Properties of the Natural Binary Code Assignment for Uniform Scalar Quantizers under Channel Errors

G. Ben-David and D. Malah
Technion - Israel Institute of Technology,
Department of Electrical Engineering,
Haifa 32000, ISRAEL
Tel: 972-4-294751
FAX: 972-4-323041
e-mail: gal@neptun.technion.ac.il

Abstract

The Uniform Scalar Quantizer is a versatile and easy to implement digital coding system. The quantization process introduces distortion due to the representation of an analog signal by a finite number of levels. Transmitting the digital code over a noisy channel may result in incorrect reconstruction of samples at the receiver. In this paper, the problem of the assignment of digital codes to quantization levels is examined. It is shown that for uniform quantizers, the channel distortion under Natural Binary Code assignment is smaller than the ensemble average distortion over all possible assignments, for mid and high resolution quantizers. For the uniform source, a new proof for the optimality of the Natural Binary Code assignment is given. An upper bound on the distortion over all assignments in this case is also introduced.

I. Introduction

The Scalar Quantizer (also referred to as PCM coding) is a simple representation technique of signals by a finite number of levels. Each level is assigned a digital code, which is transmitted via a digital channel. When transmitting the digital information over a non-ideal channel, channel-errors are introduced and incorrect analog levels may be reconstructed at the receiver (We assume that no error detection/correction means are used). An inherent outcome of the quantization process is the Quantization Distortion, which is the difference between the amplitude and its reconstruction level. In the presence of channel errors, Channel Distortion is caused by the difference between the actual reconstruction level and the one that would have been reproduced under ideal channel conditions. The difference between the input amplitude and the actual reconstruction level is referred to as the Overall Distortion.

The effect of channel errors on quantizers performance is treated in [1-11]. Most of the research done assumed a Binary Symmetric Channel, where random bit errors occur at a
given rate. The Channel Distortion and hence the Overall Distortion are affected by the assignment of the binary code to each level. As mentioned above, a channel error may cause an incorrect decoding of a reconstruction level. If this incorrect level is not "far" from the level that should have been decoded, the distortion caused is small. This heuristic argument motivates the search for an optimal code assignment for the reconstruction levels indices. This is a problem of a combinatorial character. For example, a 4-bit quantizer has as many as \( 16! = 2 \cdot 10^{13} \) possible assignments. Bounds over source probabilities on the performance of quantizers under channel errors, for a given assignment, were shown in [1]. An analysis of specific systems with given assignments are presented in [2]. It is shown that coarse quantizers are more robust to channel-errors than quantizers with a higher number of levels. The relation between quantization and channel distortions are discussed in [3]. It is shown there that for Centroid quantizers and Mean Squared-Error distortion, these distortions are orthogonal, and their sum is equal to the overall distortion. Effects of channel errors in the perspective of speech applications are found in [4]. Non-linear PCM systems are treated in [5], where an ad-hoc approach is used to gain a "good" assignment. Another approach which changes the design of the quantizer levels, in order to reduce the effect of channel errors is shown in [6]. An algorithm which improves the assignment by iteratively switching the positions of codewords is presented in [7],[8],[9]. This algorithm converges, however, to a local minimum. Attempts to find a global minimum over all possible assignments, using the simulated annealing algorithm are reported in [10]. For the special case of a Uniform Source and a Uniform Scalar Quantizer it is shown in [11] that the optimal assignment is the Natural Binary Code (NBC). In this paper we suggest a simpler proof of this property. Also, as a by-product of this proof, a new upper bound (over all assignments) on the channel distortion is derived.

The paper is organized as follows. In the following section we show that the channel distortion for the NBC assignment is less than the ensemble-average distortion for uniform quantizers with four bits or more. Next we present our proof that the NBC is optimal for the uniform source quantized by uniform quantizer. The new upper bound, mentioned above, on the channel distortion for the uniform source is also provided.

II. The Natural Binary Code Assignment

The continuous signal at the quantizer input is assumed to be a random variable with a probability density \( p(x) \). The \( N \)-level quantizer represents input amplitudes in the interval \( [b_k, b_{k+1}] \) by a reconstruction level \( r_k, k = 0, 1, \ldots, N - 1 \), where \( b_0 = -\infty \) and \( b_N = \infty \). Assignment of channel indices to the reconstruction levels is a one-to-one permutation mapping \( \Lambda : \{0, 1, \ldots, N - 1\} \rightarrow \{0, 1, \ldots, N - 1\} \). The channel index \( i \) corresponds to the reconstruction level \( r_i \) by the mapping function \( \phi(i) = r_i = r_{A(i)} \). The Natural Binary Code, for example, corresponds to the unity mapping, \( \phi(i) = r_i \).
Consider an input amplitude \( x \) which is represented by a specific reconstruction level \( \phi(i) \). However, because of channel errors, another level \( \phi(j) \) may be reconstructed. This event happens with the channel transition probability: \( q_{ij} = \text{Prob}[j \text{ received } | i \text{ transmitted}] \). The Mean Squared Error (MSE) overall distortion between the input and output signals is the expected value of the squared distance between an input sample and its reconstruction,

\[
E\{(x - \hat{x})^2\} = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} q_{ij} \int (x - \phi(j))^2 p(x) \, dx.
\]  

(1)

In the current analysis we assume that there are \( N = 2^L \) reconstruction levels, and the \( L \) bits are transmitted via a Binary Symmetric Channel (BSC). The transition probability is \( q_{ij} = q^H_i (1 - q)^{L-H_i} \), where \( H_i \) denote the Hamming distance between the binary representation of the indices \( i \) and \( j \), and \( q \) is the channel Bit Error Rate (BER). We assume that \( q < 0.5 \).

In the following analysis, only the Channel Distortion is considered, which is given by:

\[
D_c = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} q_{ij} (\phi(i) - \phi(j))^2
\]

(2)

where \( P_i \) is the probability of transmitting the channel index \( i \), corresponding to the reconstruction level \( r_{a(i)} \):

\[
P_i = \frac{b_{a(i)}}{b_{\text{total}}}
\]

(3)

Without loss of generality, we assume that the quantizer under consideration is normalized to the interval \([-1, 1]\). For the Natural Binary Code the reconstruction levels are \( \phi(i) = (i - N/2 + \Delta) / N/2, \ i = 0,1,\ldots,N-1 \). The offset \( \Delta \) can represent midrise (\( \Delta = 1/N \)) or midtread quantizers (\( \Delta = 0 \)) [2], and does not affect the current analysis.

The Natural Binary Code (NBC) has the special property that for small Bit-Error-Rates (BER), the MSE channel-distortion normalized by \( q \) tends to the same limit for all possible source statistics [1], given by:

\[
\lim_{q \to 0} \frac{D_c}{q} = \sum_{i=0}^{N-1} \left( \frac{1}{2^i} \right)^2.
\]

(4)

The series in (4) converges rapidly and is equal to 1.328 for a 4-bit quantizer and tends to 1.333 as the number of bits \( L \to \infty \). It is straightforward to compare the performance of the NBC with the ensemble-average distortion (The average distortion over all \( N! \) possible assignments). It is shown in [10] that the ensemble-average distortion \( \overline{D}_c \) (for small BER) satisfies:

\[
\lim_{q \to 0} \frac{\overline{D}_c}{q} = \frac{L}{N-1} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} (r_i - r_j)^2.
\]

(5)
For the normalized uniform quantizer, using algebraic equalities, one can verify that:

\[
\lim_{q \to 0} \frac{D_q}{q} = \frac{2}{3} \frac{(2N-1)L}{N} - \frac{4}{N} \frac{L}{N-1} \sum_{i=0}^{N-1} P_i((N-1)i - i^2).
\] (6)

In (6), the second term is just a weighted average of a set of constants. Since source probabilities are positive, and their sum is unity, we can limit the asymptotic performance from above and below, as follows:

For \(i=0,1,...,N-I\), the constants are non-negative. The constants corresponding to \(i=0\) and \(i=N-I\) are zero. The largest distortion is caused, therefore, by the synthetic probability which is concentrated at either end, causing the second term to vanish.

\[
\left( \lim_{q \to 0} \frac{D_q}{q} \right)_{\text{max}} = \frac{2}{3} \frac{(2N-1)L}{N} = \frac{1}{3} \frac{L}{N} (4N-2) = \frac{4}{3} L.
\] (7)

The lower bound is achieved by the probability which is concentrated at \(i = N/2\), since \(N\) is even. The lower bound turns out to be:

\[
\left( \lim_{q \to 0} \frac{D_q}{q} \right)_{\text{min}} = \frac{2}{3} \frac{(2N-1)L}{N} - \frac{4}{N \cdot N-1} \left[ (N-1) \frac{N}{2} - \frac{N^2}{4} \right] = \frac{1}{3} L \left[ N + 1 + \frac{3}{N-1} \right] \approx \frac{1}{3} L.
\] (8)

The normalized (by \(q\)) ensemble-average distortion for a given source, is between those bounds, For uniform source probabilities, \(P_i = 1/N\), \(i=0,1,...,N-1\):

\[
\left( \lim_{q \to 0} \frac{D_q}{q} \right)_{\text{uniform}} = \frac{2}{3} \frac{(2N-1)L}{N} - \frac{4}{N^2} \frac{L}{N-1} \sum_{i=0}^{N-1} [(N-1)i - i^2] = \frac{1}{3} \frac{L}{N} (2N+2) = \frac{2}{3} L.
\] (9)
The lower and upper bounds as well as the normalized ensemble-average distortion for the uniform source are plotted in Fig. 1, as a function of the number of bits \( L \). We use line graphs for clarity. For comparison, the performance of the NBC (i.e., \( \bar{D}_c/q \) from (4)) is also shown (lower dashed line).

Fig. 1 Upper & Lower bounds on the normalized ensemble-average distortion \( (\bar{D}_c/q) \) over all possible assignments, as compared with the performance \( (D_c/q) \) of the Natural Binary Code assignment

It is seen that the NBC has a lower distortion than the ensemble-average distortion for all possible source probabilities, for quantizers of 4 bits and up. Yet, even for quantizers with less than 4 bits, the NBC channel distortion is quite close to the lower bound. It is interesting to see that although the ensemble-average distortion grows linearly with the number of bits, the NBC distortion clearly tends towards a constant level. The last observation is that the distortion for the NBC is always lower than the ensemble-average distortion for the uniform source. In the next section it is shown that the NBC is indeed the optimal assignment for that source.
III. Results for a Uniform Source

The property that the NBC is the optimal assignment for a uniform input source is shown in [11]. Nevertheless, our proof here is much shorter and in addition gives a new upper bound on the channel distortion over all possible assignments.

In the case of a uniform scalar quantizer and a uniform source, we obtain from (2):

\[
D_c = \frac{1}{N} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} q_{ij} (\phi(i) - \phi(j))^2 = \\
\frac{2}{N} \left[ \Phi^T \cdot \Phi - \Phi^T \cdot Q_N \cdot \Phi \right]
\]

where \( \Phi = [\phi(0) \quad \phi(1) \quad \cdots \quad \phi(N-1)]^T \), and \( Q_N \) is the channel transition matrix with entries \( q_{ij} \). For different assignments the entries of \( \Phi \) are permutations of the reconstruction levels.

In order to find upper and lower bounds over all possible assignments, [11] suggested to allow the entries of \( \Phi \) in (10) to be continuous (to be denoted by \( z \)), while keeping the first two moments unchanged.

The following optimization problem can now be stated:

\[
\begin{align*}
\min_{z} & \quad f(z) = \frac{2}{N} \left[ z^T \cdot z - z^T \cdot Q_N \cdot z \right] \\
\text{subject to:} & \quad 1^T \cdot z = 0; \quad z^T \cdot z = \Phi^T \cdot \Phi = k^2
\end{align*}
\]

where \( 1 \) is a \( N \)-dimensional vector with unity entries, and \( k \) is the variance of the output levels. Removing the first (constant) term in (11), and the constant factor \( 2/N \), we get:

\[
\begin{align*}
\max_{z} \min_{z} & \quad z^T \cdot Q_N \cdot z \\
\text{subject to:} & \quad 1^T \cdot z = 0; \quad z^T \cdot z = \Phi^T \cdot \Phi = k^2.
\end{align*}
\]

Using Lagrange multipliers:

\[
L(z, \mu_1, \mu_2) = z^T Q_N z + \mu_1 (z^T z - k^2) + \mu_2 (1^T z),
\]

we obtain the following first order conditions:

\[
\begin{align*}
[Q_N + \mu_1 I]z &= -\frac{\mu_2}{z} \cdot 1 \\
1^T z &= 0 \\
z^T z &= k^2
\end{align*}
\]

Using the zero average condition on \( z \) to eliminate \( \mu_1 \), and substituting \( w = z/k \), one gets the following matrix set of nonlinear equations:

-6-
\[ [Q_N - \mathbf{w}^T \mathbf{Q_N w} \cdot I] \mathbf{w} = -\frac{\mu_2}{2k} \mathbf{1}, \]  

(15)

where \( I \) is the \( N \times N \) unity matrix.

For the BSC, \( \mathbf{Q_N} \) is symmetric and therefore has a unitary diagonalization: \( \mathbf{Q_N} = \mathbf{V} \cdot \mathbf{\Lambda} \cdot \mathbf{V}^T \) with \( \mathbf{V} \mathbf{V}^T = \mathbf{I} \). Since \( \mathbf{Q_N} \) represent probabilities, the sum of its rows is one, or \( \mathbf{Q_N} \cdot \mathbf{1} = \mathbf{1} \); therefore \( \mathbf{1} \) is an eigenvector of \( \mathbf{Q_N} \). The change of coordinates: \( \mathbf{\xi} = \mathbf{V}^T \mathbf{w} \) is now used to simplify the equations. From the constraints stated in (14), \( \mathbf{\xi} \) has the following two properties:

\[ \mathbf{\xi}^T \mathbf{\xi} = \mathbf{w}^T \mathbf{w} = 1, \quad \mathbf{\xi}(0) = \mathbf{1}^T \mathbf{V} \mathbf{\xi} = \mathbf{1}^T \mathbf{w} = 0. \]

Using the orthogonality of the columns of \( \mathbf{V} \), the set of equations in (15) becomes:

\[ [\mathbf{\Lambda} - (\mathbf{\xi}^T \mathbf{\Lambda} \mathbf{\xi}) \cdot I] \mathbf{\xi} = -\frac{\mu_2}{2k} \mathbf{V}^T \cdot \mathbf{1} = \left[ -\frac{\mu_2 \sqrt{N}}{2k} \begin{array}{cc} 0 & 0 & \cdots & 0 \end{array} \right]^T, \]

(16)

where we assumed that the first column in \( \mathbf{V} \) is the eigenvector \( \mathbf{1}/\sqrt{N} \cdot \mathbf{1} \).

Defining,

\[ \beta \mathbf{\xi}^T \mathbf{\Lambda} \mathbf{\xi} = \sum_{i=0}^{N-1} \lambda_i \mathbf{\xi}^2(i) \]  

(17)

The set of equations is now:

\[ (\lambda - \beta) \cdot \mathbf{\xi}(0) = -\frac{\mu_2}{2k} \]  

(18)

\[ (\lambda_i - \beta) \cdot \mathbf{\xi}(i) = 0, \quad i = 1, 2, \ldots, N - 1 \]

Where \( \lambda_i \) are the eigenvalues of \( \mathbf{Q_N} \), and \( \lambda_0 = 1 \). Note that the first equation forces \( \mu_2 = 0 \), since \( \mathbf{\xi}(0) = 0 \), and some entries of \( \mathbf{\xi} \) must be nonzero since \( \mathbf{\xi}^T \mathbf{\xi} = 1 \).

It is shown in [11] that \( \mathbf{Q_N} \) has \( L+1 \) distinct eigenvalues, \( \lambda^{(m)} = (1-2q)^m \), \( m = 0, 1, \ldots, L \), each with multiplicity of \( \binom{L}{m} \). We assume without loss of generality that the eigenvalues in \( \mathbf{\Lambda} \) are arranged in decreasing order. Note that \( \lambda^{(0)} = \lambda_0, \lambda^{(1)} = \lambda_1, \ldots, \lambda^{(L)} = \lambda_L \). Since \( \beta = \lambda^{(m)} \), for a given \( m \), satisfies only a subset of the equations in (18), only the entries of \( \mathbf{\xi} \) corresponding to that eigenvalue may be nonzero. Therefore, the solution to the first order conditions (14), for the given \( m \), is a linear combination of the corresponding eigenvectors:

\[ \mathbf{z}_m = k \cdot \mathbf{\xi}_m = k \sum_{l: \lambda_l = \lambda^{(m)}} \mathbf{\xi}_m(l) \mathbf{V}_l \]  

(19)

where \( \sum_{l: \lambda_l = \lambda^{(m)}} \mathbf{\xi}_m(l) = 1 \), and \( \mathbf{V}_l \) is the \( l \)-th column of \( \mathbf{V} \).
The corresponding second derivative (Hessian) is:

\[
H^{(m)} = 2[Q_y + \mu I] = \\
= 2V[\Lambda - \xi^T \Lambda \xi \cdot I]V^T = \\
= 2V \cdot \text{diag}\{[(\lambda_0 - \lambda^{(m)}), (\lambda_1 - \lambda^{(m)}), \ldots, (\lambda_m - \lambda^{(m)})]\}V^T
\]  

For \( m=L \), \( \lambda^{(m)} = \lambda_y \), the smallest eigenvalue, \( H^{(m)} \) can be shown to be strictly positive-definite over the tangent-subspace [12] \( T^{(m)} = \{ y: y^T \xi = 0, \xi^T \xi = 0 \} \), hence \( z = V_y \) is a strict maximum (as we added a minus sign) of \( f(\xi) \) in the optimization problem (11). For \( m=1 \), \( \lambda^{(m)} = \lambda_1 = \cdots = \lambda_L \), \( H^{(m)} \) is positive definite over the tangent subspace \( T^{(m)} = \{ y: y^T y = 0, \xi_i^T y = 0, i = 1, 2, \ldots, L \} \) and hence the set:

\[
\left\{ \xi: \xi = k \sum_{i=1}^{L} \xi_i(l)V_i, \sum_{i=1}^{L} \xi_i^2(l) = 1 \right\}
\]  

achieves the minimum. The values of the maximum/minimum are obtained by substituting the optimal vectors into (11):

\[
f_{\max}(\xi) = \frac{2k^2}{N} \left[ 1 - \frac{\xi^T \Lambda \xi}{\xi_{\max}} \right] = \frac{2k^2}{N} \left[ 1 - (1 - 2q)^L \right] \\
f_{\min}(\xi) = \frac{2k^2}{N} \left[ 1 - \frac{\xi^T \Lambda \xi}{\xi_{\min}} \right] = \frac{2k^2}{N} \left[ 1 - (1 - 2q) \right] = \frac{4k^2}{N} q
\]

Note that the ratio between the maximum and minimum values in (22) tends to \( L \) for small BER. Finally, we need to prove that the NBC is constructed from an eigenvector of \( Q_y \) with an eigenvalue \( \lambda^{(m)} = (1 - 2q) \), corresponding to the minimum distortion. We left this proof to the Appendix. It should be mentioned that the upper bound, based on \( z = V_y \), does not construct an assignment.
In Fig. 2 the ratio between the upper bound on the channel distortion over all possible assignments, and the distortion under the NBC assignment (22) is plotted for a 3-bit uniform quantizer (8 levels). This ratio is compared with the ratio (shown by crosses) between the best (known to be the NBC) and the worst assignments, found by an exhaustive search over all 8! possible assignments.

Fig. 2 - Ratio between the upper bound on channel distortion over all possible assignments, and the distortion of NBC (22). 3-bits quantizer; uniform source, as a function of the BER $q$. Crosses mark the ratio between the distortion for the actual worst assignment and the NBC.
Conclusions

In this paper we have shown that the Natural Binary Code is a reasonable choice as an assignment for Uniform Scalar Quantizers for any input source, since it has a lower distortion than the ensemble-average distortion. It was also shown that for the uniform source, the NBC is the optimal assignment (a result obtained in [11] using a more complicated proof). An upper bound on the Channel Distortion over all assignments, for the uniform source is found. This upper bound is about $L$ (the number of bits) times larger than the optimal NBC distortion, and appears to be quite tight.

Acknowledgment
The authors are grateful to Steve McLaughlin for helpful discussions.

Appendix

We prove here, by induction, that the NBC assignment corresponds to an eigenvector of the matrix $Q_N$ with the eigenvalue $(1 - 2q)$. For a one bit quantizer, with reconstruction levels $\pm \alpha$, the verification is simple.

$$Q_2 \begin{bmatrix} -\alpha \\ \alpha \end{bmatrix} = \begin{bmatrix} 1 - q & q \\ q & 1 - q \end{bmatrix} \begin{bmatrix} -\alpha \\ \alpha \end{bmatrix} = (1 - 2q) \begin{bmatrix} -\alpha \\ \alpha \end{bmatrix}$$  \hspace{1cm} (A.1)

Assume that the NBC corresponds to an eigenvector of $Q_N$. We shall show below that this implies the same property for a uniform quantizer having one more bit (i.e., twice as many levels):

$$Q_{2N} \begin{bmatrix} -N\alpha & \ldots & -\alpha & \alpha & \ldots & N\alpha \end{bmatrix}^T =$$

$$= \begin{bmatrix} (1-q)Q_N & qQ_N \\ qQ_N & (1-q)Q_N \end{bmatrix} \begin{bmatrix} -N\alpha & \ldots & -\alpha & \alpha & \ldots & N\alpha \end{bmatrix}^T =$$

$$= \begin{bmatrix} (1-q)Q_N \begin{bmatrix} -N\alpha & \ldots & -\alpha \end{bmatrix}^T + qQ_N \begin{bmatrix} \alpha & \ldots & N\alpha \end{bmatrix}^T \\ qQ_N \begin{bmatrix} -N\alpha & \ldots & -\alpha \end{bmatrix}^T + (1-q)Q_N \begin{bmatrix} \alpha & \ldots & N\alpha \end{bmatrix}^T \end{bmatrix} =$$

$$= (1 - 2q) \begin{bmatrix} -N\alpha & \ldots & -\alpha & \alpha & \ldots & N\alpha \end{bmatrix}^T.$$

Hence: $Q_{2N}\varphi = (1 - 2q)\varphi$ for the NBC. QED.

In (A.2) we used the relations:

$$Q_N \begin{bmatrix} -N\alpha & \ldots & -\alpha \end{bmatrix}^T = \alpha Q_N \begin{bmatrix} \frac{N-1}{2} & \ldots & \frac{N-1}{2} \end{bmatrix}^T - \frac{N+1}{2} \alpha Q_N \cdot 1 =$$

$$= \alpha(1 - 2q) \begin{bmatrix} \frac{N-1}{2} & \ldots & \frac{N-1}{2} \end{bmatrix}^T - \frac{N+1}{2} \alpha \cdot 1 \hspace{1cm} (A.3)$$
\[ Q_N \cdot [\alpha \ldots N\alpha]^T = \alpha Q_N \cdot \left[ \begin{array}{ccc} -\frac{N-1}{2} & \ldots & -\frac{N-1}{2} \\ \frac{N+1}{2} & \ldots & \frac{N+1}{2} \end{array} \right] + \frac{N+1}{2} \alpha Q_N \cdot 1 = \] 
\[ = \alpha (1-2q) \left[ -\frac{N-1}{2} \ldots -\frac{N-1}{2} \right] + \frac{N+1}{2} \alpha \cdot 1 \]

REFERENCES:


