2014 Data Compression Conference

# On Optimal Coding of Hidden Markov Sources

Mehdi Salehifar, Emrah Akyol, Kumar Viswanatha, and Kenneth Rose

Department of Electrical and Computer Engineering
University of California, Samta Barbara, CA, 93106
E-mail:{salehifar,eakyol,kumar,rose}@ece.ucsb.edu

**Abstract**

The hidden Markov model (HMM) is widely used to model processes in several real world applications, including speech processing and recognition, image understanding and sensor networks. A problem of concern is that of quantization of the sequence of observations generated by an HMM, which is referred as a hidden Markov source (HMS). Despite the importance of the problem, and the well-defined structure of the process, there has been very limited work addressing the optimal quantization of HMS, and conventional approaches focus on optimization of parameters of known quantization schemes. This paper proposes a method that directly tracks the state probability distribution of the underlying source and optimizes the encoder structure according to the estimated HMS status. Unlike existing approaches, no stationarity assumption is needed, and code parameters are updated on the fly: with each observation, both the encoder and the decoder refine the estimated probability distribution over the states. The main approach is then specialized to a practical variant involving switched quantizers, and an algorithm that iteratively optimizes the quantizer codebooks is derived. Numerical results show superiority of the proposed approach over prior methods.

## 1 Introduction

The Hidden Markov model is a discrete-time finite state Markov chain observed through a memoryless channel. The random process consisting of the sequence of observations is referred to as a hidden Markov source (HMS). Markov chains are common models for information sources with memory and memoryless channel is among the simplest communication models. HMMs are widely used in image understanding and speech recognition [1], source coding [2], communications, information theory, economics, robotics, computer vision and several other disciplines. Note that most signals modeled as Markov process are in fact captured by imperfect sensors and are hence contaminated with noise, i.e., the resulting sequence is an HMS. Motivated by its modeling capability of practical sources with memory, in this paper we consider optimal quantization of the HMS . One conventional approach to design quantizers for the HMS is to employ predictive coding techniques (such as DPCM) to exploit time correlations, followed by standard scalar quantization of the residual. However, any direct prediction method cannot fully exploit the structure of the source process, i.e., the underlying Markov process. Indeed, even if the underlying process is first order Markov (depends on the past only through the previous sample), the HMS is not Markov, i.e, all prior samples are needed to optimally predict the current sample. But a naive infinite order prediction is clearly of impractical complexity. An alternative to predictive coding is encoding by a finite-state quantizer (see eg. [3, 4]).A

1068-0314/14 $31.00 © 2014 IEEE
DOI 10.1109/DCC.2014.71

finite-state quantizer is a finite-state machine used for data compression: each successive source sample is quantized using a quantizer code book that depends on the encoder state. The current encoder state and the codeword obtained by quantization then determine the next encoder state, i.e., the next quantizer codebook. In [5], the authors optimized the finite-state quantizer for a given HMS with known parameters, and analyzed its steady state performance. While a finite state quantizer is, in general, an effective approach, it does not fully exploit the underlying structure of the HMS. In other words, the finite state quantizer is agnostic to the *source state*, although it updates the *encoder state* based on the HMS observations. This subtle but important shortcoming motivates a scheme that *explicitly* considers the source states in the quantization process, which is the approach pursued in this paper.

The fundamental approach is to exploit all available information on the source state, i.e., the probability distribution over the states of the underlying Markov chain. An important distinction of this paradigm from prior work, and specifically from the finite-state quantizer, is that the state probability distribution captures all available information and depends on the entire history of observations. Hence, with each observation both encoder and decoder refine the estimate of state probability distribution and correspondingly update the coding rule. The fundamental approach is therefore optimal. We then specialize to a practical "codebook switching" variant and propose an algorithm that optimizes the decision rule, i.e., we optimize the codebook used for quantization based on the estimate of state probabilities.

The rest of the paper is organized as follows. In Section 2, we state the problem . In Section 3, we present the proposed method. We illustrate the effectiveness of our approach by comparative numerical results in Section 4, and conclude in Section 5 with future work and possible extensions to other problem settings.

## 2 Problem Definition

A hidden Markov source (HMS) is determined by five parameter sets (see eg., [6]):

1. The number of states in the model $N$. We denote the set of all possible states as $S = \{S_1, S_2, \ldots, S_N\}$, and the state at time $t$ as $q_t$.

2. The state transition probability distribution $A = \{a_{ij}\}$, where $a_{ij} = P[q_{t+1} = S_j | q_t = S_i], 1 \leq i, j \leq N$.

3. Observation symbols, in discrete case we denote the symbols as $V = \{v_1, v_2, \cdots, v_M\}$, where $M$ is the number of distinct observation symbols per state.

4. The observation (emission) probability density function (pdf) in state $j$, $g_j(x)$, in the continuous case, and the observation (emission) probability mass function (pmf), $B = \{b_j(v_k)\}$, where $b_j(v_k) = p[O_t = v_k | q_t = S_j]$ in the discrete case. $O_t$ denotes the source output at time $t$.

5. The initial state distributions $\pi = \{\pi_i\}$ where $\pi_i = P[q_1 = S_i], 1 \leq i \leq N$.

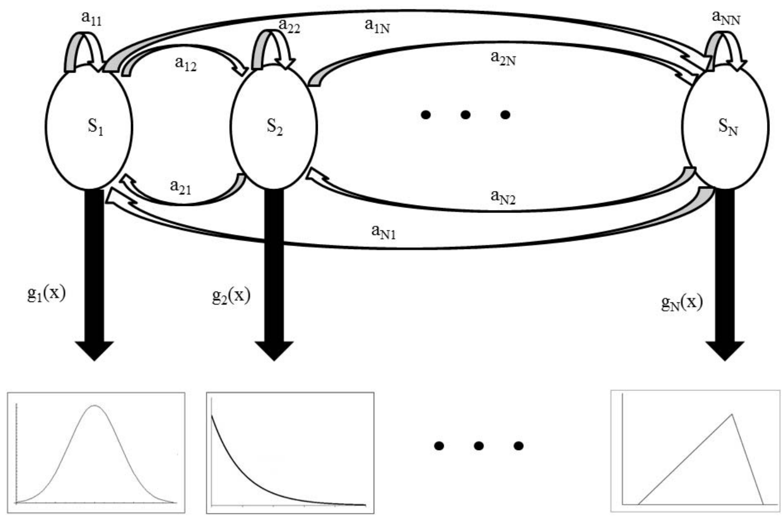



Figure 1: A continuous emission hidden Markov source with $N$ states.

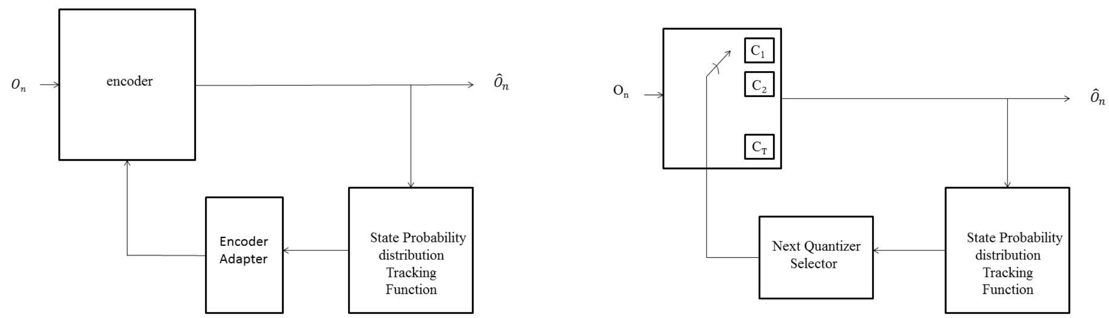
Fig. 1 depicts an example of a continuous hidden Markov source with $N$ states. In hidden Markov sources we have no access to the state of the source, only to the emitted observation, $O_n$

Our objective is to derive the optimal coding approach for the HMS.

## 3 Proposed Method

One important property of the HMS is that the state at time $t-1$, captures all past information relevant to the emission of the next source symbol. Specifically $P[O_t = v_k | q_{t-1} = S_j, O_1^{t-1}] = P[O_t = v_k | q_{t-1} = S_j]$ which implies that all observations until time $t-1$ provide no additional information on the next source symbol, beyond what is obtained by the state of the Markov chain at time $t-1$. Note further that we can not know with certainty the state of the HMS. Based on this fact the fundamental paradigm for achieving optimal coding is to track the state probability distribution of the HMS. In this approach, each output of the encoder (quantized observation) is sent into a unit called the *state probability distribution tracking function*. This unit estimates the state probability distribution, i.e., probabilities of the Markov chain being in each of the states, denoted by $\underline{\hat{p}}$. These probabilities are used in the *encoder adapter unit* to redesign the encoder optimally for the next input sample. Fig. 2(a) shows the fundamental approach for the optimal HMS encoder.

This general framework requires the encoder be redesigned for each observed sample, which entails high complexity. Therefore, to propose a practical approximation, we restrict the encoder adapter unit to select a codebook from a pre-designed set of codebooks. To facilitate this operation, we introduce a module called the 'next quantizer selector', which decides the codebook to be used for the next sample, based on




(a) Fundamental approach to optimal encoding of a hidden Markov source.

(b) Proposed practical encoder scheme.

Figure 2: Our encoder in fundamental idea and proposed method.

the output of the state probability distribution tracking function. Fig. 2(b) shows the proposed practical encoding scheme. Since the state probability distribution tracking unit uses the output of the quantizer to track states, the decoder can use the exact same procedure for computing $\underline{\hat{p}}$, and employ the next quantizer selector to determine the codebook to be used.

The main difference between the proposed scheme and a conventional finite state quantizer is in how the next quantizer is selected. The finite state quantizer makes the selection based on the current quantizer codebook and its output. However, in the proposed method, next quantizer codebook is determined by the output of the state probability distribution tracking unit, $\underline{\hat{p}}$, which captures all relevant information from the entire history of observations.

Our objective is to design the codebooks $c_1, c_2, \cdots, c_T$, the state probability distribution tracking function and the next-quantizer selector, given a training set of samples, so as to minimize the average reconstruction distortion at a given encoding rate. We first consider open loop design to derive the main concepts, and then proceed to necessary closed loop approach.

### *3.1 Open Loop Design*

The first step is to process the training set and estimate the HMS parameters. Then, we follow standard procedure in HMM analysis and define forward variable at time $t-1$ as

$$\alpha_{t-1}(i) = P(O_1^{t-1}, q_{t-1} = S_i), \tag{1}$$

i.e., the joint probability of emitting the observed source sequence $O_1^{t-1}$ up to time $t-1$ and that the state at time $t-1$ is $S_i$. The forward variables can be computed recursively:

1. $\alpha_1(i) = \pi_i g_i(O_1)$ for $1 \le i \le N$
2. $\alpha_{k+1}(j) = [\sum_{i=1}^{N} \alpha_k(i) a_{ij}] g_j(O_{k+1})$

Next, applying Bayes rule yields

$$\begin{aligned} P[q_{t-1} = S_i | O_1^{t-1}] &= \frac{P(q_{t-1} = S_i, O_1^{t-1})}{P(O_1^{t-1})} \\ &= \frac{P(q_{t-1} = S_i, O_1^{t-1})}{\sum_{j=1}^{N} P(q_{t-1} = S_j, O_1^{t-1})} \\ &= \frac{\alpha_{t-1}(i)}{\sum_{j=1}^{N} \alpha_{t-1}(j)} \end{aligned} \tag{2}$$

Let's define the vector $\underline{p}^{(t)}$ as

$$\begin{aligned} \underline{p}^{(t)} &= \{p_1, \ldots, p_{N-1}\} \\ &= \{P[q_t = S_1 | O_1^{t-1}], \cdots, P[q_t = S_{N-1} | O_1^{t-1}]\}. \end{aligned} \tag{3}$$

where, $p_i$ is the probability of being in state $i$ at time $t$, given the observation sequence up to time $t-1$ :

$$\begin{aligned} p_i &= P[q_t = S_i | O_1^{t-1}] \\ &= \sum_{j=1}^{N} P[q_{t-1} = S_j | O_1^{t-1}] a_{ij} \end{aligned} \tag{4}$$

After finding $\underline{p}^{(t)}$ for all $t$, we can divide training set to $T$ subgroups based on $\underline{p}$. In other words we vector quantize $\underline{p}$ into $T$ cells.

Next, using Lloyd's (or other) algorithm we find the best codebook and partition for the symbols in each of the training subgroups. If we allocate $R$ bits for quantizing, we have $T \times 2^R$ codewords.

$$C_n = \{x_{n1}, \ldots, x_{n(2^R)}\} \quad \text{for} \quad n = 1, 2, \ldots, T$$

In this system, the state probability distribution tracking unit finds the probability of being in each of the states i.e. $\underline{p}^{(t)}$, and the next quantizer selector decides which codebook to use next. Finally, the encoder quantizes the next symbol.

### *3.2 Closed Loop Design*

The shortcoming of the open loop system is that the decoder can not exactly mimic the encoder. To avoid this problem the state probability distribution tracking unit must track the states probability distribution, based on the output of the encoder, not based on original source symbols. In the closed loop system for estimating $\underline{p}$, instead of using source symbols we use outputs $x_{ni}$, $i = 1, 2, \ldots, 2^R$, of the quantizers, $n = 1, 2, \ldots, T$. In effect, for finding the probability to be in each of the states, we are building a new HMS based on new observation symbols, i.e. quantized values $x_{ni}$. It's obvious that the state transition probabilities and initial state distribution are the same as for the original source. So we only need to find new emission probabilities for the new observation symbols ( $x_{ni}$).

For estimating $b_j(c_{ni}) = P[O_t = c_{ni}|q_t = S_j]$, we first find the training set symbols which are in subgroup $n$ and are quantized to $c_{ni}$, then we add $p_j$ for those elements. Finally we normalize these probabilities in order to get $b_j(x_{ni})$ for $i = 1, \ldots, 2^R$, $n = 1, \ldots, T$ and $j = 1, \ldots, N-1$.

Now, to improve the codebook, state probability distribution tracking unit and next quantizer selector we do as follows:

For the first symbol of the training set we use the last quantizer, then for the other symbols state probability distribution tracking unit finds $\underline{\hat{p}}$ as

$$\begin{aligned}\underline{\hat{p}}^{(t)} &= \{\hat{p}_1, \ldots, \hat{p}_{N-1}\} \\ &= \{P[q_t = S_1|\hat{O}_1^{t-1}], \ldots, P[q_t = S_{N-1}|\hat{O}_1^{t-1}]\}\end{aligned} \tag{5}$$

where $\hat{O} \in c_{ni}$. Then the next quantizer selector chooses the next quantizer based on the $\underline{\hat{p}}$. We do this until the end of the training set. Now we have a new set of $\underline{\hat{p}}$, utilizing which we can update next quantizer selector parameters and divide the training set to $T$ subgroups and finally designing the best codebook for them. Having new codebooks we can update our new HMS parameters and $\underline{\hat{p}}$ and design the codebook from the beginning. We can repeat this iteration until there is no improvement in the performance of the system.

Before concluding this section it is important to make the following comments:

Firstly, for updating new HMM parameters we used $\underline{p}$ which is calculated only using symbols until time $t-1$. However in the design phase we have access to the entire training set whose size we denote by $F$, and we can find $\underline{p}$ using $O_1^F$ instead of $O_1^{t-1}$. This is efficiently done by employing the Backward variables, defined as follows:

$$\beta_t(i) = P[O_{t+1}^F|q_t = S_i] \tag{6}$$

i.e., the probability of emitting symbols $O_{t+1}, \ldots, O_F$, given that the source was in state $i$ at time $t$. These backward variables can be updated recursively as follows [6]:

1. $\beta_F(i) = 1$, for $1 \leq i \leq N$
2. $\beta_t(i) = \sum_{j=1}^N a_{ij} b_j(O_{t+1}) \beta_{t+1}(j)$, for $1 \leq t \leq F-1, 1 \leq i \leq N$

Finally :

$$\begin{aligned}p_i &= P[q_t = S_i|O_1^F] \\ &= \frac{\alpha_t(i)\beta_t(i)}{P[O_1^F]} \\ &= \frac{\alpha_t(i)\beta_t(i)}{\sum_{i=1}^N \alpha_F(i)}\end{aligned} \tag{7}$$

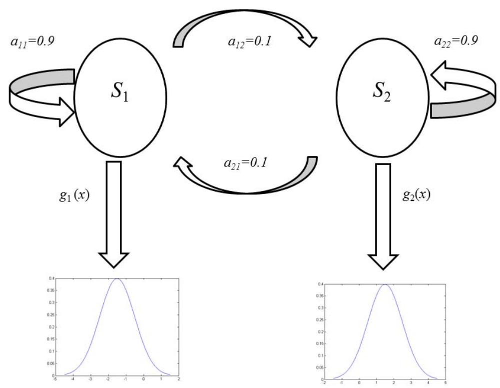



Figure 3: A Hidden Markov source with 2 states, as used in the simulations.

Note also that we use the reduced set $\underline{p} = \{p_1, \ldots, p_{N-1}\}$ for design since $p_N = 1 - \sum_{i=1}^{N-1} p_i$.

## 4 Experimental Results

Let us assume that the encoder arbitrarily uses the last quantizer when it processes the first source symbol.(Any convention can be used as long as it is known to the decoder. Alternatively we can send an initial side information to determine which quantizer is used.) For the rest of the symbols, the encoder first finds $\underline{\hat{p}}$, based on which it selects the suitable quantizer. The exact same procedure is employed by the decoder.

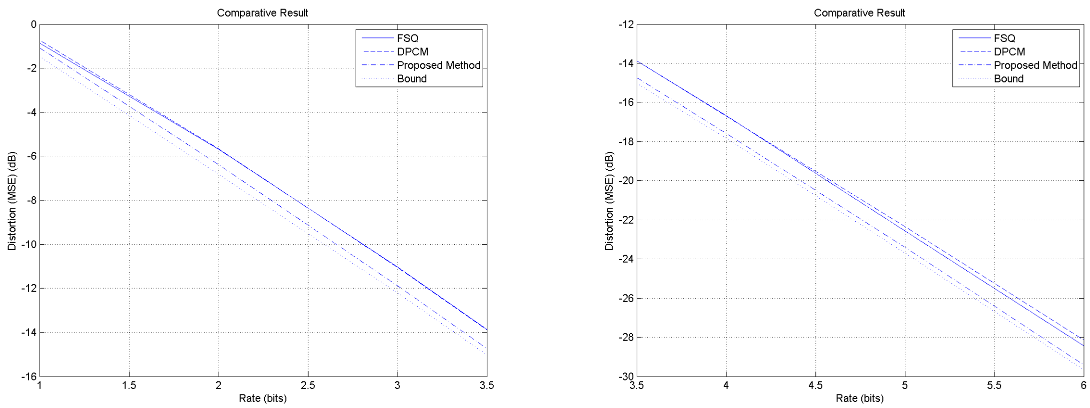
It should be noted that finding $\underline{\hat{p}}$ does not impose a suboptimal computational burden on the encoder and decoder. They simply update the forward variables (as discussed earlier) and obtain $\underline{\hat{p}}$.

A lower bound [5] on the distortion, when using a switched scalar quantizer is:

$$\hat{D} = \sum_{m=1}^{N} \rho_m \{\min_{Q} \sum_{j=1}^{N} a_{mj} E_j (x - Q(x))^2\} \tag{8}$$

where $E_j$ is expected value using $j$th subsource statistics, $\rho_m$ is the stationary probability of being in state $m$, and minimization is over all possible quantizers. Note that this bound is very optimistic, as it pretends that the Markov chain state at time $t-1$ is known exactly for selecting next quantizer.

In the first experiment the HMS has two Gaussian subsources, one of them having mean $\mu_1 = -1.5$ and variance $\sigma_1^2 = 1$ the other one having mean $\mu_2 = +1.5$ and variance $\sigma_2^2 = 1$. Moreover $a_{11} = a_{22} = 0.9$, as depicted in Fig. 3. For the simulation we set $T = 5$ (i.e., the number of codebooks are 5) and the coding rate varies from 1 to 6 bits per symbol. We compare the proposed method with finite state quantizer,as well as DPCM.




(a) Rate 1 to 3 bits per symbol

(b) Rate 4 to 6 bits per symbol

Figure 4: Rate-Distortion graphs for the proposed method, finite state quantizer (FSQ), DPCM, and the lower bound at bit rates ranging from 1 to 6.

The result is shown in Fig. 4. As is evident, the proposed method offers gains of approximately 1 dB over the finite state quantizer and DPCM. Comparison to the theoretical lower bound shows that the method leaves little room for improvement and performs very close to the bound.

Note again that this distortion bound makes the optimistic assumption that the state at time $t-1$ is known and used to obtain the best encoder for time $t$. Obviously, a real system does do not have access to source state at any time. So it is not an achievable lower bound, yet the proposed system performs fairly close to this bound.

The second experiment involves the same source except for varying the transition probability $a_{11} = a_{22}$ from 0.9 to 0.99 with coding rate of 4 bits per symbol.

The method shows consistent gains over its competitors with somewhat larger gains at high values of $a_{11}$.

## 5 Conclusion

A new fundamental source coding approach for hidden Markov sources is proposed, based on tracking the state probability distribution, and is shown to be optimal. Practical encoder and decoder schemes that leverage the main concepts are introduced, which consist of three modules: a switchable set of quantizer codebooks, a state probability distribution tracking function, and a next quantizer selector. An iterative approach is developed for optimizing the system given a training set.

The state probability distribution tracking module estimates, at each time instance, the probability that the source is in each possible state, given the quantized observations, and the next quantizer selector determines the codebook to be used, given the state probability distribution it receives at its input. The decoder mimics the encoder and switches to the same codebook. The switching decisions are made while optimally accounting for all available information from the past, and are ob-

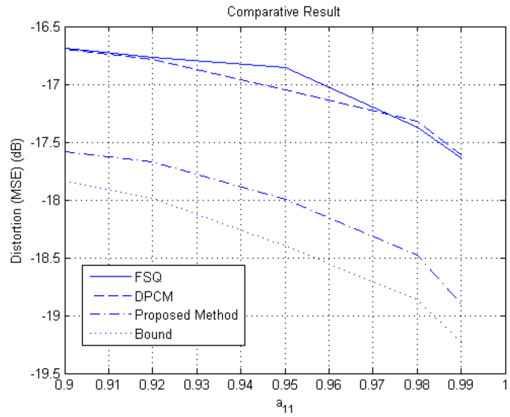



Figure 5: Rate-Distortion graph for proposed method, finite state quantizer (FSQ), DPCM, and the lower bound for bit rate 4 bits per symbol for transition probability $a_{11} = a_{22}$ from 0.9 to 0.99 .

tained efficiently by recursive computation of HMM forward variables. Simulation results confirm that the approach significantly reduces the MSE distortion relative to existing techniques. We note that these gains were obtained despite some suboptimal simplifications of the implementation, e.g., dividing the training set into equal size subsets.

We also left open the question of how large one may make $T$ (the number of codebooks) to enhance the quantizer adaptivity and improve the gains. The tradeoff is due to the increase in encoder and decoder storage complexity. This issue is exacerbated in the case of HMS with many states. An approach to solving this problem, currently under investigation, leverages the concept of the universal codebook [7], which enables increasing the number of vector quantizer codebooks in conjunction with direct control of storage complexity.

## 6 ACKNOWLEDGMENTS

The work was supported in part by the National Science Foundation under grants CCF-1016861 and CCF-1320599.

## References

[1] L. R. Rabiner, S. E. Levinson, and M. M. Sondhi, “On the application of vector quantization and hidden markov models to speaker-independent, isolated word recognition,” *Bell System Technical Journal*, vol. 62, no. 4, pp. 1075–1105, 1983.

[2] A. Gersho and R. Gray, *Vector quantization and signal compression*, Springer, 1992.

[3] J. Foster, R. Gray, and M. Dunham, “Finite-state vector quantization for waveform coding,” *Information Theory, IEEE Transactions on*, vol. 31, no. 3, pp. 348–359, 1985.

[4] M. Dunham and R. Gray, “An algorithm for the design of labeled-transition finite-state vector quantizers,” *Communications, IEEE Transactions on*, vol. 33, no. 1, pp. 83–89, 1985.

[5] D.M. Goblirsch and N. Farvardin, “Switched scalar quantizers for hidden markov sources,” *Information Theory,IEEE Transactions on*, vol. 38, no. 5, pp. 1455–1473, September 1992.

[6] L.R. Rabiner, “A tutorial on hidden markov models and selected applications in speech recognition,” *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257–286, February 1989.

[7] S. Ramakrishnan, K. Rose, and A. Gersho, “Constrained-storage vector quantization with a universal codebook,” *Image Processing,IEEE Transactions on*, vol. 7, no. 6, pp. 785–793, June 1998.