

PDF-optimized LSF Vector Quantization Based on Beta Mixture Models

Zhanyu Ma, Arne Leijon

KTH Royal Institute of Technology, Sound and Image Processing Lab
Stockholm, Sweden

zhanyu@kth.se, leijon@kth.se

Abstract

The line spectral frequencies (LSF) are known to be the most efficient representation of the linear predictive coding (LPC) parameters from both the distortion and perceptual point of view. By considering the bounded property of the LSF parameters, we apply beta mixture models (BMM) to model the distribution of the LSF parameters. Meanwhile, by following the principles of probability density function (PDF) optimized vector quantization (VQ), we derive the bit allocation strategy for the BMM. The LSF parameters are obtained from the TIMIT database and a practical VQ is designed. By taking the Bayesian information criterion (BIC), the square error (SE) and the spectral distortion (SD) as the criteria, the BMM based VQ outperforms the Gaussian mixture model based VQ with uncorrelated Gaussian component (UGMVQ) by about 1 ~ 2 bits/vector.

Index Terms: speech coding, vector quantization, line spectral frequencies, beta mixture model, Gaussian mixture model

1. Introduction

In the application of audio signal processing and speech processing, the linear predictive coding (LPC) is the mostly used tool for representing the spectral envelope of a digital signal of speech in the compressed form. Among several representations of LPC parameters, the line spectral frequencies (LSF) were shown to be the most efficient representation of the LPC parameters [1, 2] from both the distortion and perceptual point of view.

Encoding and decoding the LSF parameters effectively is the essential task of speech coding and transmission. In the practical application, how to obtain the codebook from the training set efficiently is a challenge in recent studies. The probability density function (PDF) based vector quantization (VQ) can overcome the problem of training the VQ with insufficient data. Furthermore, the codebook obtained by PDF-optimized VQ can prevent the codebook from overfitting to the training set [3, 4]. The important step in PDF-optimized VQ is to find a proper parametric distribution to model the data. Most of the literature utilized Gaussian mixture model (GMM) for the distribution and reached a good performance [3, 5, 6, 7]. One of the problems of modelling the LSF parameters with GMM is the mismatch between the support range of LSF parameters $[0, \pi]$ and the support range of Gaussian distribution $[-\infty, \infty]$. Lindblom et al. [5] proposed a bounded support GMM for LSF parameter quantization to adjust the model to fit the distribution of the data. But this model is still computationally expensive when training the bounded support GMM.

Some bounded support distributions (e.g. beta mixture models (BMM), Dirichlet mixture models) were applied to model the data with bounded support range (e.g. the image pixel values) and shown to outperform the GMM [8, 9]. Since the beta

distribution is designed to model bounded support data, in this paper, we model the distribution of the LSF parameters with BMM and propose the PDF-optimized VQ based on the trained BMM. With the high rate assumption, the distortion-rate (DR) relation of the beta distribution is derived. According to the PDF optimized VQ design principles [4], we propose the bit allocation strategy for the BMM in both the fixed rate and variable rate cases. The implementation algorithm is also introduced.

One issue should be stated in this paper. Since there is no multivariate expression of the beta distribution (the Dirichlet is not the true multivariate case because of the constraint of summation to 1 for the variables), we assume the LSF parameters in each component are independent and use instead a sufficient number of mixture components to model the dependence among the LSF parameters. As pointed in [5, 6], the mixture model with uncorrelated mixture component can model the correlations by increasing the number of mixture components and the GMM based VQ with uncorrelated Gaussian component (UGMVQ) [6] could reach the similar performance as the full GMM based VQ. Thus we apply this zero correlation assumption in our paper. With the zero correlation assumption, the VQ is optimally replaced by a set of independent scalar quantizers (SQ), which reduces the computational cost of quantization.

The TIMIT database [10] is used to evaluate the performance. The 16-order LPC parameters are extracted from the database and converted to 16-dimensional LSF parameters. By taking the Bayesian information criterion (BIC) as the criterion, the BMM models the data better than the GMM. We quantize the LSF parameters with the proposed BMM based VQ and the UGMVQ at the same level of model complexity. The experimental results show that the BMM based VQ performs better than the UGMVQ from the aspect of the LSF quantization square error (SE) and spectral distortion (SD).

2. Model the LSF Parameters with BMM

The LSF parameters were introduced by Itakura [11] and the LSFs are widely used for speech coding because of the advantages compared to some other forms of representation (such as LARs, ASRCs). For a linear predictive model with order K , the LSFs are interleaved on the unit circle [1] as

$$0 = \omega_{q_0} < \omega_{p_1} < \omega_{q_1} < \dots < \omega_{q_{\frac{K}{2}}} < \omega_{p_{\frac{K}{2}+1}} = \pi. \quad (1)$$

Then the LSF parameters are defined as

$$\mathbf{x} = [x_1, x_2, \dots, x_K]^T = [\omega_{p_1}, \omega_{q_1}, \dots, \omega_{p_{\frac{K}{2}}}, \omega_{q_{\frac{K}{2}}}]^T. \quad (2)$$

Since the LSFs are strictly bounded in the interval $[0, \pi]$, they can be linearly compressed into the interval $[0, 1]$.

The beta distribution is a family of continuous probability distributions defined on the interval $[0, 1]$ with two positive real

parameters. The probability density function (PDF) of the beta distribution is

$$\text{Beta}(x; \alpha, \beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} \quad (3)$$

where $\Gamma(\cdot)$ is the Gamma function defined as

$$\Gamma(z) = \int_0^{\infty} t^{z-1} e^{-t} dt. \quad (4)$$

Several studies [9, 12, 13] showed that the beta mixture models (BMM) could model the distribution of bounded data better than some other mixtures (e.g. Gaussian mixture models (GMM)). According to [4], modelling the data distribution better could lead to a better performance in PDF-optimized VQ design. Therefore, we apply the BMM in this paper to model the distribution of the LSF parameters and design the VQ based on BMM.

The K dimensional LSF parameter \mathbf{x} can be modelled with a mixture of beta distributions as

$$f(\mathbf{x}) = \sum_{i=1}^I \pi_i \prod_{k=1}^K \text{Beta}(x_k; \alpha_{k,i}, \beta_{k,i}) \quad (5)$$

where I is the number of mixture components, π_i is the prior probability of the i th mixture component ($\sum_{i=1}^I \pi_i = 1$, $\pi_i > 0$) and $\{\alpha, \beta\}$ are the parameters.

Although the LSF parameters are correlated, the correlation among the elements x_1, \dots, x_K can be captured by the mixture model with uncorrelated mixture component and we assume that the LSF vector \mathbf{x} is conditionally uncorrelated generated from one of the mixture components. In practise, by increasing the number of mixture components, this mixture model with uncorrelated mixture component is capable of performing similarly as the mixture models with correlated mixture component and could achieve a lower complexity [5, 6]. In the remaining part of this paper, we assume the independence among the LSF parameters and design a practical scalar quantizer to replace the vector quantizer.

Several estimation methods of the parameters in the BMM were proposed in [9, 12, 14] and we apply the conventional expectation maximization (EM) method to obtain the parameters of the BMM from the LSFs.

3. PDF-optimized VQ Design

In practical VQ design, one problem is when the training set is not sufficiently large, the codebook obtained may tend to overfit to the training set and the performance of the VQ could decrease. If the PDF of the data source is known, we can design a VQ based on the PDF-optimized scheme to overcome the above problem theoretically by generating a sufficiently large training set from the trained PDF. By using a proper PDF, the performance of VQ could be improved [3, 4].

3.1. Distortion Rate Relation with High Rate Assumption

With the high rate theory [15], we can derive the D-R relation with the high rate assumption for the beta distribution. The D-R relations are derived for both the constrained resolution and the constrained entropy case. We firstly derive the D-R relation for the scalar case and then expand it to the vector case with the independence assumption.

3.1.1. Constrained Resolution

In the constrained resolution case, the number of the quantization levels is given as $N = 2^R$. The D-R relation is as

$$\begin{aligned} D_{CR}^s(R) &= C \times 2^{-2R} \left[\int \text{Beta}(x)^{\frac{1}{3}} dx \right]^3 \\ &= C \times 2^{-2R} \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \left\{ \int [x^{\alpha-1}(1-x)^{\beta-1}]^{\frac{1}{3}} dx \right\}^3 \\ &= C \times 2^{-2R} \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \left[\frac{\Gamma(\frac{\alpha+2}{3})\Gamma(\frac{\beta+2}{3})}{\Gamma(\frac{\alpha+\beta+4}{3})} \right]^3 \end{aligned} \quad (6)$$

where C is a constant.

3.1.2. Constrained Entropy

With a constraint on the entropy, given R bits, the D-R relation is

$$D_{CE}^s(R) = C \times 2^{-2[R - h_{\text{Beta}}(x)]} \quad (7)$$

where

$$\begin{aligned} h_{\text{Beta}}(x) &= \log_2 \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta)} + \log_2 e [(\alpha + \beta - 2)\psi(\alpha + \beta)] \\ &\quad - \log_2 e [(\alpha - 1)\psi(\alpha) + (\beta - 1)\psi(\beta)] \end{aligned}$$

and $\psi(\cdot)$ is the digamma function.

3.2. Inter-component Bit Allocation Strategy for BMM

By taking the advantage of the mixture models, given a total number of R_{tot} bits, the total distortion $D_{tot} = \sum_{i=1}^I \pi_i D_i^c(R_i)$ can be minimized by assigning the i th mixture component with R_i bits and taking the weighted summation of the mixture component distortion $D^c(R_i)$, $i = 1, \dots, I$. Either the D-R relation under the constrained resolution case or the D-R relation under the constrained entropy case can be used to find the optimal bit allocation strategy. We only consider the constrained entropy case in this paper. With the zero correlation assumption, the D-R relation under the constrained entropy case for the mixture component can be expanded from (7) as

$$D^c(R) \triangleq D_{CE}^c(R) = C(K) \times 2^{-\frac{2}{K}[R - \sum_{k=1}^K h_{\text{Beta}}(x_k)]} \quad (8)$$

where $C(K)$ is a constant depends on K .

3.2.1. Fixed Rate

For the fixed rate case, the constraint for minimizing the total distortion is $2^{R_{tot}} = \sum_{i=1}^I 2^{R_i}$. With a Lagrange multiplier, an objective function can be built as

$$\eta = \sum_{i=1}^I \pi_i D_i^c(R_i) + \lambda \left(\sum_{i=1}^I 2^{R_i} - 2^{R_{tot}} \right). \quad (9)$$

Taking the derivative of η with respect to R_i , we have

$$\frac{\partial \eta}{\partial R_i} = -\frac{2}{K} \ln 2C(K) \pi_i 2^{-\frac{2}{K}[R_i - \sum_{k=1}^K h_{\text{Beta},i}(x_k)]} + \ln 2 \lambda 2^{R_i} \quad (10)$$

where $h_{\text{Beta},i}(x_k)$ is the entropy of the k th dimension in the i th mixture component. By setting (10) equal to 0 and after some algebra, we have

$$2^{R_i} \propto \left\{ \pi_i 2^{\frac{2}{K}[\sum_{k=1}^K h_{\text{Beta},i}(x_k)]} \right\}^{\frac{K}{K+2}} \quad (11)$$

and finally we can obtain

$$2^{R_i} = 2^{R_{tot}} \times \frac{\left\{ \pi_i 2^{\frac{2}{K}[\sum_{k=1}^K h_{\text{Beta},i}(x_k)]} \right\}^{\frac{K}{K+2}}}{\sum_{i=1}^I \left\{ \pi_i 2^{\frac{2}{K}[\sum_{k=1}^K h_{\text{Beta},i}(x_k)]} \right\}^{\frac{K}{K+2}}}. \quad (12)$$

3.2.2. Variable Rate

For the variable rate case, the constraint is $R_{tot} = \sum_{i=1}^I \pi_i R_i + \log_2 I$. We build an objective function with a Lagrange multiplier as

$$\eta = \sum_{i=1}^I \pi_i D_i^c(R_i) + \lambda \left(\sum_{i=1}^I \pi_i R_i - R_q \right) \quad (13)$$

where $R_q = R_{tot} - \log_2 I$ is the number of available bits for quantizing the variable. Taking the derivative of η with respect to R_i , we have

$$\frac{\partial \eta}{\partial R_i} = -\frac{2}{K} \ln 2C(K) \pi_i 2^{-\frac{2}{K} [R_i - \sum_{k=1}^K h_{Beta,i}(x_k)]} + \lambda \pi_i.$$

The relation between R_i and $\sum_{k=1}^K h_{Beta,i}(x_k)$ is

$$\begin{aligned} R_i - \sum_{k=1}^K h_{Beta,i}(x_k) &= M \quad (M \text{ is a constant}) \\ \Rightarrow \sum_{i=1}^I \pi_i \left[R_i - \sum_{k=1}^K h_{Beta,i}(x_k) \right] &= \sum_{i=1}^I \pi_i M = M \\ \Rightarrow R_i - \sum_{k=1}^K h_{Beta,i}(x_k) &= \sum_{i=1}^I \pi_i \left[R_i - \sum_{k=1}^K h_{Beta,i}(x_k) \right] \\ \Rightarrow R_i &= R_q + \sum_{k=1}^K h_{Beta,i}(x_k) - \sum_{i=1}^I \pi_i \left[\sum_{k=1}^K h_{Beta,i}(x_k) \right]. \end{aligned}$$

3.3. Intra-component Bit Allocation Strategy for BMM

If we assign equal distortion weights to each dimension of the LSF vector $\mathbf{x} = [x_1, \dots, x_K]^T$, then the bits R_i allocated for the i th component can be distributed to each dimension $k = 1, \dots, K$ according to the entropy of each dimension $h_{Beta}(x_k)$. Considering the constrained entropy case and with similar steps as in section 3.2, the bits allocated to the k th dimension in the i th mixture component is

$$R_{k,i} = \frac{R_i}{K} + h_{Beta,i}(x_k) - \frac{1}{K} \sum_{k=1}^K h_{Beta,i}(x_k). \quad (14)$$

The bit allocation obtained by (14) could allocate fractional bits. In practical implementation, the number of bits for each dimension is rounded off to satisfy the total rate constraint.

3.4. Quantization Algorithm

Since we have already allocated bits for each dimension in each mixture component, a practical designing scheme of SQ based on beta distribution is needed to obtain the codebook for each dimension in each mixture component. Under the high rate assumption, the optimal distribution of the centroid could be found. But with two different beta distributions, the transformation between two optimal centroid distributions is not linear and a closed form of the optimal transformation can not be found.¹ One possible way to solve this problem is to train the codebook with Lloyd algorithm for each beta distribution separately, which is highly computationally expensive. Another way is to apply the compander scheme. The codebook can be obtained by a compress-expand process via a uniform quantizer

¹For the case of Gaussian distribution, the transformation between two optimal centroid distribution is linear. With the same quantization level, if we have obtained the codebook for one Gaussian distribution (e.g. a Gaussian distribution with unit variance), the codebook for any other Gaussian distribution could be optimally obtained by a closed form linear transformation.

for each beta distribution. This method is more efficient than exhaustive training. But when applying high rate assumption to the low rate case, the compander via the uniform distribution performs worse compared to obtaining the codebook by training [16]. A scheme of designing a practical scalar quantizer for beta distribution is proposed in [16]. That scheme takes the advantages of both training and compander. In this paper, we applied that scheme to design the scalar quantizer for each dimension in each mixture component.

The algorithm of BMM based VQ is as follows (in the variable rate case):

1. Train the BMM with EM algorithm;
2. Allocate bits to each dimension in each mixture component with the strategy proposed above;
3. For an LSF vector \mathbf{x} , quantize it with all the mixture components and choose one mixture component by the minimum square error criterion². The VQ is realized by a set of SQs;
4. Transmit the index of the mixture component as well as the indices of each dimension;

4. Experimental Results and Discussion

We evaluate the BMM based VQ with the LSF parameters extracted from the TIMIT database [10], where the speech is sampled at 16 kHz. The dimension of the LSF parameters is $K = 16$ for the wide-band data. The training set consists of 100,000 LSF parameter vectors and the test set consists of 10,000 LSF parameter vectors which is different from the training set. Both the training set and the test set are selected randomly from the database. The LSF parameters are linearly compressed into $[0, 1]$ by dividing each dimension by π . Both the BMM and GMM are composed of 64 uncorrelated mixture components. The expectation maximization algorithm is applied to train the parameters in BMM and GMM. The test set is used to evaluate the VQ performance. The mean results are reported with 25 rounds of such simulations.

Fig. 1 shows the LSF parameter distributions (the first and second dimensions) and the corresponding obtained BMM and uncorrelated GMM. It is very clear that BMM models the distribution better than uncorrelated GMM: 1. the BMM captures the shape better at the bottom of the distribution; 2. uncorrelated GMM violates the $[0, 1]$ limitation for the LSF parameters in the 1st dimension. The BIC includes the trade-off between the likelihood and the model complexity. In our experiment, the BIC values for BMM and uncorrelated GMM are $BIC_B = -3.7729 \times 10^5$ and $BIC_G = -3.6713 \times 10^5$ respectively. We prefer BMM because of the smaller BIC value.

After obtaining the parameters for BMM and uncorrelated GMM, we quantize the LSF parameters with BMM and uncorrelated GMM respectively. With the variable rate bit allocation strategy in section 3.4, we quantize the LSF parameter with each mixture component and choose one of them by the minimum SE criterion. Once the mixture component is chosen, the index of the mixture component and the indices of the corresponding SQ is transmitted. The quantized LSF parameter is used to calculate the SD. The SD for the n th LSF frame is given as

$$SD_n = \sqrt{\frac{1}{F_s} \int_0^{F_s} \left[10 \log_{10} P_n(f) - 10 \log_{10} \hat{P}_n(f) \right]^2 df}$$

where F_s is the sampling frequency in Hz, $P_n(f)$ and $\hat{P}_n(f)$ are the original and quantized LPC power spectra of the n th

²There is a monotonic relation between SD and SE. To reduce the complexity, we choose the SE as the criterion here.

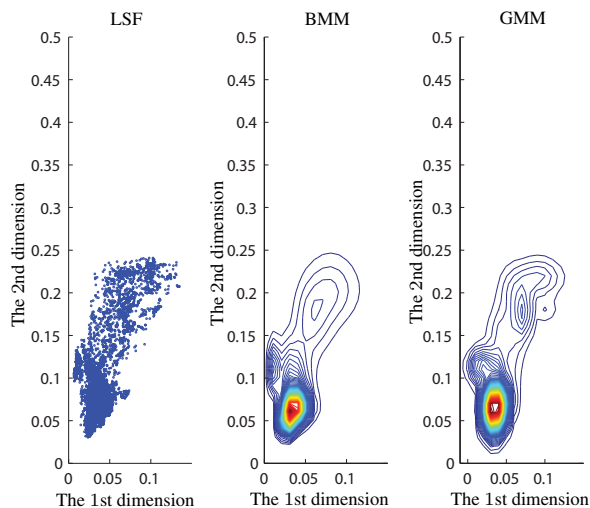


Figure 1: Comparison of modelling the LSF parameters with BMM and GMM.

frame. $P(f)$ and $\hat{P}(f)$ are calculated as

$$P_n(f) = 1/|A_n(e^{j2\pi f/F_s})|^2, \quad A(z) = 1 + \sum_{k=1}^K a_k z^{-k}$$

$$\hat{P}_n(f) = 1/|\hat{A}_n(e^{j2\pi f/F_s})|^2, \quad \hat{A}(z) = 1 + \sum_{k=1}^K \hat{a}_k z^{-k}$$

where a_k , $k = 1, \dots, K$ are the corresponding LPC parameters in the n th frame. The mean SD is evaluated from all the test LSF frames. The performance of BMM based VQ is shown in Table 1. It is observed that the BMM based VQ could reach the transparent coding with around 45 ~ 47 bits/vector. For comparison, the performance of UGMVQ is shown in Table 2. The BMM based VQ shows the advantage of 1 ~ 2 bits/vector over the UGMVQ. Although we allocate bits and selected the mixture component by SE in the quantization procedure, the monotonic relation between the SE and SD is verified in both tables. For future work, applying some other VQ design scheme (e.g. split VQ, switched split VQ) based on BMM may further improve the performance.

5. Conclusion

LSF parameter quantization is the essential part of audio/speech coding. In this paper, we proposed an algorithm for designing a practical VQ based on BMM. Since the LSF parameters are strictly bounded in $[0, \pi]$ and could be linearly compressed into the interval $[0, 1]$, the BMM was applied to model the distributions of the LSF parameters. Following the high rate theory and PDF optimized VQ design scheme, the bit allocation strategy for BMM based VQ was derived. The 16-dimensional LSFs extracted from the TIMIT database were used to evaluate the performance of the BMM based VQ. Compared to UGMVQ, BMM can model the LSFs better by taking the BIC as criterion. For transparent coding, BMM based VQ performs better than the UGMVQ. An advantage of 1 ~ 2 bits/vector could be achieved by using BMM based VQ.

6. References

- [1] F. Soong and B. Juang, "Line spectrum pair (LSP) and speech data compression," in *Acoustics, Speech, and Signal Processing, 1984. IEEE International Conference on*, vol. 9, Mar 1984, pp. 37–40.
- [2] K. Paliwal and B. Atal, "Efficient vector quantization of LPC parameters at 24 bits/frame," *Speech and Audio Processing, IEEE Transactions on*, vol. 1, no. 1, pp. 3–14, Jan 1993.

Table 1: Performance of proposed BMM based VQ

bits/vec.	avg. SD (dB)	SD outliers (in %)		SE/vec. ($\times 10^{-3}$)
		2 – 4 dB	>4 dB	
45	1.0553	0.9138	0.0229	0.2344
46	1.0053	0.5946	0.0000	0.2166
47	0.9731	0.4777	0.0000	0.1996

Table 2: Performance of UGMVQ

bits/vec.	avg. SD (dB)	SD outliers (in %)		SE/vec. ($\times 10^{-3}$)
		2 – 4 dB	>4 dB	
47	1.0417	0.5438	0.0157	0.2334
48	0.9997	0.3631	0.0000	0.2142
49	0.9619	0.2808	0.0000	0.1969

- [3] P. Hedelin and J. Skoglund, "Vector quantization based on Gaussian mixture models," *Speech and Audio Processing, IEEE Transactions on*, vol. 8, no. 4, pp. 385–401, Jul 2000.
- [4] A. D. Subramaniam and B. D. Rao, "PDF optimized parametric vector quantization of speech line spectral frequencies," *Speech and Audio Processing, IEEE Transactions on*, vol. 11, no. 2, pp. 130–142, Mar 2003.
- [5] J. Lindblom and J. Samuelsson, "Bounded support Gaussian mixture modeling of speech spectra," *Speech and Audio Processing, IEEE Transactions on*, vol. 11, no. 1, pp. 88–99, Jan 2003.
- [6] S. Chatterjee and T. V. Sreenivas, "Low complexity wide-band lsf quantization using GMM of uncorrelated Gaussian mixtures," in *16th European Signal Processing Conference (EUSIPCO 2008)*, 2008.
- [7] D. Zhao, J. Samuelsson, and M. Nilsson, "On entropy-constrained vector quantization using gaussian mixture models," *Communications, IEEE Transactions on*, vol. 56, no. 12, pp. 2094–2104, december 2008.
- [8] N. Bouguila and D. Ziou, "A Dirichlet process mixture of Dirichlet distributions for classification and prediction," in *Machine Learning for Signal Processing, 2008. IEEE Workshop on*, Oct. 2008, pp. 297–302.
- [9] Z. Ma and A. Leijon, "Beta mixture models and the application to image classification," in *Image Processing (ICIP), 2009 16th IEEE International Conference on*, nov. 2009, pp. 2045–2048.
- [10] "DARPA-TIMIT," *Acoustic-phonetic continuous speech corpus, NIST Speech Disc 1.1-1*, 1990.
- [11] F. Itakura, "Line spectrum representation of linear predictive coefficients of speech signals," *Journal of the Acoustical Society of America*, vol. 57, p. 535, 1975.
- [12] N. Bouguila, D. Ziou, and E. Monga, "Practical Bayesian estimation of a finite beta mixture through Gibbs sampling and its applications," *Statistics and Computing*, vol. 16, pp. 215–225, 2006.
- [13] Z. Ma and A. Leijon, "Human skin color detection in RGB space with Bayesian estimation of beta mixture models," in *2010 18th European Signal Processing Conference (EUSIPCO 2010)*, 2010.
- [14] —, "Bayesian estimation of beta mixture models with variational inference," submitted.
- [15] W. B. Kleijn, *A basis for source coding*, 2010, KTH lecture notes.
- [16] Z. Ma and A. Leijon, "Coding bounded support data with beta distribution," in *Network Infrastructure and Digital Content, 2010. IC-NIDC 2010. IEEE International Conference on*, 2010.