The 14th Symposium on Information Theory
and Its Applications (SITA '91)
Ibusuki, Japan, December 11-14, 1991

Asymptotic Noisy Channel Vector Quantization Via Random Coding

Kenneth Zeger
Dept. of Electrical Engineering
University of Hawaii
Honolulu, HI 96822

Victor Manzella
Dept. of Electrical Engineering
University of Rhode Island
Kingston, RI 02881

Abstract: The asymptotically minimal distortion of vector quantizers (VQ's) in the presence of channel noise is studied via random coding techniques. First, an upper bound is derived for the average r^{th}-power distortion of channel optimized VQ. The upper bound decays approximately as $O(N^{-r/r+k})$ for small bit error probabilities, as compared to $O(N^{-r/k})$ in Zador's generalization of Bennett's integral for noiseless channels. Next, we consider high resolution regular VQ's, with randomly chosen index assignments. A useful formula is derived that provides a lower bound on the SNR of optimal noisy channel VQ's. In addition, it is shown that the mean-square distortion of a noisy channel VQ is in probability asymptotically bounded away from zero.

1. Introduction

Recent interest in combined source/channel coding for bandlimited radio channels has motivated research toward quantifying the affects of channel noise on quantization systems. In [1,2] algorithms are introduced for finding locally optimal codevector index assignments (or labelings) to quantizer codevectors, so as to minimize the average distortion resulting from a particular assignment. Using the assumption of a "greedy" index assignment, some numerical high resolution bounds for noisy channel vector quantization were given in [3] in terms of integrals of point density functions.

Zador's [4] generalization of Bennett's integral shows that the asymptotic mean r^{th}-power distortion of an optimal N-point, k-dimensional vector quantizer decays as $O(N^{-r/k})$, assuming a noiseless channel. No such explicit formula has yet been displayed for quantizers in the presence of channel noise.

In the study of asymptotic VQ, it is implicitly assumed that as N grows, there exists an ever increasing bandwidth available for transmission. The binary transmission from N-point quantization corresponds to making $\log_2 N$ uses of a channel. On one hand, Shannon's channel coding theorem indicates that one could hope to reliably convey at most $C \log_2 N$ of these bits, where C is the channel's capacity. However, minimizing average quantization error does not necessarily imply one should convey the maximal amount of binary data. In fact, it might be desirable to tolerate some bit errors in order to increase the effective resolution of the quantization component. It can be shown that the minimal average r^{th}-power distortion for VQ on a noisy channel decays to zero as the number of transmitted bits per sample grows, provided one is willing to block together multiple input samples before transmission and thus incur delay. It has been an open problem, however, to find the rate of decay of the minimum distortion for zero delay quantizers (if it decays at all). Part of the difficulty in determining this lies in the complexity of mathematically analyzing the index assignment problem.

In this paper we present several results that help to answer these questions. First, we show that for optimal high resolution vector quantization on a Binary Symmetric Channel with bit error probability ε the MSE decays to zero at least as fast as some negative power of N. The decay rate bound approaches $O(N^{-r/r+k})$ as $\varepsilon > 0$ decreases, as opposed to the $O(N^{-r/k})$ decay rate when $\varepsilon = 0$. As the vector dimension k grows, the decay rate of the noisy channel bound approaches that of the noiseless channel bound.

Second, we introduce a random coding technique to analyze the MSE of noisy channel VQ's. Here we average the MSE of a given *regular* VQ over all possible index assignments. (Recall that regular VQ's, such as those designed optimally for noiseless channels, have the property that every codevector lies in its associated encoder cell.) The *expected MSE* thus obtained gives an asymptotic upper bound on the MSE of any VQ having a *better than average* index assignment. It thus provides a mathematical tool analogous to Zador's formula for analytically describing the MSE. Of theoretical interest, we also show that the average mean-square distortion on a BSC is, in probability, asymptotically bounded away from zero.

2. Upper Bound on Asymptotic Channel Optimized VQ Distortion

Bennett's formula provides a useful rule of thumb of "6 dB/bit" increase in SNR for each bit added to a scalar quantizer. It turns out that this is reasonably accurate for many low resolution cases as well. Below, it is shown that on a noisy channel (and small ε) an optimal quantizer's average distortion decreases asymptotically at least as fast as about $O(N^{-r/r+k})$. It is important to point out that this result does not assume that the centroid condition is necessarily satisfied.

Theorem 1: Let f be the pdf of a k-dimensional random vector with compact support. The minimum average r^{th}-power distortion of an N-point noisy channel vector quantizer on a Binary Symmetric Channel with crossover probability ε is asymptotically bounded above by $O(N^{-g(\varepsilon)})$, where g is a continuous and monotonically decreasing function from $g(0) = r/r+k$ to $g(1/2) = 0$.

Proof: To derive the stated upper bound it suffices to exhibit any noisy channel quantizer that satisfies the bound.

This work was supported in part by the National Science Foundation and Hewlett-Packard Co.

Assume for convenience that $N = 2^n$ and $M = 2^m$, for positive integers $n \geq m$, and consider an N-point, k-dimensional vector quantizer Q_N that is a cascade of an optimal noiseless channel M-point quantizer Q_M and an (n,m) error correcting coder $\psi_{n,m}:\{0,1,\}^m \rightarrow \{0,1,\}^n$. More precisely, if E_N and D_N are the encoder and decoder of Q_N (similarly for Q_M), then $E_N = \psi_{n,m} \circ E_M$, and $D_N = \phi_{n,m} \circ D_M$, where ϕ is the channel decoder, typically a maximum likelihood decoder and $Q_N = D_N \circ E_N$.

The capacity of a BSC with crossover probability ε is $C = \ln 2 + \varepsilon \ln \varepsilon + (1-\varepsilon)\ln(1-\varepsilon)$ in nats per channel use. For any $R < C$, Shannon's channel coding theorem guarantees the existence of functions $\psi_{n,m}$ such that $m = [Rn/\ln 2]$ information bits can be reliably transmitted for every block of n total bits sent, as n becomes asymptotically large. More precisely, $[Rn/\ln 2]$ bits can be conveyed with a probability of error

$$P_e \leq e^{-nE_r(R)},$$

where $E_r(R)$ is the *error exponent* function [5]. On a BSC with

$$R < \ln 2 - H\left(\frac{\sqrt{\varepsilon}}{\sqrt{\varepsilon}+\sqrt{1-\varepsilon}}\right)$$

$$H(\delta) = -\delta\log\delta - (1-\delta)\log(1-\delta)$$

the error exponent is given by

$$E_r(R) = \ln 2 - 2\ln(\sqrt{\varepsilon}+\sqrt{1-\varepsilon}) - R. \qquad (1)$$

The mean r^{th}-power error of a noisy channel vector quantizer, averaged over both the source and channel statistics is given by $D = E\,\|\mathbf{x}-\mathbf{y}_j\|^r$. By conditioning the expectation over the events that either a channel error does or doesn't occur, the MSE can be written as

$$\begin{aligned} D &= E[\|\mathbf{x}-\mathbf{y}_j\|^r \mid j=i](1-P_e) + E[\|\mathbf{x}-\mathbf{y}_j\|^r \mid j \neq i]P_e \\ &\leq E\,\|\mathbf{x}-\mathbf{y}_i\|^r + E[\|\mathbf{x}-\mathbf{y}_j\|^r \mid i \neq j]P_e \\ &\leq D_u \overset{\Delta}{=} G_1 2^{-mr/k} + G_2 e^{-nE_r(R)}. \end{aligned} \qquad (2)$$

where

$$G_1 = b_{r,k}\,\|f\|_{k/k+r},$$

$$G_2 = \text{diam}(\text{supp}(f))^r,$$

$b_{r,k}$ is a constant independent of f and m, $\text{diam}(\text{supp}(f))$ is the diameter of the support of f, and

$$\|f\|_p \overset{\Delta}{=} \left[\int_{\mathbf{R}^k} |f|^p\right]^{1/p}.$$

For each n we minimize the upper bound D_u over all possible rates R. If one transmits at a rate R very close to capacity C, then the number of information bits $Rn/\ln 2$ will be large and thus the quantization error, $E\,\|\mathbf{x}-\mathbf{y}_i\|^r$, will be small; however, the probability of a channel error going uncorrected can not be as tightly upper bounded for large R's so that the term P_e will contribute more to the overall distortion D. Thus, there is an important tradeoff in this case between (1) designating more of the transmitted bits as information bits to reduce quantization error, and (2) devoting more of the transmitted bits toward error control coding to drive the probability of an uncorrected channel error to zero faster. We find the rate R that optimizes this tradeoff in the bound D_u and show that it is best in this case to transmit information at a rate slightly less than $k/k+r$ bits per channel use for small values of ε. Setting $\partial D_u/\partial R = 0$ gives

$$\frac{G_1}{G_2}\frac{r}{k}e^{-n(rR/k - E_r(R))} = -\frac{\partial E_r}{\partial R} = 1.$$

Solving for R using (1) yields

$$R = \frac{k}{k+r}\left[\ln 2 - 2\ln(\sqrt{\varepsilon}+\sqrt{1-\varepsilon}) + \frac{1}{n}\ln\left[\frac{G_1}{G_2}\frac{r}{k}\right]\right] \qquad (3)$$

and substituting this optimal value of R into (2) gives

$$D_u = \left[\left(\frac{G_1}{k}\right)^k\left(\frac{G_2}{r}\right)^r\right]^{\frac{1}{k+r}} \cdot N^{-g(\varepsilon)}$$

where

$$g(\varepsilon) = \left(\frac{r}{k+r}\right)\left[1 - \frac{2\ln(\sqrt{\varepsilon}+\sqrt{1-\varepsilon})}{\ln 2}\right].$$

The function g is clearly continuous and it is easy to show that $g'(\varepsilon) < 0$ for $\varepsilon \in [0, 1/2)$, implying that g is monotonic as stated.

□

On channels with low bit error probability ε, a Taylor series approximation of g gives

$$g(\varepsilon) \approx \frac{r}{k+r}\left[1 - \frac{2\sqrt{\varepsilon}}{\ln 2}\right] \approx \frac{r}{k+r}.$$

Furthermore, from (3) it can be seen that to achieve the average distortion upper bound D_u, we must convey information across the channel at a rate of $R \approx g(\varepsilon)$ which is approximately $r/k+r$ bits per channel use for good channels.

As the dimension k of the input vectors grow, the upper bound for small ε, $D_u \approx O(N^{-r/r+k})$, approaches $O(N^{-r/k})$, the asymptotic decay rate of optimal VQ distortion on a noiseless channel. An interesting observation is that for a scalar quantizer on a BSC with small ε, the MSE drops at least as fast as about $O(N^{-2/3})$, which is the same rate of decay as an optimal scalar quantizer on a noiseless channel, but using only one third as many bits.

3. Coding Theorems for Quantizers without Channel Optimization

The asymptotic performance of a k-dimensional VQ in the presence of channel noise is examined for *typical* index assignments. To do this we use a model in which an N point quantizer Q_N has an arbitrary fixed codevector labeling and index assignment is handled by a randomly selected permutation from S_N, the symmetric group on N elements.

Let $\mathbf{X}$ denote the source random vector (at the VQ input), and let f be the pdf of $\mathbf{X}$, with compact and connected support, $S_f \subset \mathbf{R}^k$. Let $\mathbf{y}_i$ denote the i^{th} codevector of Q_N and $R_i^{(N)}$ the i^{th} partition cell, abbreviated as R_i and having volume ΔV_i. The probability of a source vector lying in the i^{th} partition cell is

$$P_i = P_i(N) = \int_{\Delta V_i} f(\mathbf{x})\,d\mathbf{x}$$

Finally, we define

$$d(N) \overset{\Delta}{=} \max_{i=1,\ldots,N} diam(R_i^{(N)})$$

In addition to the random source $\mathbf{X}$, we have Π, the random permutation from S_N, and $\mathbf{C}$, the random mapping of input indices to output indices across the channel. Note that $\mathbf{X}$ has range contained in $\mathbf{S}_f \subset \mathbf{R}^k$, Π has range S_N, which can be viewed as a subset of $\mathbf{Z}^N$ of size $N!$, and $\mathbf{C}$ has range which can be viewed as $\mathbf{Z}_N{}^2$. With this model the source encoder output $\mathbf{i}$ is a random variable depending only on $\mathbf{X}$, while the decoder input $\mathbf{j}$ depends on $\mathbf{X}, \Pi$, and $\mathbf{C}$. The total distortion $\mathbf{D} = \|\mathbf{X}-\mathbf{y_j}\|^2$ also depends on $\mathbf{X}, \Pi$, and $\mathbf{C}$.

To obtain the mean squared distortion $E_{\mathbf{X},\Pi,\mathbf{C}}[\mathbf{D}]$ we may write $D = D_s + 2D_{sc} + D_c$ where $D_s = \|X - y_I\|^2$, $D_{sc} = (X - y_I)^t (y_I - y_J)$, and $D_c = \|y_I - y_J\|^2$. Since D_s depends only on X and $I = i(X)$ we have:

$$E_{X,\Pi,C}[D] = E_X[D_s] + 2E_{X,\Pi,C}[D_{sc}] + E_{X,\Pi,C}[D_c]$$

Assuming each π is equally likely and selected independently of the source random variable and channel errors, we get

$$E_{X,\Pi,C}[D_{sc}] = \sum_{\pi} \frac{1}{N!} E_{X,C}[D_{sc}(X,\pi,C)]$$

Next, we compute

$$E_{X,\Pi,C}[D_{sc}] = E_{\Pi}\left[\sum_{i=1}^{N}\sum_{j=1}^{N} (c_i - y_i)^t (y_i - y_j) P_{\Pi(j)|\Pi(i)} P_i\right]$$

$$\triangleq E_{\Pi}[D_{sc}(\Pi)].$$

Similarly, defining

$$D_c(\Pi) \triangleq \sum_{i=1}^{N}\sum_{j=1}^{N} \|y_i - y_j\|^2 P_{\Pi(j)|\Pi(i)} P_i$$

gives $E_{X,\Pi,C}[D_c] = E_{\Pi}[D_c(\Pi)]$, and defining $D(\Pi) \triangleq E_X[D_s] + 2D_{sc}(\Pi) + D_c(\Pi)$ we get

$$E_{X,\Pi,C}[D] = E_X[D_s] + 2E_{\Pi}[D_{sc}(\Pi)] + E_{\Pi}[D_c(\Pi)]$$

i.e. the expected distortion may be written as the expected value over Π of three distortion terms already averaged over X and C. We are interested in the asymptotic behavior of $E_{\Pi}[D(\Pi)]$ as well as that of $Var_{\Pi}[D(\Pi)]$.

We now consider $E_{\Pi}[D]$ and note that the first of its three terms is asymptotically upper bounded by Zador's formula [4]. The cross-term can be shown to converge to zero whenever we restrict our attention to sequences of regular quantizers.

$$|E_{\Pi}[D_{sc}(\Pi)]| = \left|\sum_{i=1}^{N}\sum_{j=1}^{N} (c_i - y_i)^t (y_i - y_j) P_i \sum_{\pi} P_{\pi(j)|\pi(i)}\right|$$

$$\le \frac{1-(1-\varepsilon)^{\log_2 N}}{N-1} \sum_{i=1}^{N}\sum_{j=1}^{N} P_i \|c_i - y_i\| \|y_i - y_j\|$$

$$\le \left[1-(1-\varepsilon)^{\log_2 N}\right] \frac{N}{N-1}\, diam(S_f)\, d(N)$$

$$\to 0 \quad as \ N \to \infty$$

(In particular if we have a sequence of quantizers, each of which satisfies the *centroid condition*, then this cross-term is identically zero for all N.) So for sequences of regular quantizers employing random index assignment, the total distortion is asymptotically upper bounded by Zador's formula (the asymptotic form for $E_X[D_s]$), plus the term $E_{\Pi}[D_c(\Pi)]$, whose asymptotic behavior will be determined in the following proposition. We first state a useful lemma.

Lemma 1: If $N \ge 2$, then for every pair (i, j) with $i \ne j$

$$\sum_{\pi \in S_N} P_{\pi(j)|\pi(i)} = [1-(1-\varepsilon)^L]\, N\, (N-2)!$$

Proposition 1: Suppose that the source random vector X has mean m_X and components X_n with variances $\sigma_{X_n}^2$, $n = 1, 2, \cdots, k$. If $\varepsilon > 0$, then

$$\lim_{N\to\infty} E_{\Pi}[D_c(\Pi)] = \sum_{n=1}^{k} \sigma_{X_n}^2 + \int_{S_f} \|x - m_X\|^2 \lambda(x)\, dx$$

Proof: Recall that

$$E_{\Pi}[D_c(\Pi)] = \frac{1}{N!} \sum_{i=1}^{N}\sum_{j=1}^{N} \|y_i - y_j\|^2 P_i \sum_{\pi} P_{\pi(j)|\pi(i)}$$

$$= \left[(1-(1-\varepsilon)^{\log_2 N}) \frac{N}{N-1}\right] \sum_{i=1}^{N}\sum_{j=1}^{N} P_i \frac{1}{N} \|y_i - y_j\|^2$$

by Lemma 1. Note that the bracketed term converges to 1 as $N \to \infty$ (since $\varepsilon > 0$). Defining $h(N)$ by

$$h(N) = \sum_{i=1}^{N}\sum_{j=1}^{N} P_i \frac{1}{N} \|y_i - y_j\|^2 ,$$

we then observe that $h(N) = E[\|X^{(N)} - Z^{(N)}\|^2]$ where $X^{(N)}$ and $Z^{(N)}$ are independent random vectors for each N, $X^{(N)}$ has pmf $p_{X^{(N)}}(y_i) = P_i = P_i(N)$, $1 \le i \le N$, and $Z^{(N)}$ has pmf $p_{Z^{(N)}}(y_j) = 1/N$, $1 \le j \le N$. So for each N,

$$h(N) = E[\|X^{(N)} - m_X\|^2] + 2E[(X^{(N)} - m_X)^t (m_X - Z^{(N)})] + E[\|Z^{(N)} - m_X\|^2]$$

Note that the middle term is zero by the independence of $X^{(N)}$ and $Z^{(N)}$. For the other two terms we get

$$E[\|X^{(N)} - m_X\|^2] \to \int_{S_f} \|x - m_X\|^2 f(x)\, dx = \sum_{n=1}^{k} \sigma_{X_n}^2$$

$$E[\|Z^{(N)} - m_X\|^2] \to \int_{S_f} \|x - m_X\|^2 \lambda(x)\, dx$$

□

An interesting feature of this noisy channel result is that for any source the expected distortion for a typical index assignment tends asymptotically to a strictly positive value. Note that the *regularity* assumption played a key role in this conclusion since it guaranteed the decay to zero of the cross-term $E_{\Pi}[D_{sc}(\Pi)]$. We propose an approximate asymptotic form for the channel distortion, in order to mathematically predict the MSE behavior:

$$E_{\Pi}[D_c(\Pi)] = \varepsilon(\log_2 N) \frac{N}{N-1} \left[\sum_{n=1}^{k} \sigma_{X_n}^2 + \int_{S_f} \|x - m_X\|^2 \lambda(x)\, dx\right] \qquad (4)$$

Eq. (4) provides an asymptotic approximation for the channel MSE when a *typical index assignment* is used. Thus it serves as an asymptotic *upper bound* on the channel MSE for all *better than average* index assignments. For *realistic* large N, one expects this bound to be most useful for quantizers satisfying the centroid condition, for then the cross-term is guaranteed to be zero for *all* N. We next examine the asymptotic behavior of the variance,

$$Var_{\Pi}[D(\Pi)] = E_{\Pi}[(2D_{sc}(\Pi) + D_c(\Pi))^2] - E_{\Pi}^2[2D_{sc}(\Pi) + D_c(\Pi)]$$

As noted in the discussion preceding Proposition 1, $E_{\Pi}[D_{sc}(\Pi)] \to 0$ as $N \to \infty$, so the second term is given asymptotically by $(\lim_{N\to\infty} E_{\Pi}[D_c(\Pi)])^2$, whose existence was established by Proposition 1. We state the following proposition but omit its lengthy proof. This proposition, together with Proposition 1 and Chebychev's Inequality yield the following corollary.

Proposition 2: The variance of the total distortion (over randomly chosen index assignments) decays to zero as $N \to \infty$.

Corollary 1: Let f be the pdf of a random vector with compact support and let Q_N be an N-level regular VQ, used on a BSC with crossover probability ε. Then, in probability, the MSE of Q_N is asymptotically bounded away from zero.

4. Experimental Results

Experiments were performed in an attempt to deterreasonable values of N. A scalar quantizer was considered for values $N = 2^L$, where $L = 4, 5, \cdots, 10$, and codebooks were designed using the *LBG* algorithm on Gaussian *i.i.d.* samples. For each quantizer Q_N, ε was fixed (to a value between 10^{-5} and 10^{-1}), and the channel distortion was computed for 20 randomly selected index assignments. These distortion values were averaged to give an estimate of $E_\Pi(D_c)$ which was then used to obtain the overall output SNR.

The scalar results for $N = 32$ are shown in Figure 1.

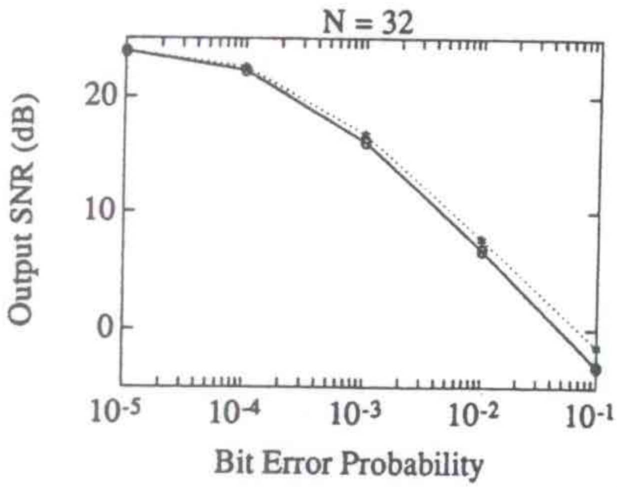

Figure 1: SNR vs. channel error probability for Lloyd-Max scalar quantizer at rate = 3 bits/sample. Solid line = theoretical, Dotted line = experimental.

The solid line represents the theoretical asymptotic form given by Eq. (4) and the dotted line the experimentally obtained SNR's. As can be seen, the form lies very close to the experimentally determined SNR's for values of ε between 10^{-5} and 10^{-1}. The cause of the difference between the theoretical and experimental curves is currently under investigation.

5. Conclusion

An upper bound is given on the rate of decay of the average distortion of a channel optimized VQ. A technique is then introduced for analyzing the asymptotic performance of VQ's in the presence of channel noise. By randomizing the binary index assignments, a useful MSE formula is obtained that closely matches experimental results.

REFERENCES

[1] J. De Marca and N. Jayant, "An Algorithm for Assigning Binary Indices to the Codevectors of a Multi-Dimensional Quantizer," *Proceedings IEEE Inter. Conf. on Communications*, Seattle, June 1987.

[2] K. Zeger and A. Gersho, "Pseudo-Gray Coding," *IEEE Trans. Communications*, vol. COM-38, no. 12, pp. 2147-2158, December 1990.

[3] S. McLaughlin and D. Neuhoff, "Asymptotic Bounds in Source-Channel Coding," *IEEE Int. Symp. Info. Theory (ISIT)*, Budapest, Hungary, June 1991.

[4] P. Zador, "Asymptotic Quantization Error of Continuous Signals and the Quantization Dimension," *IEEE Trans. Inf. Theory.*, vol. IT-28, no. 2, pp. 139-149, March 1982.

[5] R. Gallagher, *Information Theory and Reliable Communication*, John Wiley and Sons, New York, 1968.