entropy

MDPI

Article

Entropy-Constrained Scalar Quantization with a Lossy-Compressed Bit

Melanie F. Pradier [1,2,*], Pablo M. Olmos [1,2,*] and Fernando Perez-Cruz [1,3]

1 Universidad Carlos III de Madrid, Madrid 28911, Spain
2 Gregorio Marañón Health Research Institute, Madrid 28007, Spain
3 Stevens Institute of Technology, Hoboken, NJ 07030, USA; fperezcr@stevens.edu
* Correspondence: melanie@tsc.uc3m.es (M.F.P.); olmos@tsc.uc3m.es (P.M.O.);
Tel.: +34-916-246-005 (M.F.P.); +34-916-249-073 (P.M.O.)

Academic Editor: Raúl Alcaraz Martínez
Received: 8 September 2016; Accepted: 12 December 2016; Published: 16 December 2016

Abstract: We consider the compression of a continuous real-valued source X using scalar quantizers and average squared error distortion D. Using lossless compression of the quantizer's output, Gish and Pierce showed that uniform quantizing yields the smallest output entropy in the limit $D \to 0$, resulting in a rate penalty of 0.255 bits/sample above the Shannon Lower Bound (SLB). We present a scalar quantization scheme named lossy-bit entropy-constrained scalar quantization (Lb-ECSQ) that is able to reduce the $D \to 0$ gap to SLB to 0.251 bits/sample by combining both lossless and binary lossy compression of the quantizer's output. We also study the low-resolution regime and show that Lb-ECSQ significantly outperforms ECSQ in the case of 1-bit quantization.

Keywords: source coding; scalar quantization

1. Introduction

Entropy-constrained scalar quantization (ECSQ) is a well-known compression scheme where a scalar quantizer $q(\cdot)$ is followed by a block lossless entropy-constrained encoder [1,2]. The two main quantities characterizing ECSQ are its distortion D and rate R. For a real-valued input source X, the most common distortion measure is the mean squared error between the source X and its reconstruction $\hat{X}$. As the quantizer $q(\cdot)$ is followed by entropy coding, the rate R is usually defined as the entropy of the random variable at the output of the quantizer, denoted by $q(X)$.

A natural design problem is how to design $q(\cdot)$ to achieve the lowest possible rate with distortion not greater than D. While this problem can be solved numerically with various quantizer optimization algorithms [3–5], the expressions are only known when X follows an exponential [4] or uniform [6] distribution. The asymptotic limit $D \to 0$ constitutes an exception, as it is well known that an infinite-level uniform quantizer is optimal for a broad class of source distributions [1,7,8]. Further, as $D \to 0$, ECSQ with uniform quantizing is only 0.255 bits above Shannon's lower bound (SLB) to the rate distortion function $R(D)$. SLB tends to $R(D)$ as $D \to 0$, and is equal to $R(D)$ for a Gaussian distributed source. Beyond scalar quantization, *vector quantization* (VQ) is the most common option to improve ECSQ; i.e., to achieve rates closer to $R(D)$ at the same distortion level [9].

In this communication, we introduce a scalar quantization scheme that, in the limit $D \to 0$, reduces the gap to Shannon's lower bound to the rate distortion function $R(D)$ to 0.251 bits. Furthermore, we show that in the low-resolution regime (1-bit quantization), the proposed scheme can remarkably improve ECSQ. The main idea of the proposed scheme is to encode the quantizer output by combining both lossless compression and binary lossy compression at a given Hamming distortion D_H, which offers an additional degree of freedom.

The compression scheme is straightforward, as we only need to expand ECSQ with an additional bit that encodes if the source symbol was in the left half or the right half of the quantization region that contained the source symbol. In other words, this scheme codes the least significant quantization bit lossily, allowing a certain Hamming distortion D_H.

We refer to the proposed method as lossy-bit ECSQ (Lb-ECSQ). Note that Lb-ECSQ contains ECSQ as a particular solution, as ECSQ is recovered when the allowed distortion at the least significant quantization bit is set to zero.

The Lb-ECSQ method resembles works in the field of source-channel coding—namely, channel-optimized quantization [10]. Interestingly, when the output of a scalar quantizer is coded and transmitted via a very noisy channel, quantizers with a small number of levels (higher distortion) may yield better performance than those with a larger number of levels (lower distortion) [11]. Several works have addressed the design of scalar quantizers for noisy channels (e.g., [10–12]). All these works present conditions and algorithms to optimize the scalar quantizer given that it is followed by a noisy channel. This is similar to the Lb-ECSQ setup, where the lossy binary encoder behaves like a "noisy channel", with an important and critical difference: in our problem, the distortion introduced by the lossy encoder (the"error probability" of the channel) is a parameter to be optimized, and acts as an additional degree of freedom. Note also that we solely consider the problem of source coding of a continuous source; encoded symbols are transmitted errorless to the receiver that aims at reconstructing the source.

We also study the low-resolution regime, in which we only encode the source with the lossy-bit—namely, 1-bit quantizer followed by a lossy entropy encoder. Results are distribution-dependent for the low-resolution regime, and we focus on the uniform and Gaussian distributions, which are interesting cases that show different behaviors. For example, in this low-resolution regime, the distortion can be reduced by 10% for a uniform distribution when we use 0.2 bits/sample.

In Section 2 of the paper, we review the analysis of ECSQ for an infinite-level uniform quantizer in the limit $D \to 0$. The asymptotic analysis of Lb-ECSQ for the same quantizer and same limit is presented in Section 3. In Section 4, we move to the opposite limit and compare both scalar quantization schemes with 1-bit quantizers.

2. ECSQ and the Uniform Quantizer

Suppose that a source produces the sequence of independent and identically distributed (i.i.d.) real-valued random variables $\{X_k, k \in \mathbb{Z}\}$ according to the distribution $p_X(x)$. A scalar quantizer is defined as a deterministic mapping $q(\cdot)$ from the source alphabet $\mathcal{X} \subseteq \mathbb{R}$ to the reconstruction alphabet $\hat{\mathcal{X}}$, which is assumed to be countable. By Shannon's source coding theorem, $q(X)$ can be losslessly described by a variable-length code whose expected length is roughly equal to its entropy $H(q(X))$. In ECSQ, this quantity constitutes the rate of the quantizer $q(\cdot)$. Additionally, the mean squared-error distortion incurred by the scalar quantizer $q(\cdot)$ is given by

$$\mathbb{E}_X[(X - \hat{X})^2], \tag{1}$$

where $\mathbb{E}_X$ denotes that the expectation is computed w.r.t. the source distribution $p_X(x)$. Consider the set of quantizers $q(\cdot)$ for which the squared distortion in Equation (1) is smaller or equal to $D \in \mathbb{R}_+$, and let $R_s(D)$ be the smallest rate achievable among this set; more precisely

$$R_s(D) \triangleq \inf_{q(\cdot)} H(q(X)) \quad \text{s.t.} \;\; \mathbb{E}_X[(X - q(X))^2] \leq D. \tag{2}$$

Under some constraints on the continuity and decay of $p_X(x)$, Gish and Pierce showed that in the limit $D \to 0$, $R_s(D)$ can asymptotically be achieved by the infinite-level uniform

quantizer, whose quantization regions partition the real line into intervals of equal lengths [1]. Further, they showed that

$$\lim_{D\to 0}\{R_{\mathrm{s}}(D)-R(D)\}=\frac{1}{2}\log_2\frac{\pi \mathrm{e}}{6}, \tag{3}$$

where $R(D)$ is the rate-distortion function of the source [13]. In the rest of this section, we briefly review the asymptotic analysis of ECSQ with uniform quantization following the approach described in [7,8]. We later rely on intermediate results to analyze the Lb-ECSQ scheme for the uniform quantizer. The following conditions are assumed for the source [7,8]:

C1 $p_X(x)\log p_X(x)$ is integrable, ensuring that the differential entropy $h(X)$ is well-defined and finite; and

C2 the integer part of the source X has finite entropy; i.e.,

$$H(\lfloor X\rfloor)<\infty \tag{4}$$

otherwise, $R(D)$ is infinite [14].

Denote the infinite-level uniform quantizer by $q_{\mathrm{u}}(\cdot)$, and let δ be the interval length. For $x\in\mathbb{R}$, we have

$$q_{\mathrm{u}}(x)=\sum_n\left(n+\frac{1}{2}\right)\delta\,\mathbb{1}\left[n\delta<x\le(n+1)\delta\right], \tag{5}$$

where $(n+\frac{1}{2})\delta$ is the reconstruction value for interval n, and $\mathbb{1}[\cdot]$ denotes the indicator function. We define the piecewise-constant probability density function $p_X^{(\delta)}(x)$ as follows:

$$p_X^{(\delta)}(x)=\sum_n\frac{p_n}{\delta}\mathbb{1}\left[n\delta<x\le(n+1)\delta\right], \tag{6}$$

where $p_n\triangleq\int_{n\delta}^{(n+1)\delta}p_X(u)\mathrm{d}u$ is the probability that x belongs to that interval, and $\sum_n p_n=1$. To evaluate the squared error distortion, we first decompose $\mathbb{E}[(X-q_{\mathrm{u}}(X))^2]$ as follows:

$$\begin{aligned}&\mathbb{E}_X[(X-q_{\mathrm{u}}(X))^2]=\textstyle\sum_n\int_{n\delta}^{(n+1)\delta}\left(x-(n+\frac{1}{2})\delta\right)^2p_X(x)\mathrm{d}x\\&=\textstyle\sum_n\frac{p_n}{\delta}\int_{n\delta}^{(n+1)\delta}\left(x-(n+\frac{1}{2})\delta\right)^2\mathrm{d}x-\sum_n\int_{n\delta}^{(n+1)\delta}\left[\frac{p_n}{\delta}-p_X(x)\right]\left(x-(n+\frac{1}{2})\delta\right)^2\mathrm{d}x.\end{aligned}$$

As shown in [7,8], the absolute value of the second term in the above equation can be upper-bounded by $\int\left|p_X^{(\delta)}(x)-p_X(x)\right|\mathrm{d}x$, and this term vanishes as $\delta\to 0$ according to Lebesgue's differentiation theorem and Scheffe's lemma (Th. 16.12) [15]. Thus,

$$\lim_{\delta\to 0}\frac{\mathbb{E}_X[(X-q_{\mathrm{u}}(X))^2]}{\delta^2}=\delta^{-2}\sum_n\frac{p_n}{\delta}\int_{n\delta}^{(n+1)\delta}\left(x-(n+\frac{1}{2})\delta\right)^2\mathrm{d}x=\frac{1}{12}. \tag{7}$$

On the other hand, following [1], we express the entropy of the quantizer's output $H(q_{\mathrm{u}}(X))$ as follows:

$$H(q_{\mathrm{u}}(X))=\int p_X^{(\delta)}(x)\log_2(p_X^{(\delta)}(x))\mathrm{d}x-\log_2(\delta). \tag{8}$$

As shown in [16], the integral in the above expression converges to $h(X)$ as $\delta\to 0$, hence

$$H(q_{\mathrm{u}}(X))=h(X)-\log_2(\delta)+o(1), \tag{9}$$

where $o(1)$ refers to error terms that vanish as δ tends to zero. We conclude that the uniform quantizer $q_{\mathrm{u}}(\cdot)$ with quadratic distortion $D = \mathbb{E}[(X - q_{\mathrm{u}}(X))^2]$ and rate $R_{\mathrm{u}}(D) \triangleq H(q_{\mathrm{u}}(X))$ achieves

$$R_{\mathrm{u}}(D) = h(X) + \frac{1}{2}\log_2\frac{1}{D} - \frac{1}{2}\log_2(12) + o(1) \tag{10}$$

bits per sample, where $o(1)$ comprises error terms that vanish as D tends to zero. Further, for sources X satisfying conditions C1 and C2, the rate-distortion function $R(D)$ can be approximated as [17]

$$R(D) = h(X) + \frac{1}{2}\log_2(\frac{1}{D}) - \frac{1}{2}\log_2(2\pi\mathrm{e}) + o(1). \tag{11}$$

Without the $o(1)$ term, the right-hand side (RHS) of Equation (11) is referred to the *Shannon lower bound* (SLB). By combining Equations (10) and (11), we obtain

$$\lim_{D\to 0}\{R_{\mathrm{u}}(D) - R(D)\} = \frac{1}{2}\log_2(2\pi\mathrm{e}) - \frac{1}{2}\log_2(12) \approx 0.255 \text{ bits/sample} \tag{12}$$

3. Uniform Quantization with a Lossy-Compressed Bit

The above results show that—according to Equation (2)—uniform quantizers are asymptotically optimal as the allowed distortion D vanishes. In the following, we present a simple scheme that—while maintaining the scalar uniform quantizer—reduces the gap to the rate distortion function of the source below Equation (12). To this end, the quantizer's output is compressed using both lossless and lossy compression, and thus the compression rate is no longer measured by the entropy of the quantizer's output. Unlike in [1], we do not claim that uniform quantization is optimal according to the proposed definition of compression rate. Consider again the uniform quantizer $q_{\mathrm{u}}(\cdot)$ with interval length δ. Given X and $q_{\mathrm{u}}(X)$, let $b(X)$ be a binary random variable such that

$$b(x) = \begin{cases} 1, & x \le q_{\mathrm{u}}(x) \\ 0, & x > q_{\mathrm{u}}(x) \end{cases}. \tag{13}$$

3.1. Compression with a Lossy-Compressed Bit

Given the random variable $(q_{\mathrm{u}}(X), b(X))$, we maintain the lossless variable-length encoder to compress $q_{\mathrm{u}}(X)$. Moreover, the binary random variable $b(X)$ is lossy compressed with a certain Hamming distortion D_H, which is a free parameter to be tuned to minimize the squared error distortion. We refer to this compression scheme as ECSQ with a lossy-compressed bit (Lb-ECSQ).

We assume that lossy-compression of $b(X)$ at a Hamming distortion D_H is optimally done, achieving the rate distortion function for a Bernoulli source with probability $P_b \triangleq P(b(X) = 1)$. While this assumption is somewhat unrealistic, our main goal in this paper is to analyze the fundamental limits of the proposed scheme, as one would do in ECSQ when assuming that the scalar quantizer's output is compressed at a rate equal to its entropy. For the actual implementation of Lb-ESCQ, practical schemes based on low-density generator-matrix (LDGM) [18] or lattice codes [19] could be investigated.

Under the assumption of optimal lossy binary compression, we define the Lb-ECSQ rate of the uniform quantizer $q_{\mathrm{u}}(\cdot)$ as

$$R_{\text{Lb-u}}(D, D_H) \triangleq H(q_{\mathrm{u}}(X)) + R(D_H, P_b) = H(q_{\mathrm{u}}(X)) + h_2(P_b) - h_2(D_H), \tag{14}$$

where with a slight abuse of notation we use $R(D_H, P_b)$ to denote the rate distortion function of a Bernoulli source with probability P_b, and $h_2(\cdot)$ is the binary entropy function. We are interested in evaluating Equation (14) in the limit $\delta \to 0$. In this regime, it is straightforward to show that for

any source distribution $p_X(x)$ satisfying C1, then $\lim_{\delta\to 0} P_b = \frac{1}{2}$. Using this result and Equation (9), we have

$$R_{\text{Lb-u}}(D, D_H) = h(X) - \log_2(\delta) + 1 - h_2(D_H) + o(1), \tag{15}$$

where $o(1)$ comprises error terms that vanish as δ tends to zero. Observe that if we take $D_H = 0$ (i.e., lossless compression is used for both $q_{\text{u}}(X)$ and $b(X)$), in the limit $\delta \to 0$ the rate $R_{\text{Lb-u}}$ coincides with the entropy of the uniform quantizer in Equation (9) with half the interval length—i.e., $\delta' = \delta/2$.

3.2. Reconstruction Values and Squared Distortion with a Lossy-Compressed Bit

Since $q_{\text{u}}(X)$ is losslessly compressed, upon decompression, it is recovered with no error. Let $\hat{b}(X)$ be a binary random variable representing the reconstructed value for $b(X)$. Due to the lossy compression at a certain Hamming distortion, there exists a non-zero reconstruction error; namely, $P(\hat{b}(x) \neq b(x)|X = x) > 0$ for $D_H > 0$. Given the pair $(q_{\text{u}}(X), \hat{b}(X))$, we compute the source reconstruction value $\hat{X}$ as follows

$$\hat{x} = q_{\text{u}}(x) + (1 - 2\hat{b}(x))c = \begin{cases} (n + \frac{1}{2})\delta - c & \hat{b}(x) = 1 \\ (n + \frac{1}{2})\delta + c & \hat{b}(x) = 0 \end{cases}, \tag{16}$$

where $c \in [0, \frac{\delta}{2}]$ is a parameter that—along with D_H—will be optimized to minimize the squared error distortion $D = \mathbb{E}_{X,\hat{X}}[(X - \hat{X})^2]$. We note that the reconstruction rule in Equation (16) is possibly suboptimal.

Before evaluating D as a function of δ, D_H, and c, we first need to compute the error probabilities for the $b(X)$ bit. Following [20] (Chapter 10), if $(b(X), \hat{b}(X))$ are jointly distributed according to the binary symmetric channel shown in Figure 1, then the mutual information $I(b(X); \hat{b}(X))$ actually coincides with the Bernoulli rate distortion function $R(D_H, P_b) = h_2(P_b) - h_2(D_H)$. Moreover, by using random coding in [20] (Chapter 10), it is shown that there exist encoding/decoding schemes that asymptotically (in the block-length) meet the input–output distribution in Figure 1. Consequently, under the assumption of optimal lossy compression of $b(X)$ with prior probability P_b, we can compute the error reconstruction probabilities by applying Bayes' rule in Figure 1,

$$\mathrm{P}\left[\hat{b}(X) = 1 | b(X) = 0\right] = \frac{D_H}{1 - P_b} \frac{P_b - D_H}{1 - 2D_H}, \tag{17}$$

$$\mathrm{P}\left[\hat{b}(X) = 0 | b(X) = 1\right] = \frac{D_H}{P_b} \frac{1 - P_b - D_H}{1 - 2D_H}. \tag{18}$$

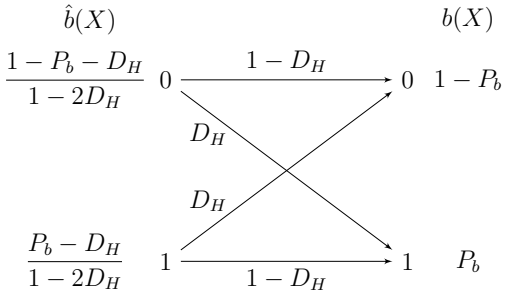

Figure 1. Binary Source Channel model of the joint probability distribution between a Bernoulli source $b(X)$ with prior probability P_b and its reconstruction $\hat{b}(X)$ after lossy compression at Hamming distortion D_H, assuming that the Bernoulli rate distortion function is achieved.

Note that in the limit $\delta \to 0$, we have $P_b = 0.5$, and thus Equations (17) and (18) are equal to D_H.

Lemma 1. *For any source X with a distribution $p_X(x)$ that satisfies conditions C1 and C2,*

$$\lim_{\delta\to 0}\frac{\mathbb{E}_{X,\hat{X}}[(X-\hat{X})^2]}{\delta^2}=\frac{1}{48}\left(1+12D_H-12D_H^2\right) \tag{19}$$

under the assumption that the binary random variable $b(X)$ defined in Equation (13) is optimally lossy compressed at a Hamming distortion D_H.

Proof. Assuming optimal lossy compression, in the limit $\delta\to 0$, $\hat{b}(X)$ is in error with probability D_H. Further, for any $\delta>0$, it is straightforward to check that both Equations (17) and (18) are upper bounded by D_H. Therefore, the squared error distortion can be computed as follows:

$$\mathbb{E}_{X,\hat{X}}[(X-\hat{X})^2]\leq\int\sum_{\hat{x}}(x-\hat{x})^2\,p_{\hat{X}|X=x}(\hat{x})\,p_X(x)\mathrm{d}x, \tag{20}$$

where $p_{\hat{X}|X=x}(\hat{x})$ is the conditional distribution of the reconstruction value for $X=x$, assuming that the reconstruction error probabilities are equal to D_H. Equality is achieved at $\delta=0$. According to Equation (16), $p_{\hat{X}|X=x}(\hat{x})$ can be expressed as follows: for $n\delta<x\leq(n+\frac{1}{2})\delta$, then $b(X)=1$, and hence

$$p_{\hat{X}|X=x}(\hat{x})=\begin{cases}(1-D_H) & \hat{x}=(n+\frac{1}{2})\delta-c\\ D_H & \hat{x}=(n+\frac{1}{2})\delta+c\\ 0 & \text{otherwise}\end{cases} \tag{21}$$

and similarly, if $(n+\frac{1}{2})\delta<x\leq(n+1)\delta$, then $b(x)=0$, and

$$p_{\hat{X}|X=x}(\hat{x})=\begin{cases}(1-D_H) & \hat{x}=(n+\frac{1}{2})\delta+c\\ D_H & \hat{x}=(n+\frac{1}{2})\delta-c\\ 0 & \text{otherwise}\end{cases}. \tag{22}$$

As in Equation (7), we expand the integral in Equation (20) using the piecewise-constant distribution $p_X^{(\delta)}(x)$

$$\begin{aligned}\mathbb{E}_{X,\hat{X}}[(X-\hat{X})^2]\leq&\sum_n\frac{p_n}{\delta}\int_{n\delta}^{(n+1)\delta}\sum_{\hat{x}}(x-\hat{x})^2p_{\hat{X}|X=x}(\hat{x})\,\mathrm{d}x\\&-\sum_n\int_{n\delta}^{(n+1)\delta}\sum_{\hat{x}}\left[\frac{p_n}{\delta}-p_X(x)\right](x-\hat{x})^2p_{\hat{X}|X=x}(\hat{x})\,\mathrm{d}x,\end{aligned} \tag{23}$$

where it can be check that the absolute value of the second term is upper bounded by $\delta^2\int\left|p_X^{(\delta)}(x)-p_X(x)\right|\mathrm{d}x$, which vanishes as $\delta\to 0$.

Using Equations (21) and (22), the first term in Equation (23) reads:

$$\sum_n\frac{p_n}{\delta}\int_{n\delta}^{(n+1)\delta}\sum_{\hat{x}}(x-\hat{x})^2p_{\hat{X}|X=x}(\hat{x})\,\mathrm{d}x=\frac{\delta^2-6\delta r+12r^2-D_H(12\delta r-6\delta^2)}{12}, \tag{24}$$

where $r=\frac{\delta}{2}-c$. The equality is obtained after straight-forward manipulation. The latter expression is minimized if we choose $r=\frac{\delta}{4}(1+2D_H)$, Equation (19) being the corresponding distortion. Note that for $D_H=0$, the reconstruction value is at the center of the interval, $c=\delta/4$. Conversely, if $D_H>0$, the reconstruction point moves closer to the center of the next largest interval, such that the distortion caused by an erroneous transmission is reduced. □

3.3. Asymptotic Gap to the Shannon Lower Bound

The following lemma jointly characterizes the rate $R_{\text{Lb-u}}$ and squared distortion D of the Lb-ECSQ scheme for the uniform quantizer $q_{\text{u}}(\cdot)$ in the limit $D \to 0$:

Lemma 2. *For any source X with a distribution $p_X(x)$ that satisfies conditions C1 and C2, the uniform quantizer $q_u(\cdot)$ with interval length δ and Lb-ECSQ compression with quadratic distortion $D = \mathbb{E}_{X,\hat{X}}[(X-\hat{X})^2]$ and Hamming distortion D_H of the bit $b(X)$ achieves*

$$R_{Lb\text{-}u}(D, D_H) = h(X) + \frac{1}{2}\log_2\frac{1}{D} - \frac{1}{2}\log_2(12) + \Delta(D_H) + o(1), \tag{25}$$

where $o(1)$ comprises error terms that vanish as D tends to zero, and

$$\Delta(D_H) = \frac{1}{2}\log_2(1 + 12D_H - 12D_H^2) - h_2(D_H). \tag{26}$$

Proof. The proof is straightforward by combining Equations (15) and (19). More precisely, from Equation (19), we get that as $\delta \to 0$, the following equality holds

$$\delta = \left(48\frac{\mathbb{E}_{X,\hat{X}}[(X-\hat{X})^2]}{(1 + 12D_H - 12D_H^2)}\right)^{1/2}. \tag{27}$$

By plugging this equality into Equation (15), we get Equation (28), where $D = \mathbb{E}_{X,\hat{X}}[(X-\hat{X})^2]$. □

In Figure 2, we plot $\Delta(D_H)$ for $D_H \in [0, 1/2]$. Observe that $\Delta(D_H)$ is equal to zero at $D_H = 0$ and $D_H = 1/2$. However, for small values of D_H, $\Delta(D_H)$ is actually smaller than zero, achieving its minimum at $D_H^* \approx 3.2 \times 10^{-3}$.

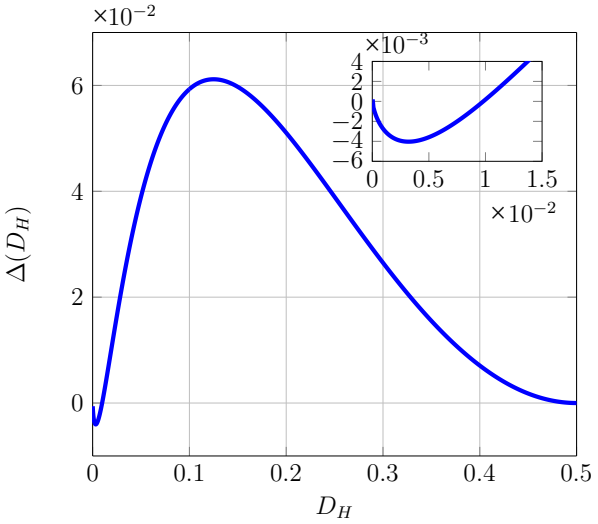

Figure 2. $\Delta(D_H)$ function from Equation (26).

Corollary 1. *The uniform quantizer $q_u(\cdot)$ with interval length δ and Lb-ECSQ compression with quadratic distortion $D = \mathbb{E}[(X-\hat{X})^2]$ achieves*

$$R_{Lb\text{-}u}(D, D_H^*) = h(X) + \frac{1}{2}\log_2\frac{1}{D} - \frac{1}{2}\log_2(12) - \Delta(D_H^*) + o(1), \tag{28}$$

bits/sample, where $\Delta(D_H^*) \approx 0.004$.

Finally, by combining Equations (11) and (28),

$$\lim_{D\to 0}\{R_{\text{Lb-u}}(D, D_H^*) - R(D)\} = \frac{1}{2}\log_2(2\pi\text{e}) - \frac{1}{2}\log_2(12) - \Delta(D*_H) \approx 0.251 \tag{29}$$

bits/sample, which proves that Lb-ECSQ is able to outperform ECSQ in the limit $D \to 0$ using the uniform quantizer $q_{\text{u}}(\cdot)$.

4. Lb-ECSQ in the High Distortion Regime

The above results demonstrate that the use of lossy compression can reduce the gap to SLB in the limit $D \to 0$ with respect to ECSQ. Improvements can also be observed for low-to-moderate compression rates. The case of a quantizer $q(\cdot)$ with only two quantization levels plays a special role that we analyze in this section. While the extension to an arbitrary number N of quantization levels is interesting, preliminary results show that the biggest gain is achieved for a 2-level quantizer, and that the Lb-ECSQ performance tend with N very quickly to the asymptotic gain ($N \to \infty$) described in the previous section. We consider a two-level quantizer $q(\cdot)$ with quantization regions $A_1 = \{x : x \leq \alpha\}$ and $A_2 = \{x : x > \alpha\}$ for some $\alpha \in \mathbb{R}$ and two possible source distributions, $X \sim U \in [-\delta/2, \delta/2]$ and $X \sim \mathcal{N}(0, \sigma^2)$.

4.1. Two-Level Quantization of a Uniform Source

For the uniform source $X \sim U \in [-\delta/2, \delta/2]$, the ECSQ rate is

$$R_q \triangleq H(q(X)) = h_2(\frac{\alpha'}{\delta}), \tag{30}$$

where $\alpha' = \alpha + \delta/2$ and $\alpha \in [-\delta/2, \delta/2]$. The squared distortion is minimized if reconstruction points are placed at the center of each quantization region [6]; i.e., $q(x) = \alpha/2 - \delta/4$ if $x \in A_1$ and $q(x) = \alpha/2 + \delta/4$ if $x \in A_2$, and the distortion incurred is

$$\mathbb{E}_X[(X - q(X))^2] = \frac{1}{\delta}\left(\frac{\alpha'^3}{12} + \frac{(\delta - \alpha'^3)}{12}\right). \tag{31}$$

In Figure 3, we plot $\mathbb{E}[(X - q(X))^2]$ vs. R_q as we vary $\alpha' \in [0, \delta]$ for $\delta = \sqrt{12}$ (red curve with $\circ$ marker). As presented in Section 3, Lb-ECSQ combines lossless compression of $q(X)$ with lossy compression of a random variable $b(X)$ that gathers additional information of the source X within the quantization region. As now $q(\cdot)$ only partitions the real line in two quantization regions, we implement Lb-ECSQ by directly lossy compressing the quantizer's output $q(X)$. To this end, we define the binary R.V. $b(X) = 1$ if $x \in A_1$ and zero-otherwise. The Lb-ECSQ rate is given by the compression rate of $b(X)$ at a certain Hamming distortion D_H:

$$R_{\text{Lb-}q} \triangleq h_2(\frac{\alpha'}{\delta}) - h_2(D_H). \tag{32}$$

Further, we fix the quantizer threshold to $\alpha = 0$, which implies that $q(X)$ takes value either $-\frac{\delta}{4}$ or $\frac{\delta}{4}$ with uniform probability, and thus $R_{\text{Lb-}q} = 1 - h_2(D_H)$. Under optimal lossy compression, the reconstructed bit $\hat{b}(X)$ is in error with the same probability model described in Equations (17) and (18); namely, $\hat{b}(X)$ is in error with probability D_H. We set the source reconstruction $\hat{X} = -c$ if $\hat{b}(X) = 1$ and $\hat{X} = c$ if $\hat{b}(X) = 0$, where c is a positive quantity optimized to minimize $\mathbb{E}[(X - \hat{X})^2]$.

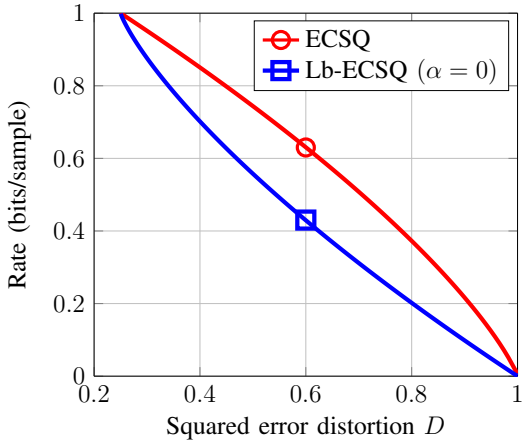

Figure 3. Entropy-constrained scalar quantization (ECSQ) and lossy-bit ECSQ (Lb-ECSQ) rate distortion function for a 1-bit quantizer and uniform source, $X \sim U \in [-\delta/2, \delta/2]$ for $\delta = \sqrt{12}$.

Lemma 3. *Given the source $X \sim U \in [-\delta/2, \delta/2]$, the 1-bit quantizer $q(\cdot)$ with threshold $\alpha = 0$ and Lb-ECSQ compression achieves a squared distortion*

$$\mathbb{E}_{X,\hat{X}}[(X - q(X))^2] = \frac{\delta^2}{48}(1 + 12D_H - 12D_H^2) \tag{33}$$

under the assumption that $q(X)$ is optimally lossy compressed at a Hamming distortion D_H.

Proof. The proof is similar to that of Lemma 1, expanding $\mathbb{E}[(X - \hat{X})^2]$ as done for every quantization region done in and minimizing w.r.t. the reconstruction point c. □

In Figure 3, we plot $\mathbb{E}_{X,\hat{X}}[(X - \hat{X}))^2]$ in Equation (37) vs. $R_{\text{Lb-}q}$ in Equation (32) for $\delta = \sqrt{12}$ as we vary $D_H \in [0, 1/2]$ (blue curve with □ marker). Observe that Lb-ECSQ improves ECSQ at all points, except for $D_H = 0$ and $D_H = 1/2$, as we know they must be equivalent at these two points. The Lb-ECSQ analysis proposed for $\alpha = 0$ can be generalized to an arbitrary threshold $\alpha \in [-\delta/2, \delta/2]$, but simulations for $\alpha \neq 0$ using numerical optimization show that the obtained rate-distortion function coincides with the one computed for $\alpha = 0$. This result is dependent on the source distribution, as shown for the Gaussian source case.

4.2. Two-Level Quantization of a Gaussian Source

Now consider the same quantizer $q(\cdot)$ and $X \sim \mathcal{N}(0, \sigma^2)$. Low-resolution ECSQ for a Gaussian input source was studied in [21], where the authors showed that the minimum rate is achieved by a quantizer whose unique threshold α goes either to $-\infty$ or to ∞ as $D \to \sigma^2$, and the two reconstruction points are the centroids of the quantization regions. The ECSQ rate distortion function for this source is given by the following parametric curve

$$R_q = h_2(\Phi(\alpha)), \tag{34}$$

$$\mathbb{E}_X[(X - q(X))^2] = \int_{-\infty}^{\alpha} p_X(x)(x - c_1(\alpha))^2 \mathrm{d}x + \int_{\alpha}^{\infty} p_X(x)(x - c_2(\alpha))^2 \mathrm{d}x, \tag{35}$$

where $\Phi(\alpha)$ is the cumulative density function of the Gaussian distribution, and

$$c_1(\alpha) = \frac{1}{\Phi(\alpha)} \int_{-\infty}^{\alpha} x\, p_X(x) \mathrm{d}x, \quad c_2(\alpha) = \frac{1}{1 - \Phi(\alpha)} \int_{\alpha}^{\infty} x\, p_X(x) \mathrm{d}x. \tag{36}$$

We now study the same Lb-ECSQ scheme analyzed before for the uniform source. First, we fix the quantizer threshold to $\alpha = 0$ and define $b(X) = 1$ if $X \leq \alpha$, and zero otherwise. Note that Lb-ECSQ rate is given in Equation (32).

Lemma 4. *Given the source $X \sim \mathcal{N}(0, \sigma^2)$, the quantizer $q(\cdot)$ with $\alpha = 0$ and Lb-ECSQ compression achieves a squared distortion*

$$\mathbb{E}_{X,\hat{X}}[(X - \hat{X})^2] = \frac{1}{\sigma} - \frac{2}{\sigma\pi}(1 + 4D_H^2 - 4D_H) \tag{37}$$

for $D_H \in [0, 1/2]$.

Proof. The proof is based on expanding $\mathbb{E}[(X - \hat{X})^2]$ as done for every quantization region in Equation (24) and minimizing w.r.t. the reconstruction point c. □

In Figure 4, we plot the gap between the ECSQ and Lb-ECSQ rate distortion function for a 1-bit quantizer and the rate distortion function for the source; i.e., $R(D) = 0.5 \log_2(\sigma^2/D)$. Observe that, unlike the case of a uniform source, for $D/\sigma^2 \to 1$, Lb-ECSQ is slightly worse than ECSQ. As discussed before, in the ECSQ solution for a Gaussian input source, the threshold α goes to infinity in the limit $D/\sigma^2 \to 1$ [21]. By fixing the threshold α to 0 in Lb-ECSQ, we are restraining to an equivalent solution. This can be tackled by generalizing the above equations to an arbitrary threshold α. While the methodology is equivalent, we have to rely on numerical optimization to find the optimal choice of α, $c_1(\alpha)$, and $c_2(\alpha)$ for each value of D_H. In this case, the bit error reconstruction probabilities take the form given in Equations (17) and (18). Additionally, for an arbitrary threshold α, $b(X)$ is a Bernoulli source with probability $p = \Phi(\alpha)$, and hence the compression rate is $R_{\text{Lb-}q} = h_2(p) - h_2(D_H)$. A numerical optimization (gradient descend) procedure has been used to find the minimum distortion for each $R_{\text{Lb-}q}$. The results are shown in Figure 4, where we can see that now Lb-ECSQ is able to perform equally to or better than ECSQ in the whole range.

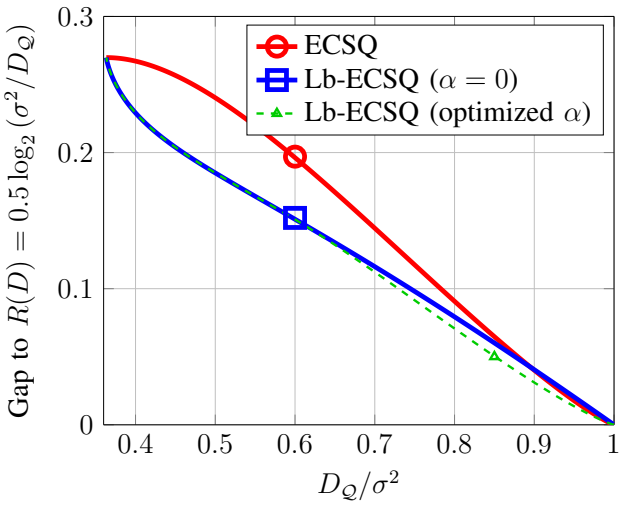

Figure 4. For $X \sim \mathcal{N}(0, \sigma^2)$, we plot the gap between the ECSQ and Lb-ECSQ rate distortion function for a 1-bit quantizer and the rate distortion function for the source; i.e., $R(D) = 0.5 \log_2(\sigma^2/D)$.

Acknowledgments: The authors wish to thank Tobias Koch and Gonzalo Vázquez Vilar for fruitful discussions and helpful comments to the manuscript. This work has been supported in part by the European Union 7th Framework Programme through the Marie Curie Initial Training Network "Machine Learning for Personalized Medicine" MLPM2012, Grant No. 316861, by the Spanish Ministry of Economy and Competitiveness and Ministry of Education under grants TEC2016-78434-C3-3-R (MINECO/FEDER, EU) and IJCI-2014-19150, and by Comunidad de Madrid (project 'CASI-CAM-CM', id. S2013/ICE-2845).

Author Contributions: All authors have contributed equally to conceive and design the experiments, to analyze the data, and to write the paper. All authors have read and approved the final manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Gish, H.; Pierce, J. Asymptotically efficient quantizing. *IEEE Trans. Inf. Theory* **1968**, *14*, 676–683.
2. Goblick, T.; Holsinger, J. Analog source digitization: A comparison of theory and practice (Corresp.). *IEEE Trans. Inf. Theory* **1967**, *13*, 323–326.
3. Farvardin, N.; Modestino, J. Optimum quantizer performance for a class of non-Gaussian memoryless sources. *IEEE Trans. Inf. Theory* **1984**, *30*, 485–497.
4. Sullivan, G. Efficient scalar quantization of exponential and Laplacian random variables. *IEEE Trans. Inf. Theory* **1996**, *42*, 1365–1374.
5. Noll, P.; Zelinski, R. Bounds on Quantizer Performance in the Low Bit-Rate Region. *IEEE Trans. Inf. Theory* **1978**, *26*, 300–304.
6. Gyorgy, A.; Linder, T. Optimal entropy-constrained scalar quantization of a uniform source. *IEEE Trans. Inf. Theory* **2000**, *46*, 2704–2711.
7. Linder, T.; Zeger, K. Asymptotic entropy-constrained performance of tessellating and universal randomized lattice quantization. *IEEE Trans. Inf. Theory* **1994**, *40*, 575–579.
8. Koch, T.; Vazquez-Vilar, G. Rate-Distortion Bounds for High-Resolution Vector Quantization via Gibbs's Inequality. *arxiv* **2015**, arXiv:1507.08349.
9. Chou, P.A.; Lookabaugh, T.; Gray, R.M. Entropy-constrained vector quantization. *IEEE Trans. Acoust. Speech Signal Process.* **1989**, *37*, 31–42.
10. Farvardin, N.; Vaishampayan, V. Optimal quantizer design for noisy channels: An approach to combined source-channel coding. *IEEE Trans. Inf. Theory* **1987**, *33*, 827–838.
11. Spilker, J.J. *Digital Communications by Satellite*; Prentice-Hall: Upper Saddle River, NJ, USA, 1977.
12. Kurtenbach, A.J.; Wintz, P.A. Quantizing for noisy channels. *IEEE Trans. Inf. Theory* **1969**, *17*, 291–302.
13. Shannon, C.E. A mathematical theory of communication. *Bell Syst. Tech. J.* **1948**, *27*, 379–423.
14. Koch, T. The shannon lower bound is asymptotically tight for sources with finite renyi information dimension. *IEEE Trans. Inf. Theory* **2015**, doi:10.1109/TIT.2016.2604254.
15. Billingsley, P. *Convergence of Probability Measures*; John Wiley: New York, NY, USA, 1968.
16. Rényi, A. *Probability Theory*; North-Holland Series in Applied Mathematics and Mechanics; Elsevier: Budapest, Hungary, 1970.
17. Linder, T.; Zamir, R. On the asymptotic tightness of the Shannon lower bound. *IEEE Trans. Inf. Theory* **1994**, *40*, 2026–2031.
18. Aref, V.; Macris, N.; Vuffray, M. Approaching the rate-distortion limit with spatial coupling, belief propagation, and decimation. *IEEE Trans. Inf. Theory* **2015**, *61*, 3954–3979.
19. Calderbank, A.R.; Fishburn, P.C.; Rabinovich, A. Covering properties of convolutional codes and associated lattices. *IEEE Trans. Inf. Theory* **1995**, *41*, 732–746.
20. Cover, T.M.; Thomas, J.A. *Elements of Information Theory*; Wiley Series in Telecommunications and Signal Processing; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2006.
21. Marco, D.; Neuhoff, D. Low-resolution scalar quantization for Gaussian sources and squared error. *IEEE Trans. Inf. Theory* **2006**, *52*, 1689–1697.