Simple Compression Algorithm for Memoryless Laplacian Source Based on the Optimal Companding Technique

Zoran H. PERIĆ[1], Marko D. PETKOVIĆ[2], Milan R. DINČIĆ[1]

[1]*Faculty of Electronic Engineering, Universsity of Nis*
Aleksandra Medvedeva 14, Nis, Serbia
e-mail: peric@elfak.ni.ac.yu

[2]*Faculty of Sciences and Mathematics, Universsity of Nis*
Višegradska 33, Nis, Serbia
e-mail: dexterofnis@gmail.com

Abstract. This paper has two achievements. The first aim of this paper is optimization of the lossy compression coder realized as companding quantizer with optimal compression law. This optimization is achieved by optimizing maximal amplitude for that optimal companding quantizer for Laplacian source. Approximate expression in closed form for optimal maximal amplitude is found. Although this expression is very simple and suitable for practical implementation, it satisfy optimality criterion for Lloyd–Max quantizer (for R >= 6 bits/sample). In the second part of this paper novel simple lossless compression method is presented. This method is much simpler than Huffman method, but it gives better results. Finally, at the end of the paper, we join optimal companding quantizer and lossless coding method together in one generalized compression method. This method is applied on the concrete still image and good results are obtained. Besides still images, this method also could be used for compression speech and bio-medical signals.

Keywords: simple lossless compression algorithm, companding quantization, optimal maximal amplitude.

1. Introduction

Quantizers play an important role in every A/D conversion system. They are applied for the purpose of storage and transmission of continual signals. A vast of research was done in this topic. An efficient algorithm for design of the optimal quantizer for the source with known distribution was developed by Lloyd and Max (Max, 1960). This is an iterative method which gives the sequence of quantizers converging to the optimal quantizer.

Lloyd–Max algorithm is efficient for the small number N of quantization points (usually for $N \leqslant 32$). However, for the large number of quantization points, it is time consuming. Also the realization complexity of the optimal quantizers with many quantization levels is very high. One solution which overcomes these difficulties is the companding model (Jayant and Noll, 1984, Chapter 4, pp. 129–139; Gray, 2004). Quantizers based on

the companding model (also known as companding quantizers) have simple realization structure and performances close to the optimal. Also the design of such quantizers is more efficient than Lloyd–Max algorithm since it does not require the iterative method. This difference is very notable for some commonly used sources including the Laplacian source. We will consider companding quantizer, because of less complexity in design and implementation.

Memoryless Laplacian source is commonly used model in many applications, due to its simplicity and fact that many parameters and characteristics can be found as the closed form relations (including the optimal compression function). Examples are speech signal, images, video signal, bio-medical images (for example, computed tomography scanner), etc. For transmission, processing and storing that signals, simple and fast compression algorithms are desirable. One solution is given in (Starosolsky, 2007) for uniform quantizer. In this paper we give simpler solution for non-uniform quantizer.

In paper (Ramalho, 2002) lossless compression algorithm was given that provided only the additional compression of the digitized signal (PCM), but without providing a quality improvement. Our algorithm gives quality improvement as well as further compression. In paper (Nikolic and Peric, 2008) lossy compression and Lloyd–Max quantizer are applied. In our paper simple algorithm for lossy and lossless compression is given. For lossy compression simpler quantizer is used without deterioration of performances. Using this algorithm for speech signal compression, improvement of 0.33 bits/sample is achieved in relation to algorithm used in (Nikolic and Peric, 2008).

In this paper we want to accomplish two goals. Our first goal is to optimize performances of the companding quantizer for memoryless Laplacian source by optimizing maximal amplitude. In the paper (Na and Neuhoff, 2001) it is given one heuristic solution for the support region of the optimal quantizer. Also paper (Na, 2004) deals with the upper bound of this support region. The heuristic solution for the maximal signal amplitude in the companding model is given in (Nikolic and Peric, 2008). An intension of this paper is to generalize these results by giving several approximate analytical expressions for the maximal signal amplitude and its numerical computation. As it is shown, one of the analytical expressions almost coincides with the optimal value.

The second goal of this paper is the construction of the lossless coder for Laplacian source with simple realization structure. The well-known efficient algorithm for lossless coding of the information sources with known symbol probability is Huffman algorithm (Hankerson *et al.*, 2004; Sayood, 2006). It requires very complex realization structure and also is time consuming. Hence we give the simple coding algorithm for M-ary extended information source and give the optimization of the extension order M. Our algorithm has better performances (smaller bit-rate) and drastically simpler realization complexity than Huffman algorithm. Our algorithm is also simpler than algorithm proposed in (Starosolsky, 2007), because we have 1 bit in prefix and $(r-1)$bits or r bits in suffix, that is, we have constant number of bits in prefix. On the other hand, algorithm in (Starosolsky, 2007) uses variable number of bits in prefix, which requires much complex coder, and especially decoder**.** Prefix defines region where symbol of M-ary source extension falls (M is a number of samples in frame, and we can consider frame as M-ary source extension). Suffix represents codeword for symbol of M-ary source extension.

Finally, we have joined optimal companding quantizer and simple lossless coding in a generalized compression method. This generalized method is applied on a still image and good results are obtained, as it shown in the paper. Beside still images, this compression method can be applied on speech and bio-medical signals.

This paper is organized as follows. Section 2 recalls some basic theory of quantizers and companding model. In Section 3, optimization of companding quantizer was done by finding optimal value for the maximal signal amplitude $t_{\max}$. We gave three approximate expressions in closed form for $t_{\max}$ and also one numerical method for finding $t_{\max}$ which is based on minimization of approximate expression for distortion. In Section 4, we compare the approximations with numerically obtained maximal signal amplitude. Section 5 deals with the construction and performance analysis of the simple lossless coding algorithm. In Section 6 we gave generalized compression algorithm which represents combination of lossy and lossless compression. In this section application of this generalized method on a still image is presented. Section 7 concludes the paper.

2. Fixed Rate Scalar Quantizers and Companding Technique

An N-point, fixed rate, scalar quantizer is characterized by the set of real numbers $t_1, t_2, \ldots, t_{N-1}$, called *decision thresholds*, which satisfy

$$-\infty = t_0 < t_1 < \ldots < t_{N-1} < t_N = +\infty,$$

and set $y_1, \ldots, y_N$, called *representation levels*, which satisfy $y_j \in \alpha_j = (t_{j-1}, t_j]$, for $j = 1, \ldots, N$. Sets $\alpha_1, \alpha_2, \ldots, \alpha_N$ form the partition of the set of real numbers R and are called *quantization cells*. The quantizer is defined as many-to-one mapping $Q: R \to R$ defined by $Q(x) = y_j$ where $x \in \alpha_j$. In practice, input signal value x is discretized (quantized) to the value y_j. Cells $\alpha_2, \ldots, \alpha_{N-1}$ are called *inner cells* (or *granular cells*) while α_1 and α_N are *outer cells* (or *overload cells*). In such way, cells $\alpha_2, \ldots, \alpha_{N-1}$ form granular while cells α_1 and α_N form an overload region. As fixed-rate and scalar are only types of quantizers considered in the paper, for brevity we will ordinarily omit these adjectives.

Suppose that an input signal is characterized by continuous random variable X with probability density function (pdf) $p(x)$. In the rest of the paper we will suppose that information source is Laplacian source with memoryless property and zero mean value. The pdf of this source is given by

$$p(x) = \frac{1}{\sqrt{2\sigma^2}} e^{-\frac{|x|\sqrt{2}}{\sigma}}.$$

The sources with exponential and Laplacian pdf are commonly encountered and the methods for designing quantizers for these sources are very similar. Without loosing of generality we can suppose that $\sigma = 1$ and last expression becomes

$$p(x) = \frac{1}{\sqrt{2}} e^{-\sqrt{2}|x|}.$$

Since $p(x)$ is an even function, an optimal quantizer Q must satisfy $t_i = -t_{N-i}$ and $y_i = -y_{N-i}$ for arbitrary $i = 1, \ldots, N$.

The quality of the quantizer is measured by distortion of resulting reproduction in comparison to the original. Mostly used measure of distortion is mean-squared error. It is defined as

$$D(Q) = E(X - Q(x))^2 = \sum_{i=1}^{N} \int_{t_{i-1}}^{t_i} (x - y_i)^2 p(x) \, dx.$$

The N-point quantizer Q is *optimal* for the source X if there is no other N-point quantizer Q_1 such that $D(Q_1) < D(Q)$. We also define granular $D_g(Q)$ and overload $D_{ol}(Q)$ distortion by

$$D_g(Q) = \sum_{j=2}^{N-1} \int_{t_{j-1}}^{t_j} (x - y_j)^2 p(x) \, dx,$$

$$D_{ol}(Q) = \int_{-\infty}^{t_1} (x - y_1)^2 p(x) \, dx + \int_{t_{N-1}}^{+\infty} (x - y_N)^2 p(x) \, dx. \tag{1}$$

Obviously holds $D(Q) = D_g(Q) + D_{ol}(Q)$. Denote by D_N^* distortion of an optimal N-point quantizer. Considerable amount of work has been focused on the design of optimal quantizers for compression sources in image, speech, and other applications. As first discovered by Panter and Dite (1951), for large N holds $D_N^* = c_\infty / N^2$. Here c_∞ is the Panter–Dite constant

$$c_\infty = \frac{1}{12} \left(\int_{-\infty}^{+\infty} p^{1/3}(x) \, dx \right)^3.$$

The general method for design of an optimal N-point quantizer for the given source X is Lloyd–Max algorithm (Max, 1960; Jayant and Noll, 1984, Chapter 4, pp. 129–139). This is iterative process which starts from some initial quantizer $Q^{(0)}$ and constructs the sequence $Q^{(n)}$, $n = 1, 2, \ldots$, of quantizers which converge to the optimal quantizer Q^* (values of thresholds $t_i^{(n)}$ converge to the thresholds t_i^* of the optimal quantizer Q^*). Due to the computational complexity of this method, it is not suitable for the design of optimal quantizers with more than 128 levels. Hence there are developed other methods for construction of nearly optimal quantizers for large number of quantization levels.

One of the such techniques which is commonly used is companding technique (Judell and Scharf, 1986). It forms the core of the ITU-T G.711 PCM standard, recommended for coding speech signals. Companding technique consists of the following steps:

- compress the input signal x by applying the compressor function $c(x)$;
- apply the uniform quantizer Q_u on the compressed signal $c(x)$;
- expand the quantized version of the compressed signal using an inverse compressor function $c^{-1}(x)$.

The corresponding structure of a nonuniform quantizer consisting of a compressor, a uniform quantizer, and expandor in cascade is called *compandor*.

Optimal compression law is used in this paper. Compression function for that law in dependence of pdf is known, but optimal maximal amplitude for that compression law hasn't determined up to now. In this paper this optimal maximal amplitude for optimal compression law will be determined. There are several ways to define the compressor function for optimal compression law. Originally, in (Judell and Scharf, 1986) and also in (Na, 2004) it is the function $c\colon R \to (-1, 1)$ defined by

$$c_0(x) = \frac{\int_{-\infty}^{x} p^{1/3}(x)\,\mathrm{d}x}{\int_{-\infty}^{+\infty} p^{1/3}(x)\,\mathrm{d}x}.$$

We will use the similar definition as above

$$c(x; t_{\max}) = \begin{cases} -1, & x < -t_{\max}, \\ -1 + 2\dfrac{\int_{-t_{\max}}^{x} p^{1/3}(x)\,\mathrm{d}x}{\int_{-t_{\max}}^{t_{\max}} p^{1/3}(x)\,\mathrm{d}x}, & x \in [-t_{\max}, t_{\max}], \\ 1, & x > t_{\max}. \end{cases} \tag{2}$$

Value $t_{\max}$ is called *maximal signal amplitude*. Since an image of the function $c(x; t_{\max})$ is $(-1, 1)$, decision thresholds $t_{u,i}$ and representation levels $y_{u,i}$ of the uniform quantizer are defined by

$$t_{u,i} = -1 + \frac{2i}{N}, \ \ (i = 0, 1, \ldots, N), \quad y_{u,i} = -1 + \frac{2i-1}{N}, \ \ i = 1, 2, \ldots, N.$$

Hence, decision thresholds and representation levels of equivalent non-uniform quantizer $Q(x) = c^{-1}(Q_u(c(x)))$ can be determined as the solutions of the following equations

$$c(t_i; t_{\max}) = t_{u,i}, \quad c(y_i; t_{\max}) = y_{u,i}. \tag{3}$$

By solving the last equations we obtain the values t_i and y_i as the functions of the maximal amplitude $t_{\max}$

$$t_i = \frac{3}{\sqrt{2}} \log\left(\frac{2i + (N - 2i)\exp\left(-\frac{\sqrt{2}}{3}t_{\max}\right)}{N} \right), \quad 0 \leqslant i \leqslant N/2,$$

$$t_i = \frac{3}{\sqrt{2}} \log\left(\frac{N}{2N - 2i + (2i - N)\exp\left(-\frac{\sqrt{2}}{3}t_{\max}\right)} \right), \quad N/2 < i \leqslant N,$$

$$y_i = \frac{3}{\sqrt{2}} \log\left(\frac{2i - 1 + (N - 2i + 1)\exp\left(-\frac{\sqrt{2}}{3}t_{\max}\right)}{N} \right), \quad 1 \leqslant i \leqslant N/2,$$

$$y_i = \frac{3}{\sqrt{2}} \log\left(\frac{N}{2N - 2i + 1 + (2i - 1 - N)\exp\left(-\frac{\sqrt{2}}{3}t_{\max}\right)} \right), \quad N/2 < i \leqslant N, \tag{4}$$

where $\log x$ represents natural logarithm of x.

Expressions for t_{N-1} and y_N will be used in further consideration and hence we will give them explicitly

$$t_{N-1} = \frac{3}{\sqrt{2}} \log \left(\frac{N}{2 + (N-2) \exp\left(-\frac{\sqrt{2}t_{\max}}{3}\right)} \right), \tag{5}$$

$$y_N = \frac{3}{\sqrt{2}} \log \left(\frac{N}{1 + (N-1) \exp\left(-\frac{\sqrt{2}t_{\max}}{3}\right)} \right). \tag{6}$$

For the ease of notation, we will denote the distortion, granular distortion and overload distortion by D, D_g and D_{ol} respectively (we will omit the quantizer Q). Using (1) we obtain that distortion D is function of one parameter $t_{\max}$. Same holds also for D_g and D_{ol}. Our goal is to find the optimal value $t_{\max}^{comp}$ which minimizes distortion D. One approximate solution is given in (Peric *et al.*, 2007).

3. Estimation of the Optimal Maximal Signal Amplitude

Closed-form expression for the exact distortion $D = D(t_{\max})$, as the function of the maximal signal amplitude $D = D(t_{\max})$ can be obtained by combining relations (1) and (4). Obtained functional dependence is very complicated, since $t_{\max}$ appears in bounds of integrals and also in functions under integrations. Hence, the exact minimization of $D(t_{\max})$ is very complex (see Section 4), our first goal in this paper is to find some approximate closed-form expressions for optimal maximal amplitude $t_{\max}$.

According to the (Peric and Nikolic, 2007), granular distortion D_g can be approximated by the value $\tilde{D}_g$ as follows

$$\tilde{D}_g = \frac{1}{12(N-2)^2} \left(\int_{-t_{N-1}}^{t_{N-1}} p^{1/3}(x)\,\mathrm{d}x \right)^3$$
$$= \frac{9}{2(N-2)^2} \left(1 - \exp\left(-\frac{\sqrt{2}t_{N-1}}{3}\right) \right)^3. \tag{7}$$

On the other side, we can explicitly compute overload distortion by

$$D_{ol} = 2 \int_{t_{N-1}}^{+\infty} (x - y_N)^2 p(x)\,\mathrm{d}x$$
$$= \mathrm{e}^{-\sqrt{2}t_{N-1}} \left(1 + \sqrt{2}t_{N-1} + t_{N-1}^2 - \sqrt{2}y_N - 2t_{N-1}y_N + y_N^2 \right). \tag{8}$$

Hence we will consider an approximate distortion $\tilde{D} = \tilde{D}_g + D_{ol}$. It is worth mentioning that bounds of the integral (7) are usually set to $-t_{\max}$ and $t_{\max}$ respectively, and therefore maximum signal amplitude is equal to the support region. Maximal error in this approximation is obtained on the interval $(t_{N-1}, t_{\max})$. Therefore, in order to improve the approximation we set the corresponding bound to t_{N-1}, and therefore lower bound of integral in (8) is also t_{N-1}.

Value $\tilde{D}$ is the function of t_{N-1} and y_N and since these values are both functions of $t_{\max}$ (expressions (5) and (6)), also $\tilde{D}$ can be expressed as the function of $t_{\max}$. Note that approximation of D by $\tilde{D}$ is very accurate, especially for the large values of N.

On Figs. 1a, 1b, 1c we show graphs of D and $\tilde{D}$ for $N = 128$, $N = 64$ and $N = 32$ as the functions of $t_{\max}$, and graph of relative error $\frac{D-\tilde{D}}{D}$.

We will give three estimate expressions for the optimal maximal signal amplitude $t_{\max}$.

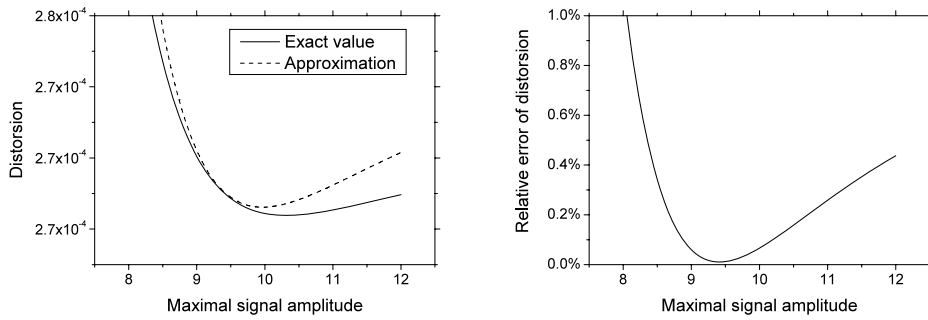The first estimation is based on the fact that the value of the last representation

Fig. 1a. Graphs of D and $\tilde{D}$ (left) and relative error (right) as the function of $t_{\max}$ for $N = 128$.
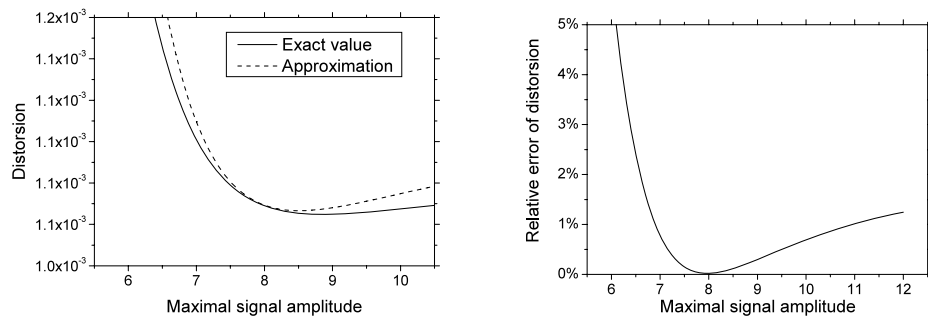

Fig. 1b. Graphs of D and $\tilde{D}$ (left) and relative error (right) as the function of $t_{\max}$ for $N = 64$.
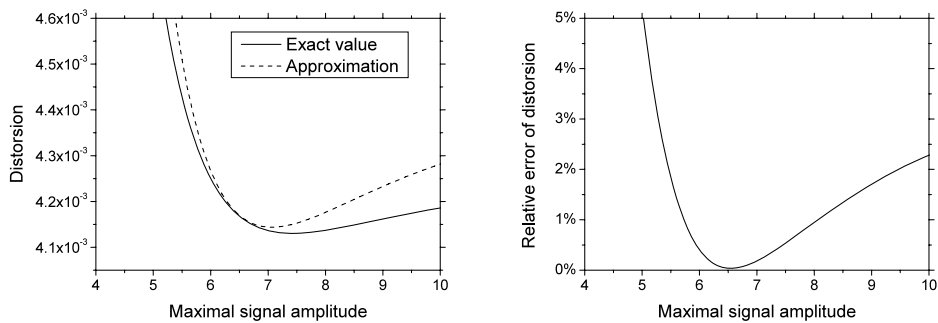

Fig. 1c. Graphs of D and $\tilde{D}$ (left) and relative error (right) as the function of $t_{\max}$ for $N = 32$.

level y_N, as determined by (6) is not optimal in general. For fixed t_{N-1}, an optimal value of y_N is given by the following expression

$$y_N^{opt} = \frac{\int_{t_{N-1}}^{+\infty} xp(x)\,dx}{\int_{t_{N-1}}^{+\infty} p(x)\,dx} = t_{N-1} + \frac{1}{\sqrt{2}}. \tag{9}$$

On the other hand, for the fixed y_N, value of t_{N-1} minimizing $\tilde{D}$ is obtained as the solution of the following equation:

$$\frac{\partial \tilde{D}}{\partial t_{N-1}} = 0. \tag{10}$$

The solution of the last equation is given by

$$\hat{t}_{N-1} = \frac{3}{\sqrt{2}} \log\left(\frac{N+1}{3}\right). \tag{11}$$

Value $t_{\max}$ is determined such that next condition holds

$$\hat{t}_{N-1} + \frac{1}{\sqrt{2}} = y_N. \tag{12}$$

Using (11) and (6) we obtain the following equation in terms of $t_{\max}$

$$\frac{3}{\sqrt{2}} \log\left(\frac{N+1}{3}\right) + \frac{1}{\sqrt{2}} = \frac{3}{\sqrt{2}} \log\left(\frac{N}{1 + (N-1)\exp\left(-\frac{\sqrt{2}t_{\max}}{3}\right)}\right)$$

whose solution is given by

$$t_{\max}^{(1)} = \frac{3}{\sqrt{2}} \log\left(\frac{N-1}{\frac{3N}{N+1}\exp(-1/3) - 1}\right). \tag{13}$$

In the second approach we consider the similar condition as (12). Namely, instead of $\hat{t}_{N-1}$ we will put the value $\bar{t}_{N-1} = \frac{3}{\sqrt{2}} \log(\frac{N}{2})$ given by Na (2004) as the upper bound of the support region. In such way we obtain the following estimation of the maximal signal amplitude

$$t_{\max}^{(2)} = \frac{3}{\sqrt{2}} \log\left(\frac{N-1}{2\exp(-1/3) - 1}\right). \tag{14}$$

Finally, in the third approach, we will search for the value of $t_{\max}$ such that $\hat{t}_{N-1} = t_{N-1}$. By solving this equation we obtain

$$t_{\max}^{(3)} = \frac{3}{\sqrt{2}} \log(N+1). \tag{15}$$

It can be seen in the last section that quantizers corresponding to the maximal signal amplitudes $t_{\max}^{(1)}, t_{\max}^{(2)}$ and $t_{\max}^{(3)}$ has almost optimal distortion, especially for the large values of N.

Now we will describe the iterative method for the minimization of the approximate distortion $\tilde{D}$. As we have already seen, total distortion $\tilde{D}$ is the function of t_{N-1} and y_N (according to the (7) and (8)). Moreover, relations (5) and (6) explicitly determine these two quantities as the function of the optimal maximal signal amplitude $t_{\max}$. Hence we obtain closed form analytical expression for total distortion $\tilde{D}$ as the function of number of quantization levels N and optimal maximal signal amplitude $t_{\max}$, i.e., $\tilde{D} = \tilde{D}(N, t_{\max})$. An optimal value $\tilde{t}_{\max}$ can be obtained by applying Newton iterative method

$$t_{max,i+1} = t_{max,i} - \frac{\frac{\mathrm{d}\tilde{D}}{\mathrm{d}t_{\max}}(t_{max,i})}{\frac{\mathrm{d}^2\tilde{D}}{\mathrm{d}t_{\max}^2}(t_{max,i})}. \tag{16}$$

Estimates $t_{\max}^{(1)}, t_{\max}^{(2)}$ and $t_{\max}^{(3)}$ can be used as the starting points. We can obtain $\tilde{D}(t_{\max})$ as the closed-form expression and then perform the symbolic differentiation to obtain closed-form expressions for the first and second derivative.

Similar procedure (Newton method) can be repeated for the numerical computation of $t_{\max}^{comp}$, which is the minimum of the exact distortion $D(t_{\max})$. Since in (16) we need the second derivation of the exact distortion $D(t_{\max})$, produced expressions will be too large. Instead of that, it is more suitable to use some non-gradient optimization method for minimization of $D(t_{\max})$. A good choice is Simplex method (Chong and Zak, 2001) due to its simplicity and ease of implementation. Recall that exact expression for $D(t_{\max})$ is obtained by combining equations (4) with equation (1). So, the procedure for computation of $t_{\max}^{comp}$ can be describes as follows: first we find the exact expression for $D(t_{\max})$ substituting (4) in (1), and then apply Simplex optimization method.

From the Fig. 1 we can see that both D and $\tilde{D}$ has only one global minimum, and hence both Newton iterative method and non-gradient optimization methods will converge to $\tilde{t}_{\max}$ and $t_{\max}^{comp}$ respectively.

At the end of this section we recall that $t_{\max}^{comp}$ is the exact optimal maximal amplitude obtained by minimization of the exact distortion D, but that minimization procedure is very complicated. Because of that, we found four approximate values for the optimal maximal amplitude: $\tilde{t}_{\max}, t_{\max}^{(1)}, t_{\max}^{(2)}, t_{\max}^{(3)}$. $\tilde{t}_{\max}$ was obtained by minimization of approximate expression for distortion, $\tilde{D}$, but that minimization procedure is also very complicated. On the other hand, expressions for $t_{\max}^{(1)}, t_{\max}^{(2)}, t_{\max}^{(3)}$ are very simple and in closed-form, and because of that they are very useful in practical applications. As it would be seen from the next section, $t_{\max}^{(1)}$ and $t_{\max}^{(3)}$ are very closed to $t_{\max}^{comp}$, especially $t_{\max}^{(3)}$.

4. Numerical Examples

In this section we will compare the approximate values $t_{\max}^{(1)}, t_{\max}^{(2)}, t_{\max}^{(3)}$ and $\tilde{t}_{\max}$ with the exact optimal value $t_{\max}^{comp}$. $t_{\max}^{(1)}, t_{\max}^{(2)}, t_{\max}^{(3)}$ is calculated using expressions (13),

(14) and (15) respectively, while $\tilde{t}_{\max}$ and $t_{\max}^{comp}$ is found by minimization of $\tilde{D}$ and D as it was described in previous section. Minimization procedures for D and $\tilde{D}$ are implemented in the symbolic programming package MATHEMATICA (see for example (Wolfram, 2003)). The performance of a quantizer is often specified in terms of signal-to-noise ratio ($SNRQ$), which is directly obtained from the distortion D using the following relation

$$SNRQ = 10 \log_{10} \left(\frac{\sigma^2}{D} \right).$$

Note that $SNRQ$ is descending function of D, and hence the quantizer is better when $SNRQ$ has higher value. In Table 1 we show the maximal signal amplitude approximate values $t_{\max}^{(1)}, t_{\max}^{(2)}, t_{\max}^{(3)}$ and $\tilde{t}_{\max}$ and exact value $t_{\max}^{comp}$ for different values of number of quantization levels N. It can be seen that values $t_{\max}^{(3)}$ and $t_{\max}^{comp}$ are very close.

Table 2 gives the optimal $SNRQ$ of companding based quantizer ($SNRQ^{comp}$, corresponding to the $t_{\max}^{comp}$) as well as the differences between $SNRQ$ value corresponding to the $t_{\max}^{(1)}, t_{\max}^{(2)}, t_{\max}^{(3)}$ and $\tilde{t}_{\max}$($SNRQ^{(1)}, SNRQ^{(2)}, SNRQ^{(3)}$, and $S\tilde{N}RQ$ respectively), and $SNRQ^{comp}$. We denoted $\Delta SNRQ^{(i)} = SNRQ^{comp} - SNRQ^{(i)}$ for $i = 1, 2, 3$ and $\Delta S\tilde{N}RQ = SNRQ^{comp} - S\tilde{N}RQ$.

It is evident from Table 2 that companding quantizer with maximal signal amplitude $t_{\max}^{(3)}$ has $SNRQ$ almost equal to the maximum possible value $SNRQ^{comp}$ of the quantizer based on the companding model (difference is comparable with the machine precision). This estimation is even better than $\tilde{t}_{\max}$ obtained by minimization of the approximate distortion $\tilde{D}$. Other two estimates have also close values of $SNRQ$ to the optimal. Therefore, we can conclude that estimation $t_{\max}^{(3)}$ is very accurate approximation of the optimal maximal signal amplitude and can be efficiently used in the practical applications.

Note that for $N \geqslant 64$ all considered companding models (with maximal signal amplitude equal to $t_{\max}^{(1)}, t_{\max}^{(3)}, \tilde{t}_{\max}$ and $t_{\max}^{comp}$ satisfies the stopping criterion of the Lloyd–Max

Table 1
$t_{\max}^{(1)}, t_{\max}^{(2)}, t_{\max}^{(3)}, \tilde{t}_{\max}$ and $t_{\max}^{comp}$ for different values of N

N	$t_{\max}^{(1)}$	$t_{\max}^{(2)}$	$t_{\max}^{(3)}$	$\tilde{t}_{\max}$	$t_{\max}^{comp}$
16	5.44891	7.51992	6.01015	5.64467	6.01886
32	6.98886	9.05986	7.41721	7.06473	7.42437
64	8.49319	10.5642	8.85521	8.50929	8.86145
128	9.98034	12.0513	10.3092	9.96659	10.3151
256	11.4591	13.5301	11.7714	11.4304	11.7764
512	12.9336	15.0046	13.2376	12.8975	13.2480
1024	14.4061	16.4771	14.7059	14.3662	14.7109
2048	15.8775	17.9485	16.1753	15.8358	16.1875
4096	17.3484	19.4194	17.6452	17.3057	17.7500
8192	18.819	20.8901	19.1153	18.7759	18.7500

Table 2

$SNRQ^{comp}$, $\Delta SNRQ^{(1)}$, $\Delta SNRQ^{(2)}$, $\Delta SNRQ^{(3)}$, and $\Delta S\tilde{N}RQ$ for different values of N

N	$SNRQ^{comp}$	$\Delta SNRQ^{(1)}$	$\Delta SNRQ^{(2)}$	$\Delta SNRQ^{(3)}$	$\Delta S\tilde{N}RQ$
16	18.0802	0.0227864	0.0576701	$3.98027*10^{-6}$	0.00886233
32	23.8401	0.00668196	0.034428	$1.43247*10^{-6}$	0.00437507
64	29.7272	0.00238848	0.0186859	$5.70152*10^{-6}$	0.00216523
128	35.6802	0.000985415	0.00971984	$2.49387*10^{-6}$	0.00107601
256	41.6668	0.000444215	0.00495525	$1.187*10^{-6}$	0.00053623
512	47.6704	0.00021043	0.00250157	$3.03133*10^{-6}$	0.000267631
1024	53.6824	0.000102473	0.00125698	$6.54602*10^{-6}$	0.000133861
2048	59.6987	0.0000507707	0.000630517	$5.94234*10^{-6}$	0.0000674533
4096	65.7172	0.000025261	0.000316148	$7.79579*10^{-6}$	0.0000347439
8192	71.7367	0.0227864	$1.576701*10^{-6}$	$9.12923*10^{-6}$	0.0000197884

algorithm. According to the (Gray, 2004), this stopping criterion can be expressed as follows

$$\delta = \frac{D - D_N^*}{D_N^*} < 0.005,$$

where D_N^* is the distortion of the optimal N-level quantizer. Comparing $SNRQ$ values, stopping criterion can be expressed as

$$\frac{SNRQ - SNRQ_N^*}{SNRQ_N^*} < 0.02,$$

where $SNRQ_N^* = 10 \log_{10}(1/D_N^*)$. This approves that quantizers based on the companding model with optimized maximal signal amplitude, are close to the optimal quantizers. On the other hand, let us remember that in practice, companding quantizers has much simpler structure than optimal quantizers, especially for the large number of quantization levels N. Hence they are very suitable for the practical applications.

Optimization of the lossy algorithm based on optimal compression law was done in this and previous section. Also, we gave some simple approximate expressions in closed form for the maximal amplitude $t_{\max}$ which satisfied stop criterion of Max-Lloyd algorithm. The best approximate expression is $t_{\max}^{(3)}$ which is very simple and gives negligible error. The aim was to choose approximate expression with minimal error in comparison to optimal solution. Importance of the good choice of $t_{\max}$ can be seen from Fig. 1: for small values for N (e.g., $N = 32$), error of distortion can be greater than 5%.

Optimal companding quantizer, presented in this and previous section, will be used as a part of generalized compression algorithm, which will be presented in Section 6.

5. Novel Simple Lossless Compression Method

In this section, novel simple lossless compression method will be presented. This is our second goal in this paper.

Suppose that the output of the N-level quantizer are the indices $i = 0, \ldots, N - 1$. In the rest of this section we will suppose that $N = 2^r$ where r is positive integer. Hence, all indices $i = 0, \ldots, N - 1$ can be coded by the r-bit codeword and average bit-rate (average number of bits per sample) is equal to $R = r$. We will develop the simple compression algorithm which improves that bit-rate, but taking into account the properties of the source X. Since X has Laplacian distribution, most probable indices are middle ones, i.e., indices from the segment $I = N/4, \ldots, 3N/4 - 1$. The corresponding range for the source signal is $[-x_1, x_1]$, where

$$x_1 = -\frac{3}{\sqrt{2}} \log \left[\frac{1}{2} \left(1 + \exp \left(-\frac{\sqrt{2}}{3} t_{\max} \right) \right) \right].$$

Since I consists of $N/2$ indices, required number of bits for the representation of each index $i \in I$ is $r - 1$.

Now we will describe the structure of the coder. First, indices are grouped into the frames consisting of M indices. First bit of each frame is control bit. If all indices in the group belongs to I, they are all coded with $r - 1$ bits each, and the first bit is set to 1. Otherwise, first bit is set to 0 and all indices are coded with r bits each.

Denote by p_1 the probability that output index i belongs to I (i.e., that sample of the signal belongs to $(-x_1, x_1)$). It is equal to $p_1 = 1 - \exp(-\sqrt{2}x_1)$. Now the probability that M consecutive indices belongs to I is equal to p_1^M. Hence, we can compute average bit-rate as follows

$$\overline{R} = p_1^M (r - 1) + \left(1 - p_1^M \right) r + \frac{1}{M}, \tag{17}$$

where the last term corresponds to the first bit of the frame.

For example, by taking $M = 3$, $N = 128$ and $t_{\max} = t_{\max}^{(3)}$, we obtain $x_1 = 1.45401$, $p_1 = 0.87207$ and $\overline{R} = 6.67012$. We will compare this result with the entropy H of the source and average bit-rate R_h obtained by using Huffman algorithm (Hankerson *et al.*, 2004). Let us remember that last two quantities are given by

$$H = \sum_{i=1}^{N} P_i \log_2 \left(\frac{1}{P_i} \right), \quad R_h = \sum_{i=1}^{N} P_i \left\lceil \log_2 \frac{1}{P_i} \right\rceil.$$

The length of source extension M can be optimized using relation (16). For $N = 128$ it can be shown that optimal value of M is $M = 3$. Also holds $H = 6.38823$ and $R_h = 6.8427$. Let us notice that holds $H < \overline{R} < H + 1$ and $\overline{R} < R_h$. Similarly holds for the other values of N. Hence we can conclude that our lossless coder gives better performances than Huffman algorithm. It is worth mentioning that our coder is drastically simpler for the realization than Huffman code. Hence it is suitable for implementations where simplicity of the coder structure is important.

6. Generalized Compression Algorithm – Combination of Lossy and Lossless Compression

Generalized compression algorithm which is combination of lossy and lossless compression method, is presented in this section. Optimization of lossy method with optimal compression function was done by optimization of $t_{\max}$ in Sections 3 and 4. An simple lossless compression method was proposed in Section 5, where also optimization of extended information source was done by optimizing M, where M is the order of the source extension.

It is assumed that stack S has length M and it is empty on algorithm start. Also assume that $t_{\max}$ is equal to $t_{\max}^{(3)}$ and this value is precomputed.

Algorithm 1. (Coder)
Input. Signal sample x.
Step 1. Compute $y = c(x)$.
Step 2. Apply uniform quantizer on y and find the index i of the cell where y belongs.
Step 3. Put i on the top of the stack S.
Step 4. If S is not full, go to the **Step 1** and continue with the next sample. Otherwise continue.
Step 5. Take M indices $i_1, \ldots, i_M$ from the stack S.
Step 6. If $i_j \in I$ for every $j = 1, \ldots, M$ then code every index with $r - 1$ bits and produce the binary codeword $C = (c_1, \ldots, c_{(r-1)M}$ of the length $(r - 1)M$. Transmit $1C$. Otherwise code every index with r bits and produce the binary codeword $C = (c_1, \ldots, c_{rM})$ of the length rM. Transmit $0C$.

Algorithm 2. (Decoder)
Step 1. Take one bit c from the input.
Step 2. If $c = 1$, take $(r-1)M$ bits from the input and decode indices $i_1, \ldots, i_M$, each from $r - 1$ bits. Otherwise take rM bits from the input and decode indices $i_1, \ldots, i_M$, each from r bits.
Step 3. For each $j = 1$ to M perform:
>   *Step* 3.1. Set y to the value of representation level of i_j-th cell of uniform quantizer. Return $x = c^{-1}(y)$.

At the end of this section we present practical example of usage of generalized compression method for compression of a still image. Fig. 2 (left) presents original image and Fig. 2 (right) presents image after compression. We can see that compressed image is visually very close to the original image. First, we found differences of the consecutive image samples because they have Laplacian distribution, and then applied compression method on those differences. To prevent error propagation, beside differences we also transmit one original image sample on every 128 samples. We use companding quantizer with $N = 64$ levels to quantizing differences of the samples, and for this quantizer $SNRQ = 28.038dB$. We use lossless algorithm with $M = 3$. Average bit-rate for compressed image is $R_{compr.} = 5.4702$ bits/pixel and for original image is $R_{orig.} = 8$ bits/pixel, so the compression ratio is $R_{orig.}/R_{compr.} = 1.462$.

Fig. 2. Original image (left) and compressed image (right).

7. Conclusion

This paper provides the simple structure coder for memoryless Laplacian source. We used companding quantizer and perform the optimization of the maximal signal amplitude $t_{\max}$. There are derived 3 analytical estimates of the optimal maximal signal amplitude as well as one numerical estimate (based on the minimization of approximate distortion) and exact numerical method for its computation. It is shown that one of the estimates has almost optimal distortion. Due to the exact simple analytical expression, this estimate is very suitable for the practical applications. Also we develop simple lossless coder algorithm for Laplacian source and compared with the Huffman code and source entropy. Main advantage of our method is simple realization structure and less bit-rate, compared with the Huffman method. Finally, we joined lossy and lossless compression methods in one generalized compression algorithms. We applied this generalized algorithm on the one still image and obtained good results.

Generally our coder gives the very simple realization structure and performances close to optimal and hence it is very useful in practical applications, such as speech signals, images, bio-medical images, etc.

References

Chong, E.K.P., and S.H. Zak (2001). *An Introduction to Optimization*. John Wiley & Sons inc., New York, Chichester, Weinheim, Brisbane, Singapore, Toronto.

Gray, R. (2004). *Quantization and Data Compression, Lecture Notes*. Stanford University.

Hankerson, D., G.A. Harris and P.D. Johnson Jr. (2004). *Introduction Information Theory and Data Compression*. 2nd ed. Chapman&HALL/CRC.

Jayant, N.S., and P. Noll (1984). *Digital Coding of Waveforms*. Prentice-Hall, New Jersey.

Judell, N., and L. Scharf (1986). A simple derivation of Lloyd's classical result for the optimum scalar quantizer. *IEEE Transactions on Information Theory*, **32**(2), 326–328.

Max, J. (1960). Quantizing for minimum distortion. *IRE, Transactions on Information Theory*, **IT-6**, 7–12.

Na, S. (2004). On the support of fixed-rate minimum mean-squared error scalar quantizers for Laplacian source. *IEEE Transactions on Information Theory*, **50**(5), 937–944.

Na, S., and D.L. Neuhoff (2001). On the support of MSE-optimal, fixed-rate, scalar quantizers. *IEEE Transactions on Information Theory*, **47**(7), 2972–2982.

Nikolic, J., and Z. Peric (2008). Lloyd–Max's algorithm implementation in speech coding algorithm based on forward adaptive technique. *Informatica*, **19**(2), 255–270.

Panter, P.F., and W. Dite (1951). Quantization distortion in pulse count modulation with nonuniform spacing of levels. In *Proc. IRE*. pp. 44–48.

Perić, Z.H., and J.R. Nikolić (2007). An effective method for initialization of Lloyd–Max's algorithm of optimal scalar quantization for Laplacian source. *Informatica*, **18**(2), 1–10.

Perić, Z.H., J.R. Nikolić and D.M. Pokrajac (2007). Estimation of the support region for Laplacian source scalar quantizers. *Journal of Electrical Engineering*, **58**(1), 47–51.

Ramalho, M. (2002). *Ramalho G.711 Lossless (RGL) Codec Whitepaper*. Cysco Systems, Inc.

Sayood, K. (2006). *Introduction to data Compression*, 3rd ed. Elsevier Inc.

Starosolsky, R. (2007). Simple fast and adaptive lossless image compression algorithm. *Software-Practice and Experience*, **37**, 65–91.

Wolfram, S. (2003). *Mathematica Book*, 5th. ed. Wolfram Media.

Z.H. Perić was born in Nis, Serbia, in 1964. He received the BSc degree in electronics and telecommunications from the Faculty of Electronic Engineering, Nis, Serbia, Yugoslavia, in 1989, and MSc degree in telecommunications from the University of Nis, in 1994. He received the PhD degree from the University of Nis, also, in 1999. He is currently professor at the Department of Telecommunications and vicedean of the Faculty of Electronic Engineering, University of Nis, Serbia. His current research interests include the information theory, source and channel coding and signal processing. He is particulary working on scalar and vector quantization techniques in speech and image coding. He was author and coauthor in over 100 papers in digital communications. Dr. Peric has been a reviewer for *IEEE Transactions on Information Theory*. He is member Editorial Board of journal *Electronics and Electrical Engineering*.

M.D. Petković was born in Nis, Serbia, in 1984. He graduated mathematics and computer science at the Faculty of Sciences and Mathematics, Nis, Serbia in 2006. He also graduated telecommunications at Faculty of Electronic Engineering, Nis, Serbia in 2007. Currently he is the PhD student in computer science at the Faculty of Sciences and Mathematics, Nis, Serbia. His research interests include the source and channel coding, symbolic computation and optimization methods. He is the author of about 20 papers (about 10 of them in peer-reviewed international journals). He is supported by Ministry of Science, Republic of Serbia, Grant No. 144011.

M.R. Dinčić was born in Nis, Serbia, in 1983. He received the BSc degree in telecommunications from the Faculty of Electronic Engineering, Nis, in 2007. He is currently on doctoral studies on the same faculty, and he is scholar of Ministry of Science Republic of Serbia. His current research interests include source coding and quantization of speech signals and images.

Apie glaudinimo algoritmą, skirtą neturinčiam atminties
Laplaso šaltiniui bei grindžiamą optimalia glaudinimo procedūra

Zoran H. PERIĆ, Marko D. PETKOVIĆ, Milan DINČIĆ

Pirmoje straipsnio dalyje siekiama optimizuoti glaudinimo su informacijos praradimu koderį, realizuotą kompandiniu signalų keitikliu pagal diskrečiuosius lygius ir veikiantį pagal optimalų glaudinimo dėsnį. Darbe šis uždavinys išspręstas optimizuojant tokiam keitikliui maksimalią amplitudę Laplaso šaltinio atveju. Rasta uždaroje formoje aproksimuojanti išraiška optimaliai amplitudei gauti. Ji tenkina Loido–Makso signalų dydžių keitikliams optimalumo kriterijų, kai $R \geqslant 6$ bitai/atskaitai. Antroje straipsnio dalyje pateiktas naujas paprastas glaudinimo metodas be informacijos praradimo. Šis metodas esti žymiai paprastesnis nei Hafmano metodas. Jis pranašesnis už pastarąjį ir tuo, kad jį taikant gaunami geresni taikymų rezultatai. Straipsnio pabaigoje autoriai apjungia optimalų kompaundinį signalo keitiklį pagal diskrečiuosius lygius su kodavimo metodu be informacijos praradimo į apibendrintą glaudinimo procedūrą, kuri panaudota tam tikro skaitmeninio vaizdo apdorojimui. Eksperimentiniai apibendrintos procedūros patikros rezultatai pateikti grafikų ir lentelių pavidalu.