DESIGN OF CHANNEL OPTIMIZED VECTOR QUANTIZERS IN THE PRESENCE OF CHANNEL MISMATCH

Hamid Jafarkhani

AT&T Labs - Research
Red Bank, NJ
email: hamid@research.att.com

Nariman Farvardin

Department of Electrical Engineering and
Institute for Systems Research
University of Maryland, College Park, MD
email: farvar@isr.umd.edu

ABSTRACT

We propose algorithms to design channel-optimized vector quantizers in the presence of channel mismatch. We consider two cases: (i) no information about the statistics of the channel bit error rate is available and (ii) the probability density function of the channel bit error rate is known. We also consider the use of an estimate of the channel signal-to-noise ratio to improve performance. Simulation results demonstrate the advantages of new design algorithms.

1. INTRODUCTION

Source coding applications which involve transmission over noisy channels have been the main motivation for studying the sensitivity of a vector quantizer (VQ) to channel noise. These studies have led to the development of techniques for making a VQ robust with respect to channel noise, either by an appropriate binary codeword assignment [1, 2] or by a complete redesign of the VQ partition and codebook, resulting in the so-called channel-optimized vector quantizer (COVQ) [3, 4].

Previous studies on the subject have concentrated on the design of a COVQ for memoryless binary symmetric channels (BSCs) [3, 4] and finite-state channels (FSCs) [5, 6, 7]. In both cases, the exact knowledge of the characteristics of the channel is needed to design the quantizer. In many practical situations, e.g. wireless communication applications, the characteristics of the channel are not known. In fact, the behavior of a fading channel is time-varying. To apply the idea of a COVQ to slow fading channels, one may model the channel by an FSC [8, 9, 10] and design a COVQ for the resulting FSC [7]. The basic idea is to compute the average bit error rate (BER) for a given statistics of the channel (or a state of the channel) and apply the COVQ design procedure to a hypothetical BSC with the calculated BER. The rational behind this approach is that given a modulation scheme, e.g. uncoded binary phase-shift keying (BPSK), and received channel signal-to-noise ratio (CSNR), the channel behaves like a BSC. So, the BSC model is valid at each time and it is reasonable to assume that the BSC model works for the overall channel as long as the variations are small (the channel is in the same "state").

In this paper, we derive the optimal solution of the problem and show that the aforementioned approach is not optimal. Then, we demonstrate, through simulation, how much gain can be achieved by adopting our new scheme. The organization of the paper is as follows. Section 2 defines the notation and formulates the problem. Section 3 proposes the minimization of a worst-case scenario when the probability density function (p.d.f.) of BERs is not known. Section 4 assumes a p.d.f. for BERs and provides an optimal solution to minimize the average distortion. In Section 5, the estimation of the received CSNR is used to improve the overall performance of the system. A low complexity, table-lookup implementation is proposed in Section 6. Simulation results are presented in Section 7.

2. PROBLEM FORMULATION

Let us assume that the behavior of the channel for each channel use is like that of a BSC. For an uncoded BPSK signaling system, the BER is given by

$$BER = \frac{1}{2}\text{erfc}(\sqrt{\gamma_b}), \tag{1}$$

where γ_b is the received CSNR per bit and erfc is the error function. The actual BER, ϵ_a, depends on the received CSNR which is time-varying for a fading channel. For non-binary modulation schemes, the channel behaves like a discrete memoryless channel (DMC) at each time instance. All algorithms presented in this paper can be generalized to the

Prepared through collaborative participation in the Advanced Telecommunications/Information Distribution Research Program (ATRIP) Consortium sponsored by the ARL under Cooperative Agreement DAAL01-96-2-0002. Hamid Jafarkhani was with the University of Maryland.

case of DMC (instead of BSC); however, we do not discuss DMCs for the sake of brevity. Let us assume that the set of all possible BERs is I ($\epsilon_a \in I$). I is usually a subset in [0,0.5]. If the transmitter (encoder) and receiver (decoder) both know the exact value of ϵ_a for each channel use, a COVQ can be designed for every possible value of $\epsilon_a \in I$ and the transmitter and receiver can adapt themselves by switching to the optimal COVQ encoder and decoder. Such a scheme is implementable for FSCs [7] due to the fact that $|I|$ (cardinality of I) is finite. If I is a continuous set ($|I|=\infty$), the required memory of such a system is infinite and therefore it is impractical. So, usually there is a mismatch between the characteristics of the channel and the assumptions for which the quantizer has been designed. A simple example of such a channel mismatch corresponds to a quantizer designed for a noiseless channel and applied to a BSC or an FSC. Simulation shows that the performance degradation due to mismatch is significant for very noisy channels. A more general example is a quantizer designed for a BSC with BER equal to ϵ_d and applied to a BSC with BER equal to ϵ_a (previous example is a special case when $\epsilon_d = 0$). As we discussed earlier such a mismatch is unavoidable in many practical scenarios. In the following sections, we present optimal design strategies for COVQs in the presence of channel mismatch.

3. BEST STRATEGY WITH NO KNOWLEDGE ABOUT THE STATISTICS OF THE BER

In this section, we assume that the channel is a BSC and we do not have any information about the statistics of the channel BER (not even its p.d.f.). Let us define $D(\epsilon_d, \epsilon_a)$ as the average end-to-end distortion when we transmit the codewords of a COVQ designed for a BSC with BER equal to ϵ_d over a BSC with BER equal to ϵ_a. Then for a given ϵ_d, the worst performance is calculated by

$$\sup_{\epsilon_a \in I} D(\epsilon_d, \epsilon_a). \tag{2}$$

One design objective in this case is to find ϵ_d^* which provides

$$D^* = \inf_{\epsilon_d} \sup_{\epsilon_a \in I} D(\epsilon_d, \epsilon_a). \tag{3}$$

It can be easily shown that

$$D^* = \inf_{\epsilon_d} \sup_{\epsilon_a \in I} D(\epsilon_d, \epsilon_a) = \inf_{\epsilon_d} D(\epsilon_d, \sup_{\epsilon_a \in I} \epsilon_a). \tag{4}$$

And, by the definition of COVQ:

$$\epsilon_d^* = \sup_{\epsilon_a \in I} \epsilon_a. \tag{5}$$

4. BEST STRATEGY WITH KNOWN P.D.F. OF THE BER

In this section, we assume that the p.d.f. of the BER, denoted by $f_\epsilon(.)$, is known apriori. In other words,

$$p(A < \epsilon_a \leq B) = \int_A^B f_\epsilon(x) dx. \tag{6}$$

The objective is to minimize the average distortion

$$\overline{D} = \int_I D(\epsilon_a) f_\epsilon(\epsilon_a) d\epsilon_a, \tag{7}$$

where $D(\epsilon_a)$ is the end-to-end distortion when BER=ϵ_a. The optimization is done for a given source, a fixed dimension m and a fixed codebook size M. Let us denote the i^{th} encoding cell and codevector by S_i and $\mathbf{c}_i$, respectively. Distortion $D(\epsilon_a)$ in Equation (7) can be calculated by

$$D(\epsilon_a) = \sum_{i=0}^{M-1} \sum_{j=0}^{M-1} P_{\epsilon_a}(j|i) \int_{S_i} p_{\mathbf{x}}(\mathbf{x}) \parallel \mathbf{x} - \mathbf{c}_j \parallel^2 d\mathbf{x}, \tag{8}$$

where $P_{\epsilon_a}(j|i)$ denotes the probability that j is received given that i is transmitted over a BSC with BER=ϵ_a, $p_{\mathbf{x}}(\mathbf{x})$ is the m-fold p.d.f. of the source and $P_{\epsilon_a}(j|i) = \epsilon_a^{h(i,j)}(1-\epsilon_a)^{log_2 M - h(i,j)}$, where $h(i,j)$ is the Hamming distance between i and j, $i, j \in \{0, 1, \cdots, M-1\}$. Then, substituting Equation (8) in Equation (7) results in

$$\overline{D} = \int_I \sum_{i,j} P_{\epsilon_a}(j|i) \int_{S_i} p_{\mathbf{x}}(\mathbf{x}) \parallel \mathbf{x} - \mathbf{c}_j \parallel^2 d\mathbf{x} f_\epsilon(\epsilon_a) d\epsilon_a, \tag{9}$$

$$\overline{D} = \sum_{i,j} \int_I P_{\epsilon_a}(j|i) f_\epsilon(\epsilon_a) d\epsilon_a \int_{S_i} p_{\mathbf{x}}(\mathbf{x}) \parallel \mathbf{x} - \mathbf{c}_j \parallel^2 d\mathbf{x}, \tag{10}$$

$$\overline{D} = \sum_{i=0}^{M-1} \sum_{j=0}^{M-1} \overline{P}(j|i) \int_{S_i} p_{\mathbf{x}}(\mathbf{x}) \parallel \mathbf{x} - \mathbf{c}_j \parallel^2 d\mathbf{x}, \tag{11}$$

where

$$\overline{P}(j|i) = \int_I P_{\epsilon_a}(j|i) f_\epsilon(\epsilon_a) d\epsilon_a. \tag{12}$$

For the squared error distortion measure, the design algorithm is a modified generalized Lloyd algorithm [11], with the following expressions for the optimal partition, $\mathcal{P} = \{S_0, S_1, \ldots, S_{M-1}\}$, and the optimal codebook, $\mathcal{C} = \{\mathbf{c}_0, \mathbf{c}_1, \ldots, \mathbf{c}_{M-1}\}$:

$$S_i = \{\mathbf{x} : \sum_{j=0}^{M-1} \overline{P}(j|i) \parallel \mathbf{x} - \mathbf{c}_j \parallel^2 \leq \sum_{j=0}^{M-1} \overline{P}(j|l) \parallel \mathbf{x} - \mathbf{c}_j \parallel^2, \ \forall l\}, \tag{13}$$

and

$$\mathbf{c}_j = \frac{\sum_{i=0}^{M-1} \overline{P}(j|i) \int_{S_i} \mathbf{x} p_{\mathbf{x}}(\mathbf{x}) d\mathbf{x}}{\sum_{i=0}^{M-1} \overline{P}(j|i) \int_{S_i} p_{\mathbf{x}}(\mathbf{x}) d\mathbf{x}}. \quad (14)$$

The design algorithm is based on an iterative application of Equations (13) and (14).

Equation (11) shows that designing a COVQ for $\epsilon_d = \overline{\epsilon_a} = \int_I \epsilon_a f_\epsilon(\epsilon_a) d\epsilon_a$ is not optimal. The performance improvement depends on the encoding rate, the level of noise in the channel and the statistics of the source. We will have more to say on this in following sections.

5. BEST STRATEGY WHEN WE HAVE AN ESTIMATE OF BER AT THE DECODER

Usually in practice, the decoder can estimate the value of ϵ_a (using the received CSNR). This information can be used to improve the performance of the quantizer. We use a single encoder and K sets of codevectors at the decoder. At the decoder, at each time instance, we use $\mathbf{c}_{j,k}$ to reconstruct the vector when the received index is j and $\epsilon_a \in I_k,\ k = 1, \ldots, K$. The average distortion can be calculated by

$$\overline{D} = \sum_{k=1}^{K} \int_{I_k} \sum_{i=0}^{M-1} \sum_{j=0}^{M-1} P_{\epsilon_a}(j|i) \int_{S_i} p_{\mathbf{x}}(\mathbf{x}) \parallel \mathbf{x} - \mathbf{c}_{j,k} \parallel^2 d\mathbf{x} f_\epsilon(\epsilon_a) d\epsilon_a, \quad (15)$$

$$\overline{D} = \sum_{k=1}^{K} \sum_{i=0}^{M-1} \sum_{j=0}^{M-1} \overline{P}_k(j|i) \int_{S_i} p_{\mathbf{x}}(\mathbf{x}) \parallel \mathbf{x} - \mathbf{c}_{j,k} \parallel^2 d\mathbf{x}, \quad (16)$$

where

$$\overline{P}_k(j|i) = \int_{I_k} P_{\epsilon_a}(j|i) f_\epsilon(\epsilon_a) d\epsilon_a. \quad (17)$$

Now, the design components are: (i) the encoder, (ii) the decoder and (iii) the intervals I_k. (Defining $I_k = (T_{k-1}, T_k]$ ($T_0 \le T_1 \le \ldots \le T_{K-1} \le T_K$), we have $K-1$ parameters to find.) The design algorithm iterates between the following steps:

- Given the intervals, use a generalized Lloyd algorithm to design the encoder and the decoder. This is accomplished by iterating between
  – *Generalized centroid:*
  When the encoding regions are fixed, we choose the best codevector, $\mathbf{c}_{j,k}$, for each received index, j, by

$$\mathbf{c}_{j,k} = \frac{\sum_{i=0}^{M-1} \overline{P}_k(j|i) \int_{S_i} \mathbf{x} p_{\mathbf{X}}(\mathbf{x}) d\mathbf{x}}{\sum_{i=0}^{M-1} \overline{P}_k(j|i) \int_{S_i} p_{\mathbf{X}}(\mathbf{x}) d\mathbf{x}}. \quad (18)$$

  – *Generalized nearest neighbor:*
  For each input vector, $\mathbf{x}$, we select the optimal encoding region corresponding to codeword i that minimizes the distortion for a fixed set of codevectors by

$$S_i = \{\mathbf{x} : \sum_{k=1}^{K} \sum_{j=0}^{M-1} \overline{P}_k(j|i) \parallel \mathbf{x} - \mathbf{c}_{j,k} \parallel^2 \le \sum_{k=1}^{K} \sum_{j=0}^{M-1} \overline{P}_k(j|l) \parallel \mathbf{x} - \mathbf{c}_{j,k} \parallel^2; \forall l\}. \quad (19)$$

- Given the encoder and the decoder, change the thresholds in a direction which reduces the distortion.

6. TABLE-LOOKUP IMPLEMENTATION OF THE ENCODER

The encoder of the quantizers proposed in Sections 4 and 5 can be implemented using table-lookups. The encoder of a channel-matched hierarchical table-lookup vector quantizer (CM-HTVQ) with a compression ratio (CR) of $2^N : 1$ consists of N stages. To design the first $(N-1)$ stages, first $(N-1)$ VQs with dimensions 2^n, $n = 1, \ldots, N-1$ are designed. The 2^n-dimensional VQ is designed with a CR $2^n : 1$. The first table is constructed by considering each pair of input samples and storing the codeword of the nearest codevector of the 2-dimensional VQ as the output of the table. The lookup table of the n^{th} stage is designed by considering all possible input pairs (outputs of the $(n-1)^{th}$ stage) and storing the corresponding codeword of the 2^n-dimensional VQ encoder in the table. The last stage, which describes the actual encoder partition, must be adapted to the characteristics of the channel and is designed based on the same principles using a COVQ (e.g. COVQs designed in Sections 4 and 5). For a detailed description of the design algorithm, the interested reader is referred to [7]. In [7] we show that CM-HTVQ simultaneously provides low encoding complexity (only lookup tables) and robustness against transmission noise for BSCs and FSCs. In next section, we examine the CM-HTVQ structure against channel mismatch.

7. SIMULATION RESULTS AND CONCLUSIONS

Simulation results for transmitting the 512×512 Lenna over a log-normal fading channel with an average channel signal to noise ratio of 10 dB is shown in Table 1. The quantizers are designed using a training sequence which consists of five images from the USC database (Couple, Crowd, Man, Woman1 and Woman2) and the peak signal-to-noise ratio (PSNR)[1] results are obtained by averaging the mean squared error over 10 runs of channel simulation. Simulated

[1] PSNR=$10 \log_{10}[255^2 / \frac{1}{N} \sum_{i=1}^{N} (x_i - \hat{x}_i)^2]$

Table 1: PSNR Results for a log-normal fading channel (average CSNR=10 dB); 512×512 Lenna.

Rate (bpp)	Dim.	VQ	COVQ Average BER	Optimal NA	DA $K=2$
2	4	24.49	26.33	27.00	27.27
1	8	23.96	25.80	26.25	26.46
0.5	16	23.54	25.14	25.59	25.84

Table 2: PSNR Results for a log-normal fading channel (average CSNR=10 dB); 512×512 Lenna.

Rate (bpp)	Dim.	HTVQ	CM-HTVQ Average BER	Optimal NA	DA $K=2$
2	4	24.46	26.28	26.94	27.20
1	8	23.88	25.69	26.11	26.31
0.5	16	23.45	24.96	25.38	25.60

annealing is used to assign binary code words [1] to the code vectors of VQ, thus making the comparison between different schemes meaningful. Note that f_ϵ and I are derived for the log-normal fading channel by assuming a BPSK modulation scheme and the corresponding BER given by Equation (1). This is straightforward due to the fact that BER as a function of received CSNR (γ_b) is a one-to-one decreasing map from $I' = [0, \infty)$ to $I = (0, 0.5]$. In Table 1, optimal NA (non-adaptive) refers to the algorithm of Section 4 and DA (decoder-adaptive) refers to the algorithm of Section 5. As expected, the performance of the optimal non-adaptive system is better than that of the COVQ designed for an average BER. The decoder-adaptive system with only two intervals outperforms the optimal non-adaptive system. Of course using more intervals results in improving the performance at the cost of increasing the required memory at the decoder. In all cases, channel-matched quantizers outperform quantizers designed for a noiseless channel by a large margin. Similar trends are observed for CM-HTVQs in Table 2. As can be seen in Table 2, CM-HTVQ results are very close to COVQ results while maintaining a low encoding complexity.

One important question is the sensitivity of the decoder-adaptive quantizer to the CSNR estimate. To study the effects of imperfect CSNR estimation, we repeat the simulations of the last column of Table 1 with 20% overestimate or underestimate in the amount of CSNR and report the results in Table 3. Simulation results show that the decoder-adaptive system is not sensitive to the CSNR estimation. This is due to the fact that 20% error in CSNR estimation, for the above example, only affects 2.5% of the probability mass. In other words, the boarder region contains a small probability, thus a minor effect on the overall performance.

In this paper, we have proposed a single solution for two different problems. One problem is the case of a BSC which its BER is unknown and a fixed channel mismatch exists. The second problem is a time-varying BSC. Although mathematically the two problems are similar – therefore the proposed unified solutions – we feel that different design criteria are appropriate for these problems. In particular, we believe that even if the p.d.f. of the BER is available for a BSC with unknown (but fixed) BER, a worst-case design criterion is more appropriate than an average distortion criterion. This is an interesting issue for future investigation.

Table 3: Effects of imperfect CSNR estimation on the performance of the DA-COVQ in Table 1.

Rate	Dim.	20% overestimate	20% underestimate
2	4	27.26	27.27
1	8	26.44	26.45
0.5	16	25.82	25.84

8. REFERENCES

[1] N. Farvardin, "A Study of Vector Quantization for Noisy Channels," *IEEE Trans. Inform. Theory*, vol. 36, pp. 799–809, July 1990.

[2] K. Zeger and A. Gersho, "Pseudo-Gray Coding," *IEEE Trans. Commun.*, vol. 38, pp. 2147–2158, Dec. 1990.

[3] H. Kumazawa, M. Kasahara, and T. Namekawa, "A Construction of Vector Quantizers for Noisy Channels," *Electron. Eng. Japan*, vol. 67-B, no. 4, pp. 39–47, 1984.

[4] N. Farvardin and V. Vaishampayan, "On the Performance and Complexity of Channel-Optimized Vector Quantizers," *IEEE Trans. Inform. Theory*, vol. 37, pp. 155–160, Jan. 1991.

[5] H. S. Wang and N. Moayeri, "Optimal Quantization for a Finite-State Noisy Channel," *Abstracts IEEE Int. Symp. Inform. Theory*, San Diego, CA, p. 79, Jan. 1990.

[6] T. M. Duman and M. Salehi, "Optimal Quantization for Finite-State Channels," *IEEE Trans. Inform. Theory*, vol. 43, pp. 758–765, Mar. 1997.

[7] H. Jafarkhani and N. Farvardin, "Channel–Matched Hierarchical Table–Lookup Vector Quantization for Finite–State Channels," submitted to *IEEE Trans. Inform. Theory*, Jan. 1997.

[8] E. N. Gilbert, "Capacity of a Burst-Noise Channel," *Bell Syst. Tech. J.*, vol. 39, pp. 1253–1265, Sept. 1960.

[9] E. O. Elliott, "Estimates of Error Rates for Codes on Burst-Noise Channels," *Bell Syst. Tech. J.*, vol. 42, pp. 1977–1997, Sept. 1963.

[10] H. S. Wang and N. Moayeri, "Finite-State Markov Channel – A Useful Model for Radio Communication Channels," *IEEE Trans. Vehicular Tech.*, vol. 44, pp. 163–171, Feb. 1995.

[11] Y. Linde, A. Buzo, and R. M. Gray, "An Algorithm for Vector Quantizer Design," *IEEE Trans. Commun.*, vol. COM-28, pp. 84–95, Jan. 1980.