows for adjusting the redundancy between the descriptions.

There exists a bijective mapping between the central lattice and the sublattices, which is called an index assignment scheme. The mapping works as follows. Each x^L is first quantized with Λ_C. The quantized value is then mapped to a corresponding codeword in each of the side codebooks resulting in a k-tuple of codewords. We refer to elements of the tuple as descriptions. The descriptions are transmitted independently through the network such that the losses of the descriptions are independent. This requires the existence of diversity within the network, which can be represented as an assumption about the availability of k independent channels. During the transmission any of the descriptions may be lost. The index assignment is designed to allow the reconstruction of x^L for any subset of received descriptions. If all the descriptions are received, the mapping can be inverted and the correct reconstruction point from Λ_C can be selected. In case of losses x^L is reconstructed with lower fidelity by means of averaging of all received descriptions [6].

The usage of the source modeling implies that θ should be known at the decoder. The quantized vector of parameters θ can be included in any of the k descriptions resulting in a two-stage description. However, the transmission of θ with each signal description may lead potentially to a large rate overhead, since only the successful reception of one two-stage description is required to reconstruct the signal. Note that the multiple description quantization cannot be applied directly to the vector of model parameters in order to reduce the overhead because a loss of any model description likely destroys the mapping that was used during the encoding. We assume that m out of k descriptions contain full information about θ (see Fig. 1). Transmitting model information as separate descriptions has low practical relevancy, therefore it is not considered here.
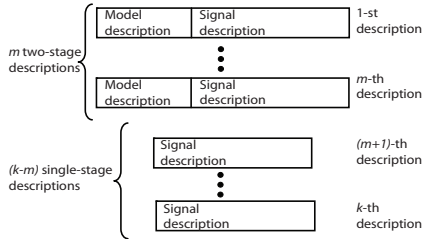

Fig. 1. Multiple-description coding scenario, k-description case.

Let $D^{(k,i)}$ be the average distortion (MSE) for the case of receiving i signal descriptions out of k including at least one two-stage description. Let $D_{k,m}$ be the weighted distortion for the system using m two-stage descriptions out of a total number of k descriptions. It is assumed that the description losses are independent for each description and that the probability of description loss is p for each of the descriptions. For the case of $m < k$ and $i \geq k - m + 1$, there are $\binom{k}{i}$ combinations of the descriptions that allow for reconstructing the x^L, for $1 \leq i \leq k - m$, there are $\binom{k}{i} - \binom{k-m}{i}$ such combinations. The remaining combinations of the descriptions make the decoding impossible due to lack of model information, and the associated distortion is equal to $\frac{\mathbb{E}_X\{\|X^L\|^2\}}{L}$. Thus, the weighted

distortion (MSE) is given by

$$D_{k,m}(v, N) = \sum_{i=k-m+1}^{k} \binom{k}{i}(1-p)^i p^{k-i} D^{(k,i)}(v, N) + \quad (1)$$

$$+ \sum_{i=1}^{k-m} \left(\binom{k}{i} - \binom{k-m}{i} \right)(1-p)^i p^{k-i} D^{(k,i)}(v, N) +$$

$$+ p^m \frac{\mathbb{E}_X\{\|X^L\|^2\}}{L}.$$

For the case of $k = m$ the second term of of the eq. (1) vanishes. $D^{(k,i)}$ can be computed as [6]

$$D^{(k,i)}(v, N) = \left(G(\Lambda_c) + \frac{k-i}{2ki} G(S_L)\phi_L^2 N^{\frac{2k}{L(k-1)}} \right) v^{\frac{2}{L}}, \quad (2)$$

where $G(\Lambda_c)$ and $G(S_L)$ are second dimensionless moments of inertia of Λ_c and an L-dimensional sphere respectively. ϕ_L is a so-called expansion factor [6].

3. OPTIMAL RATE ALLOCATION

In this section we derive the optimal rate allocation between model and signal. We consider a k-description case with m two-stage descriptions. The probability of description erasure is assumed p for each description. The rate used to describe the signal is the same for each description and equal to R_S (symmetric MDC). The two-stage descriptions have an additional R_M nats to describe the model information. The optimization problem is to find the R_M that minimizes the total rate for arbitrary v and N.

Assuming a high-rate (small central quantization cell volume v and source pdf approximately uniform within side quantization cell) for a given vector of parameters θ and sublattice index N, the length of a codeword required to create a single description of x^L is

$$L_{X|\Theta}(x^L|\theta(x^L)) = -\log\left(f_{X|\Theta}(x^L|\theta(x^L)vN) \right), \quad (3)$$

and the average rate is $R_S = \mathbb{E}_X\{L_{X|\Theta}(X^L|\theta(X^L))\}$.

A natural criterion for selecting the vector of model parameters $\hat{\theta}$ is then

$$\hat{\theta}(x^L) = \arg\max_\theta f_{X|\Theta}(x^L|\theta(x^L)), \quad (4)$$

which minimizes the codeword length required to describe x^L. However, $\hat{\theta}$ cannot be transmitted with a finite rate and it must be quantized yielding $\bar{\theta}$. Assume high-rate and that $\Theta = \theta(X^L)$ is distributed according to some probability mass function $f_\Theta(\theta)$. The codeword length required to describe $\bar{\theta}(x^L)$ is $L_{\bar{\Theta}}(\bar{\theta}(x^L)) = -\log\left(f_\Theta(\bar{\theta}(x^L)) \right)$. The average rate spent on encoding the model parameters is therefore $R_M = \mathbb{E}_X\{L_{\bar{\Theta}}(\bar{\theta}(X^L))\}$.

We finally obtain a coded version of x^L that consists of m two-stage descriptions and $k-m$ single-stage descriptions using the total rate

$$L(x^L) = mL_{\bar{\Theta}}(\bar{\theta}(x^L)) + kL_{X|\bar{\theta}}(x^L|\bar{\theta}(x^L)). \quad (5)$$

We are interested in minimizing $R_T = \mathbb{E}_X\{L(X^L)\}$ for constrained m, k and p, and finally a fixed side cell volume vN. The goal is to explicitly express the penalty to the total rate resulting from the usage of a quantized model $\bar{\theta}$ instead of the optimal $\hat{\theta}$. Thus we write the total rate as

$$R_T = \mathbb{E}_X\{\psi(\hat{\theta}, \bar{\theta}, X^L)\} - k\mathbb{E}_X\{f_{X|\hat{\Theta}}(X^L|\hat{\theta}(X^L))\} +$$
$$- k\log(Nv), \quad (6)$$

where we introduced the so-called *index of resolvability*

$$\psi(\hat{\theta}, \bar{\theta}, x^L) = -m \log f_{\Theta}(\bar{\theta}(x^L)) - k \log \frac{f_{X|\bar{\Theta}}(x^L|\bar{\theta}(x^L))}{f_{X|\hat{\Theta}}(x^L|\hat{\theta}(x^L))}. \quad (7)$$

We note that for arbitrary N and v the optimal rate for the model is minimizing the $\mathbb{E}_X\{\psi(\hat{\theta}, \bar{\theta}, X^L)\}$, since the other terms of eq. (6) do not depend on the quantized model.

3.1. Application to Autoregressive Modeling

Following the derivation in [1] we consider the case when autoregressive (AR) modeling is used to model the source. We assume that X^L has a Gaussian multivariate distribution

$$f_{X|\bar{\Theta}}(x^L|\bar{\theta}(x^L)) = \frac{1}{\sqrt{(2\pi)^L \det(R_{\Theta})}} \exp(-\frac{1}{2} x^{L^T} R_{\Theta}^{-1} x^L), \quad (8)$$

where R_{Θ} is model covariance matrix for X^L corresponding to the AR model with parameters θ. The result of [1] is that the mean rate penalty due to the quantization of the model is given by

$$\mathbb{E}_X\{\log \frac{f_{X|\bar{\Theta}}(X^L|\bar{\theta}(X^L))}{f_{X|\hat{\Theta}}(X^L|\hat{\theta}(X^L))}\} = \frac{L\alpha}{4} D(\bar{\Theta}, \hat{\Theta}), \quad (9)$$

$$\alpha \approx \frac{1}{L} x^{L^T} R_{\Theta}^{-1} x^L. \quad (10)$$

The mean model distortion $D(\bar{\Theta}, \hat{\Theta})$ is the mean log-spectral distortion. Then the index of resolvability for the MDC case takes the form

$$\mathbb{E}_X\{\psi(\hat{\theta}, \bar{\theta}, X^L)\} \approx m R_M + k \frac{L\alpha}{4} D(\bar{\Theta}, \hat{\Theta}). \quad (11)$$

The model distortion $D(\bar{\Theta}, \hat{\Theta})$ can be computed as

$$D(\bar{\Theta}, \hat{\Theta}) = C e^{-\frac{2}{d}(R_M - h(\hat{\Theta}))}, \quad (12)$$

where C is the coefficient of quantization for the model parameter quantizer, $h(\hat{\Theta})$ is the differential entropy measured in the log spectral domain and d is the dimensionality of the manifold that $\hat{\Theta}$ lies on (*cf.* [1]). Plugging eq. (12) into eq. (11) and differentiating with respect to R_M, we obtain the optimal rate allocation, for given N and v, which is given by

$$R_M^*(k, m) = h(\hat{\Theta}) + \frac{d}{2} \log\left(\frac{L\alpha}{2d} C\right) + \frac{d}{2} \log\left(\frac{k}{m}\right). \quad (13)$$

The last term of eq. (13) represents a penalty for the case of $m < k$. For $k = m$ the last term vanishes and we obtain the result of [1]. We note that for constant v and N the optimal R_M does not depend on the volume of the side lattice cell, which is true under the high-rate assumption.

3.2. Constrained weighted distortion case

The optimization problem is to find the R_M that minimizes the total rate R_T required to achieve a *constrained value of the weighted distortion* $D_{k,m}$. We first show that, given a packet loss rate, constraining the distortion and minimizing the last term in eq. (6) results in v and N that are constant. Thus, the optimal rate for the model can be obtained by minimizing the mean index of resolvability. Specifically, for the AR modeling the optimal rate for the model is given by eq. (13).

Constraining the weighted distortion $D_{k,m}$, we can compute v for arbitrary N using eq. (1) and eq. (2) obtaining

$$v(D_{k,m}, N) = \left(\frac{D_{k,m} - p^m \mathbb{E}_X \|X^L\|^2 / L}{K_1 G(\Lambda_c) + K_2 G(S_L) \phi_L^2 N^{\frac{2k}{L(k-1)}}}\right)^{\frac{L}{2}}, \quad (14)$$

where

$$K_1(k, m) = \sum_{i=k-m+1}^{k} \binom{k}{i} (1-p)^i p^{k-i} + \quad (15)$$
$$+ \sum_{i=1}^{k-m} \left(\binom{k}{i} - \binom{k-m}{i}\right) (1-p)^i p^{k-i},$$

$$K_2(k, m) = \sum_{i=k-m+1}^{k} \binom{k}{i} (1-p)^i p^{k-i} \frac{k-i}{2ki} + \quad (16)$$
$$+ \sum_{i=1}^{k-m} \left(\binom{k}{i} - \binom{k-m}{i}\right) (1-p)^i p^{k-i} \frac{k-i}{2ki}.$$

We define a side lattice cell volume $\tau(N) = N v(D_{k,m}, N)$. Minimizing the total rate (eq. (6)) with respect to N is the same as maximizing the volume $\tau(N)$. By using the expression for the weighted total distortion (eq. (1)) and eq. (2), we can derive the optimal $N_{opt} = \arg\max_N \tau(N)$ for given weighted distortion. After some algebra we find that

$$N_{opt}(k, m) = \left((k-1) \frac{K_1(k, m) G(\Lambda_c)}{K_2(k, m) G(S_L) \phi_L^2}\right)^{\frac{L(k-1)}{2k}}. \quad (17)$$

We note from eq. (17) that the optimal N is constant for a fixed k, m and p. Moreover for a fixed distortion $D_{k,m}$ and N the optimal $v(D_{k,m}, N)$ is constant and the minimization of the model rate can be still done within the mean index of resolvability. Our result shows that the optimal redundancy introduced by MDC is independent from the model rate R_M^*.

3.3. Optimal construction of the descriptions

Given a packet loss rate p, and a mean weighted distortion $D_{k,m}$ we minimize the rate R_T by finding the optimal k and m and optimizing a rate-allocation between the components of the system.

The following steps lead to the solution. For each possible combination of k and m we find the optimal N according to eq. (17). We note that the set of possible combinations is usually small due to the fact that a highest applicable k is low in practise. For each k and m we find the model rate that minimizes $\mathbb{E}_X\{\psi(\hat{\theta}, \bar{\theta}, X^L, m, k)\}$. We finally select a combination of k and m that minimizes

$$\begin{aligned} R_T(k, m) = \ & \mathbb{E}_X\{\psi(\hat{\theta}, \bar{\theta}, X^L, m, k)\} + \\ & -k \mathbb{E}_X\{f_{X|\hat{\Theta}}(X^L|\hat{\theta}(X^L))\} + \\ & -k \log(N_{opt}(m, k) v(D_{m,k}, N_{opt}(m, k))), \end{aligned} \quad (18)$$

subject to $D_{m,k} = D$. Specifically, for the case of AR modeling the optimal k and m should minimize the following cost function

$$\begin{aligned} \eta(k, m) = \ & m R_M^*(k, m) - k \mathbb{E}_X\{f_{X|\bar{\Theta}}(X^L|\bar{\theta}(X^L))\} + \\ & -k \log(N_{opt}(m, k) v(D_{m,k}, N_{opt}(m, k))), \ (19) \end{aligned}$$

subject to $D_{m,k} = D$.

4. EXPERIMENT

The experimental verification of our results was performed for the case of speech coding based on tenth-order AR modeling. Our experiment uses similar methodology as the experiment described in [1].

10000 speech blocks with 20ms length were extracted from the TIMIT database [10] and resampled to 8 kHz. A tenth-order AR model was estimated for each of the blocks resulting in a vector of linear prediction coefficients (LPC) that was first converted to the line spectral frequency (LSF) domain and scalar-quantized spending R_M bits. A diagonal sensitivity matrix was derived for each vector of coefficients and used during the quantization. Each block was divided into four non-overlapping subblocks. A subblock in the middle of the block was selected and normalized to obtain $\alpha = 1$. The KLT transform was computed using the quantized model and applied to the subblock obtaining a vector of transform coefficients.

Each vector of coefficients was then quantized using a scalar (central) quantizer and for each vector a necessary rate was estimated to obtain a certain SNR for the case of k-description quantization with m two-stage descriptions. The probability of description erasure was assumed to be $p = 0.01$ and for each considered configuration of the MDC an optimal value of N was computed to optimize the weighted distortion computed from eq. (1). The rate necessary for transmitting each of the descriptions has been estimated according to [6] and multiplied times 4 to obtain the rate for the whole block.

We estimated the differential entropy and the dimensionality of the manifold that $\hat{\Theta}$ lies on using the method of [11]. The dimensionality of the manifold was $d = 7.1$ and the differential entropy for the scalar quantization was 19.35 bits. We made the experiments for the cases of $k = m = 1$ with no losses in the channel, $k = 2, m = 2$ and $k = 3, m = 2$ for the case of lossy channel ($p = 0.01$). The aim of the experiment was to check if the optimal model rate computed according to eq. (13) indeed minimizes R_T. In the experiment we measure an excess rate over the lowest total rate that yields a certain value of SNR in a function of model rate. Our results confirm the theory that the optimal rate for the model is constant for high SNR. According to the theory, the optimal model rate for the cases of $k = m = 1$ and $k = m = 2$ is 19 bits per block and for the case $k = 3, m = 2$ is 21.1 bits per block.

Additionally we empirically observed the model rate-penalty for the case $k > m$ related to the last term of the eq. (13), which is 2.1 bits between the cases of $k = m$ and $k = 3, m = 2$.

5. CONCLUSIONS

We conclude that the rate allocation between components of the model-based lattice multiple description coder can be optimized analytically. The optimization is done subject to a distortion constraint and for a given packet loss rate. The considered setup allows for performing the optimization for a combination of two-stage and single stage descriptions. Our analysis of the rate-allocation problem shows that the optimal redundancy introduced by MDC does not depend on the model rate. We show that under the high rate condition the optimal rate for the model does not depend on the distortion constraint. Finally we provide a method to optimize the number of two-stage and single-stage descriptions.

We apply our theory to the case of AR modeling. By means of experiment we confirm that for a high rate and a fixed number of two-stage and single-stage descriptions the optimal rate for the model does not depend on the distortion constraint.

Our other experimental results indicate that using an optimized number of two-stage descriptions generally performs better than us-
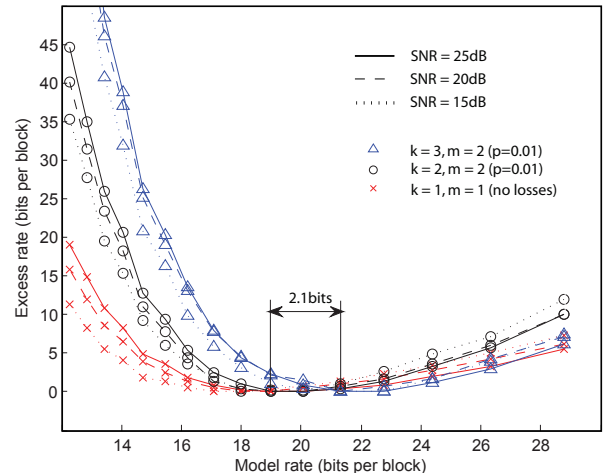

Fig. 2. Rate excess over the minimum total rate yielding a certain SNR for the cases of $k = m = 1$ and no losses, $k = m = 2$ and $p = 0.01$, $k = 3, m = 2$ and $p = 0.01$.

ing both two-stage and single-stage descriptions. It remains to be shown if this result is universal.

6. REFERENCES

[1] W. B. Kleijn and A. Ozerov, "Rate distribution between model and signal," in *IEEE Worksh. on Apps. of Signal Processing to Audio and Acoustics (WASPAA)*, Mohonk, NY, Oct. 2007, pp. 243–246.

[2] J. Rissanen, "Modeling by the shortest data description," *Automatica*, vol. 14, pp. 465–471, 1978.

[3] A. Barron, J. Rissanen, and B. Yu, "The minimum description length principle in coding and modeling," *IEEE Trans. Inform. Th.*, vol. 44, pp. 2743–2760, 1998.

[4] T. Lookabough and R. M. Gray, "High-resolution quantization theory and the vector quantizer advantage," *IEEE Trans. Inform. Th.*, vol. 35, no. 5, pp. 1020–1033, 1989.

[5] M. Y. Kim and W. B. Kleijn, "KLT-based adaptive classifeid VQ of the speech signal," *IEEE Trans. Speech Audio Processing*, vol. 12, pp. 277–289, May 2004.

[6] J. Østergaard, J. Jensen, and R. Heusdens, "n-channel entropy-constrained multiple-description lattice vector quantization," *IEEE Trans. Inform. Th.*, vol. 52, pp. 1956 – 1973, 2006.

[7] A. Gersho, "Asymptotically optimal block quantization," *IEEE Trans. Inform. Th.*, vol. 25, pp. 373–380, 1979.

[8] S. D. Servetto, V. A. Vaishampayan, and N. J. A. Sloane, "Multiple description lattice vector quantization," in *Data Compression Conference*, March 1999, pp. 13–22.

[9] J. Samuelsson and J. H. Plasberg, "Multiple description coding based on Gaussian Mixture Models," *IEEE Sig. Proc. Letters*, vol. 12, pp. 449 – 452, 2005.

[10] DARPA-TIMIT, "Acoustic-phonetic continuous speech corpus," NIST Speech Disc 1-1.1, 1990.

[11] M. Nilsson and W. B. Kleijn, "On the estimation of differential entropy from data located on embedded manifolds," *IEEE Trans. Inform. Th.*, vol. 52, pp. 2330–2341, Jul 2007.