VECTOR QUANTIZER DESIGN FOR MEMORYLESS NOISY CHANNELS

Kenneth A. Zeger and Allen Gersho

Communications Research Laboratory
Dept. of Electrical & Computer Engineering
University of California
Santa Barbara, CA 93106

ABSTRACT

To handle the effect of transmission errors on the performance of vector quantization (VQ) in source coding, a channel index assignment function can be incorporated into a source/channel model of VQ. Using this model, we obtain new conditions for the optimality of a vector quantizer for a given distortion measure which generalize the familiar centroid and nearest neighbor conditions. The optimal codevectors are linear combinations of those for the noiseless case, weighted by the a posteriori channel transition probabilities. The optimal encoder selects that codevector which minimizes a weighted sum of the distortions between the input and each codevector, where the weights are channel transition probabilities. An insightful derivation of the new conditions for a memoryless channel is given and an iterative design algorithm is described where at each step the average distortion monotonically decreases. Each iteration consists of three steps which separately modify the encoder, decoder, and the channel index assignment.

1. Introduction

Vector quantization (VQ), as a technique for analog waveform data compression has shown great promise for areas such as low rate speech and image coding (see for example [4],[5],[6], or [9]). Only recently has VQ emerged as a realizable technique in modern communications systems, such as in the 4.8 kbits/sec NASA Mobile Satellite Experiment [5]. Motivated by our work on this project, we have been concerned with a new and distinct error control problem associated specifically with the use of VQ in coding a speech waveform or an image. In VQ communications systems, codevector indices (binary words) must be transmitted over a noisy channel, often with a fairly high bit error rate, resulting in possibly incorrectly decoded codevectors. The resulting distortion adds to the degradation of reconstructed speech or image quality. In order to preserve low data rates, one would prefer to avoid the use of redundancy codes. What then can be done to improve the average distortion level of the system?

Prior work has considered the optimization of channel codewords to minimize excess distortion in signal coding due to bit errors ([2-3] and recently some specific consideration of channel errors for VQ has led to algorithms for the design of channel index permuations to reduce the effect of errors for a given vector quantizer design [1], [5], and [10]). By modifying the design procedure of vector quantizers for discrete memoryless channels, one can obtain improved performance. We generalize the notion of a quantizer to include the effects of channel noise as well as to incorporate the notion of an index assignment function, that carefully assigns channel indices to codevectors. Optimality conditions that generalize the usual centroid and nearest-neighbor conditions are derived. A three step noisy channel VQ design procedure similar in structure to the generalized Lloyd algorithm is then easily inferred.

2. Definitions

A *vector quantizer* Q is a mapping $Q : \mathbf{R}^p \rightarrow Y$ of p-dimensional Euclidean space $\mathbf{R}^p$ into a finite subset Y of $\mathbf{R}^p$, where $Y = \{\mathbf{y}_0, \mathbf{y}_1, \cdots, \mathbf{y}_{N-1}\}$ and $\mathbf{y}_i \in \mathbf{R}^p$ for $0 \le i \le N-1$. The ordered set Y is a *codebook* and the N elements of Y are called *codevectors.* The subscripts of the codevectors are the codevector *indices*, each index representing a channel word. Let the set of N vector indices be denoted by

$$I = \{0, 1, \cdots, N-1\}. \tag{1}$$

A vector quantizer Q that transmits vector indices across a noisy channel can be interpreted as the composition of four independent mappings involving a *coder* $C : \mathbf{R}^p \rightarrow I$, a *decoder* $D : I \rightarrow Y$, an index assignment function, and a random channel noise function. The decoder satisfies $D(k) = \mathbf{y}_k$ for $0 \le k \le N-1$. Corresponding to an encoder C is a partition $\{R_i\}$ of $\mathbf{R}^p$ where for each $i \in I$

$$R_i = C^{-1}(i) = \{x \in \mathbf{R}^p : C(x) = i\}. \tag{2}$$

Note that the encoder operation is fully specified by the partition. A *noisy channel vector quantizer* Q is defined as

$$Q = D \circ \pi^{-1} \circ \tau \circ \pi \circ C \tag{3}$$

where $\tau : I \rightarrow I$ is a memoryless noisy channel index mapping, and $\pi : I \rightarrow I$ is a one-to-one function that permutes the assignment of indices to codevectors. π^{-1} "unpermutes" the index assignment at the receiver end.

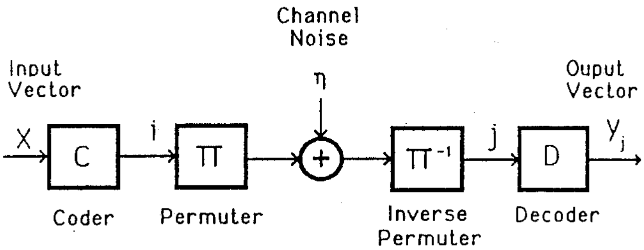

Fig. 1 A Noisy Channel Vector Quantizer

A noisy channel vector quantizer is completely specified by its partition $\{R_i\}$, output set Y, and index permutation function π. For a binary channel the noise function τ can be represented by

$$\tau(i) = i \oplus \eta \quad (i \in I) \tag{4}$$

where η is a random variable taking on values from the set I, and the operation $\oplus$ corresponds to the bitwise exclusive OR function. This is depicted in Fig. 1.

The goal in vector quantizer design is to find a quantizer that minimizes an expected distortion between an input vector and a decoded (or quantized) output vector. A distortion function d is assumed that takes as input two vectors $\mathbf{x}$ and $\mathbf{y}$ from $\mathbf{R}^p$ and produces a non-negative real value $d(\mathbf{x}, \mathbf{y})$, the distortion between $\mathbf{x}$ and $\mathbf{y}$. We wish to find a quantizer which minimizes the average distortion

$$e \equiv E[d(\mathbf{X}, Q(\mathbf{X}))]. \tag{5}$$

Minimizing e requires the specification of a codebook, a partition of $\mathbf{R}^p$, and a permutation function π that are jointly optimal. Because the global optimization of a quantizer appears to be a difficult task, we instead derive conditions for an optimal coder, decoder, or permutation, given that each of the other two is assumed fixed. This leads to a vector quantizer "design loop" in which an iterative algorithm that monotonically reduces the distortion can be implemented.

3. Optimal MSE Decoder

Let $\mathbf{x}$ be an input (random variable) vector to the quantizer. We assume a fixed encoder (partition), $\{R_i\}$, and a fixed permutation function π, and derive conditions for an optimal decoder. The distortion criterion considered in this development will be the mean-squared error distortion measure given by $d(\mathbf{x}, \mathbf{y}) = \|\mathbf{x}-\mathbf{y}\|^2$. The codevector $\mathbf{y}_j$ is decoded as the quantized value of $\mathbf{x}$ if some index $\pi(i)$ is transmitted and is received as $\pi(j)$. That is, $\mathbf{y}_j$ is selected if $\mathbf{x} \in R_i$ and $\tau(\pi(i)) = \pi(j)$ for some $i \in I$. With this in mind, we define for each $j \in I$ the *selector function* of the random vector $\mathbf{x}$ and the channel noise τ as

$$S_j(\mathbf{x}, \tau) = \begin{cases} 1 & \text{if } \bigvee_{i=0}^{N-1} [\mathbf{x} \in R_i \wedge \tau(\pi(i)) = \pi(j)] \\ 0 & \text{else} \end{cases} \tag{6}$$

where $\bigvee$ and $\wedge$ denote the boolean binary operators "OR" and "AND" respectively. The quantized value of $\mathbf{x}$ can be expressed as a linear combination of the vectors $\{\mathbf{y}_j\}$ by

$$Q(\mathbf{x}) = \sum_{j=0}^{N-1} \mathbf{y}_j A_j. \tag{7}$$

where in order to simplify notation we have let

$$A_j = S_j(\mathbf{x}, \tau). \tag{8}$$

The mean square error (MSE) of the system due to the combined effect of quantization and channel noise is given by $e = E\|\mathbf{x} - Q(\mathbf{x})\|^2$. We wish to minimize this quantity over all choices of $\mathbf{y}_i$. Since $Q(\mathbf{x})$ is a linear estimate of $\mathbf{x}$ with respect to the observables $\{A_j\}$, e is minimized using the orthogonality principle. This yields

$$E[(\mathbf{x} - Q(\mathbf{x}))A_j] = 0 \qquad 0 \le j \le N-1 \tag{9}$$

which implies that

$$E[\mathbf{x}A_j] = E[Q(\mathbf{x})A_j] \qquad 0 \le j \le N-1. \tag{10}$$

Since for each j, A_j can only take on values of 0 or 1, we have

$$E[\mathbf{x} \mid A_j = 1] \,\mathbf{Pr}[A_j = 1] = E[Q(\mathbf{x}) \mid A_j = 1] \,\mathbf{Pr}[A_j = 1]$$

so that

$$E[\mathbf{x} \mid A_j = 1] = E[Q(\mathbf{x}) \mid A_j = 1] = \mathbf{y}_j.$$

Using the definition of the random variable A_j we have

$$\mathbf{y}_j = E[\mathbf{x} \mid \bigvee_{i=0}^{N-1} (\mathbf{x} \in R_i \wedge \tau(\pi(i)) = \pi(j))] \tag{11}$$

where the expectation is conditioned over a union of N disjoint events. Expanding the conditional expectation above as a summation yields

$$\mathbf{y}_j = \frac{\sum_{i=0}^{N-1} E[\mathbf{x} \mid \mathbf{x} \in R_i] \, Pr[\mathbf{x} \in R_i \wedge \tau(\pi(i)) = \pi(j)]}{\sum_{i=0}^{N-1} Pr[\mathbf{x} \in R_i \wedge \tau(\pi(i)) = \pi(j)]} \tag{12}$$

where we have used the fact that $\mathbf{x}$ is independent of the channel noise. Regarding $\pi(i)$ as the transmitted index and $\pi(j)$ as the received index, denote the partition region probabilities by

$$P(\pi(i)) \equiv Pr[\mathbf{x} \in R_i] = Pr[\pi(i) \text{ sent}] \quad (13)$$

and the channel transition probabilities by

$$P(\pi(j) \mid \pi(i)) = Pr[\pi(j) \text{ received} \mid \pi(i) \text{ sent}]. \quad (14)$$

Further, let $P(\pi(i) \mid \pi(j))$ denote the a posteriori probability that index $\pi(i)$ was sent, given that index $\pi(j)$ was received. Then we see that

$$\mathbf{y}_j = \sum_{i=0}^{N-1} \frac{P(\pi(i))\, P(\pi(j) \mid \pi(i))\, E[\mathbf{x} \mid \mathbf{x} \in R_i]}{Pr[\pi(j) \text{ received}]} \quad (15)$$

Using Baye's Theorem we can write a simplified weighted centroid condition for the optimal codevectors in terms of the a posteriori channel transition probabilities.

Weighted Centroid Condition

$$\mathbf{y}_j = \sum_{i=0}^{N-1} E[\mathbf{x} \mid \mathbf{x} \in R_i]\, P(\pi(i) \mid \pi(j)). \quad (16)$$

The optimal codevectors for a noisy vector quantizer using the mean squared error criterion are linear combinations of the optimal codevectors in the noiseless case (i.e. the centroids of each partition region). Knowledge of the centroids and the channel transition probabilities completely specifies the choice of optimal codevectors. Thus, on a binary symmetric channel (BSC) with crossover probability ε, the optimal codevectors are chosen according to the rule

$$\mathbf{y}_j = \frac{1}{Pr[\pi(j) \text{ received}]} \cdot \sum_{i=0}^{N-1} E[\mathbf{x} \mid \mathbf{x} \in R_i]\, P(\pi(i))\, \varepsilon^{H(\pi(j),\pi(i))}\, (1-\varepsilon)^{b-H(\pi(j),\pi(i))} \quad (17)$$

where $b = \log_2 N$, assumed to be an integer, is the number of bits in a channel index, and H is the Hamming distance function between any two indices (the number of bit positions in which their binary representations differ).

As ε gets small, the likelihood of channel errors diminishes. The distortion in this case becomes increasingly due to the quantization error. In this limiting situation, since $\mathbf{y}_j$ is a continuous function of ε, it is easy to see that

$$\lim_{\varepsilon \to 0} \mathbf{y}_j = E[\mathbf{x} \mid \mathbf{x} \in R_j] \quad (18)$$

so that the optimal choice of the codevector $\{\mathbf{y}_j\}$ approaches the optimal codevector in the noiseless case as given by the usual centroid condition for optimal VQ.

4. Optimal Encoder

Given a set of codevectors $\{\mathbf{y}_i\}$ (decoder), and a fixed permutation function π, we want to find the optimal partition $\{R_i\}$ for a noisy channel vector quantizer. For the following optimality condition we no longer need to assume that the distortion measure d is the mean-squared criterion.

We seek the partition that minimizes the quantity

$$\begin{aligned} e &= E[d(\mathbf{x}, Q(\mathbf{x}))] \\ &= \sum_{i \in I} \sum_{j \in I} E[d(\mathbf{x}, Q(\mathbf{x})) \mid \pi(i) \text{ sent}, \pi(j) \text{ received}]\, P(\pi(i), \pi(j)) \\ &= \sum_{i \in I} \int_{R_i} \left[\sum_{j \in I} P(\pi(j) \mid \pi(i))\, d(\mathbf{x}, \mathbf{y}_j) \right] f_x(\mathbf{x})\, d\mathbf{x} \\ &\geq \sum_{i \in I} \int_{R_i} \min_{k \in I} \left[\sum_{j \in I} P(\pi(j) \mid p(k))\, d(\mathbf{x}, \mathbf{y}_j) \right] f_x(\mathbf{x})\, d\mathbf{x} \end{aligned} \quad (19)$$

Since the codevectors and the permutation function π are fixed, this lower bound can be attained by requiring for each $i \in I$,

Weighted Nearest Neighbor Condition

$$\left\{ \mathbf{x} \in \mathbf{R}^p : \sum_{j \in I} d(\mathbf{x}, \mathbf{y}_j)\, P(\pi(j) \mid \pi(i)) < \sum_{j \in I} d(\mathbf{x}, \mathbf{y}_j)\, P(\pi(j) \mid \pi(k)) \text{ for all } k \in I \right\} \subseteq R_i \quad (20)$$

We say that $\mathbf{y}_i$ is a *weighted nearest neighbor* of $\mathbf{x}$ if the quantity

$$\sum_{j \in I} d(\mathbf{x}, \mathbf{y}_j)\, P(\pi(j) \mid \pi(k)) \quad (21)$$

is minimized over $k \in I$ when $k = i$. The optimality condition requires that the partition be such that for each $i \in I$, R_i contains of all those vectors in $\mathbf{R}^p$ for which $\mathbf{y}_i$ is a weighted nearest neighbor. If we assume that the pdf of random variable $\mathbf{x}$ is such that the set of all $\mathbf{x}$ which have more than one weighted nearest neighbor (boundary points) has probability zero, then we can arbitrarily assign those $\mathbf{x}$ to any partition region R_i, without affecting the average distortion e.

It can be shown that the geometric structure of the optimal partition regions are convex polytopes as is known to be the case for a noiseless vector quantizer [6]. The proof is omitted here for brevity but will be presented in a future publication.

5. Optimal Index Assignment

Next, we consider the problem of selecting the best permutation function π for a fixed coder and decoder. Some locally optimal algorithms to achieve this task have been studied in [1], [10], and [5]. These techniques have shown significant improvements in the overall signal-to-noise ratio of various VQ systems. We indicate below a possible technique for finding an optimal permutation function given a training set of source data. There are at most a finite number of different permutation functions for a given index set I. The number of permutations of a set of N indices is $N!$. An exhaustive search technique for determining the best π on a BSC is indicated below. As before, suppose that $\mathbf{x} \in R_i$ and that $\pi(i)$ is transmitted and is received as $\pi(j)$. The choice of π that minimizes $e = E[d(\mathbf{x},\mathbf{y}_j)]$ can be determined in the following manner. Given a fixed codebook and a large training set $\mathbf{T}$ of vectors from $\mathbf{R}^p$, the VQ system can effectively be simulated, with channel error simulations governed by the quantity ε. The value of e can be approximated for large training sets by

$$e \approx \frac{1}{|\mathbf{T}|} \sum_{\mathbf{u} \in \mathbf{T}} d(\mathbf{u}, \mathbf{y}_j) \tag{22}$$

for each fixed codebook permutation π. For each $\mathbf{u} \in \mathbf{T}$, the partition region R_i to which it belongs can be easily determined by applying the weighted nearest neighbor optimality condition. The index $\pi(i)$ can be computed and transmission across a noisy channel simulated. The index j can then be deduced by applying π^{-1} to the received index. Hence the quantity $d(\mathbf{u}, \mathbf{y}_i)$ can be calculated for each $\mathbf{u}$ in the training set. If all $N!$ codebook permutations are run through the simulation, then the one with the smallest value of e is the optimal permutation.

This procedure, however, would require enormous computational complexity for even a relatively small number of codevector indices. In addition, the size of a training set $\mathbf{T}$ must be large enough to reasonably approximate the effects of channel errors on the codevector indices. It would be desirable to have any simulation take into account the effect of channel errors on each codevector index. The correct choice of an improved permutation function can never degrade quality. VQ design that does not consider such a permutation function in the design process, merely assumes that π is the identity function on I.

6. Quantizer Design Loop

We have shown that either the encoder, decoder, or permutation function can be chosen so as to improve the average quantizer distortion, when both of the other two are fixed. In fact one can never do worse by optimizing one of the three when the others are known. At worst, no distortion reduction could result. This suggests a three stage quantizer design loop, where at each stage only improvement can result. This design procedure essentially follows the spirit of the generalized Lloyd algorithm, and can easily be implemented using training sets.

The generalized Lloyd (or LBG [9]) design algorithm for vector quantization provides a convenient method for obtaining locally optimal VQ codebooks. At each iteration in the algorithm, a monotonic decrease in average distortion is achieved. The generalized Lloyd algorithm consists of a two step loop that redesigns a codebook for a fixed partition and repartitions the input space for a fixed codebook. The familiar centroid and nearest neighbor conditions provide rules for determining either an optimal codebook or partition when the other is fixed.

The basic structure of our algorithm is outlined below and in Fig. 2. It is assumed that a training set of input vectors is available and that the channel error probabilities are given.

Algorithm

Step 0: Pick an initial codebook and permutation function.

Step 1: Partition the training set optimally using the codevectors and the channel probabilities. Compute the centroids.

Step 2: Find an optimal permutation function for the current codebook and partition.

Step 3: Find the best codebook from the centroids and channel probabilities. Go to Step 1.

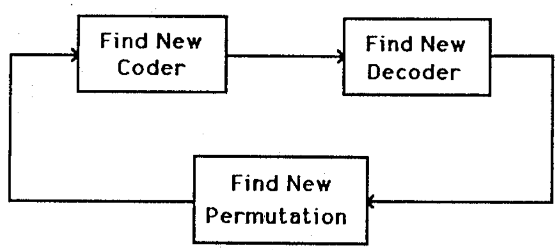

Fig. 2 Design Loop For Noisy Vector Quantizer Design.

7. Conclusions

With the knowledge that a VQ coding system will transmit data across a noisy channel, one can optimize the average performance of the system by properly designing the VQ coder and decoder. This can be achieved by adhering to the weighted nearest neighbor rule for partitioning the input space, and to the weighted centroid condition for choosing the values of the codevectors corresponding to each partition region. A further increase in performance can be achieved by properly assigning binary indices to codevectors. Together, these three design techniques are combined to form a three stage design loop which generalizes the usual LBG algorithm for vector quantizer design to the case where noisy channels will be encountered.

8. Bibliography

[1] K. Zeger and A. Gersho, "Zero Redundancy Channel Coding in Vector Quantisation", *Electronics Letters,* Vol. 23, No. 12, June 1987, pp. 654-656.

[2] N. Farvardin and V. Vaishampayan, "Optimal Quantizer Design For Noisy Channels: An Approach to Combined Source-Channel Coding", *IEEE Trans. Info. Theory,* (to appear). *System Research Center Tech. Report,* TR 85-33, University of Maryland, College Park, Maryland, November 1985.

[3] Noll, P , "Effects of Channel Errors on the Signal-to-Noise Performance of Speech Encoding Systems", *Bell Sys. Tech. J.* Vol. 54, pp. 1615-1636, November 1975.

[4] Makhoul, J., Roucos, S., and Gish, H. , "Vector quantization in Speech Coding" *Proc. of the IEEE,* Vol. 73, No. 11, pp. 1551-1588, November 1985.

[5] Chen, J. H., Davidson, G., Gersho, A. and Zeger, K. , "Speech Coding for the Mobile Satellite Experiment," *Proceedings IEEE Inter. Conf. on Communications,* Seattle, June 1987, pp. 756-763.

[6] Gersho, Allen, "On The Structure Of Vector Quantizers," *IEEE Trans. Inf. Theory.,* vol. IT-28, No. 2 , pp. 157-166, March 1982.

[7] Ayanoglu, Ender, "Trellis Encoding For Sources And Channels," *Tech. Report No. L907-86-1,* Information Systems Lab, Stanford University, August 1986.

[8] Lloyd, S.P., "Least Squares Quantization in PCM", *IEEE Trans. Infom. Theory,* Vol. IT-28, pp. 129-137, March 1982.

[9] Y. Linde, A. Buzo, and R. Gray, "An Algorithm For Vector Quantizer Design," *IEEE Trans. Comm.,* Vol. COM-28, pp. 84-95, January 1980.

[10] J.R.B. De Marca, N.S. Jayant, "An Algorithm for Assigning Binary Indices to the Codevectors of a Multi-Dimensional Quantizer," *Proceedings IEEE Inter. Conf. on Communications,* Seattle, June 1987, pp. 1128-1132.

[11] A. Kurtenbach and P. Wintz, "Quantization for Noisy Channels", *IEEE Trans. Communications,* Vol. COM-17, April 1969, pp. 291-302.