EEE431 - Digital Communications

MATLAB Assignment 1 Report

Efe Tarhan 22002840

1) Introduction and Outline

Compared to analog communications, digital communication methods are the most widely used methods around the world for services like internet, television broadcast and GSM. A digital communication system transfers “digitized” data to transmitter channel where this digitized signal can be converted to an analog signal or remain digital. Initially for a signal to be transmitted with a digital communication channel it must be digitized by sampling that discretizes the signal in time, quantization which discretizes signal amplitude and finally an encoder which is commonly used for objectives like data size reduction or security concerns. In this MATLAB assignment 2 steps of preparing a signal for transmission with a digital channel has been inspected.

The uniform and nonuniform quantization algorithms have been investigated to understand the conversion of a signal with continuous range of values to a signal with predetermined amount of discrete amplitude levels. 2 different quantization techniques have been investigated; uniform quantization is the technique where the amplitude range of the signal is divided into predetermined number of intervals and midpoint of these intervals are chosen as quantization levels whereas nonuniform quantization technique requires the Lloyd-Max Algorithm that iteratively assigns quantization levels to center of mass of quantization intervals and boundaries as the midpoint of these quantization levels. Two of these algorithms has been tested with a synthetically created DMS (Discrete Memoryless Source).

At the second part of the assignment the one of the most used encoding techniques has been investigated which is the LZW (Lempel-Ziv-Welch) Algorithm on a 500-word text taken from Wikipedia. Effects of this encoding has been investigated throughout the procedure of the assignment.

2) Part 1

In this part a random variable Y has been defined as summation of two random variables X_1 and X_2. My student ID number is 22002840, therefore the value a is equal to 8 and the value b is equal to 4. These values create the following random variables:

$$X_1 \sim \mathbb{U}[0,16]$$

$$X_2 \sim \mathbb{U}[-8,0]$$

Pdf of these random variables have been plotted using MATLAB and can be seen in Figures 1 and 2.

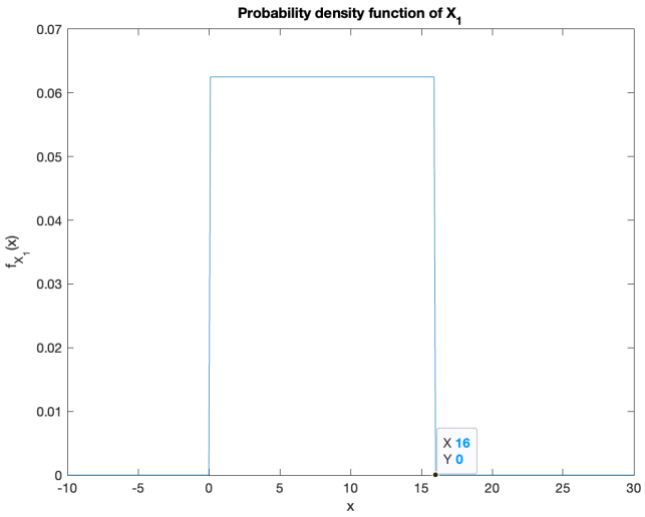

Fig.1 Probability density function of the r.v. X_1

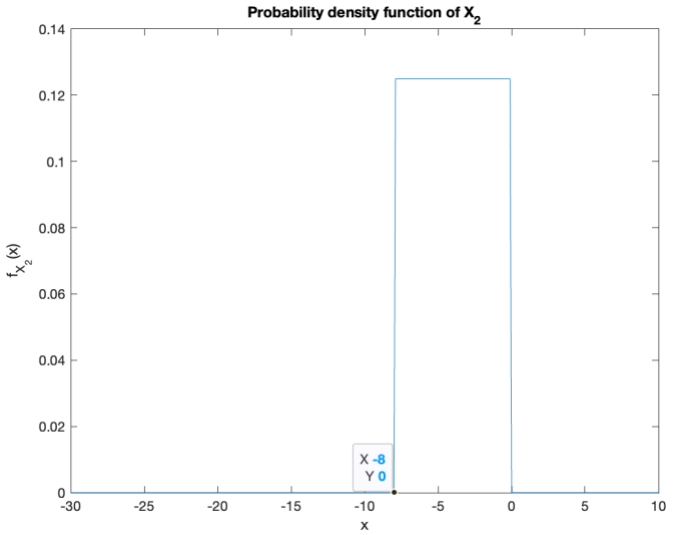

Fig.2 Probability density function of the r.v. X_2.

Probability density function of the r.v. Y can be calculated by the following operations using the pdf of X_1 and X_2.

$$Y = X_1 + X_2$$

Then,

$$P(Y < y \mid X_1 = x_1)$$
$$= P(X_1 + X_2 < y \mid X_1 = x_1) = P(X_2 < y - x_1)$$
$$\frac{d}{dy} P(X_2 < y - x_1) = f_{X_2}(y - x1) = f_{Y|X_1}(y|x_1)$$

Lastly,

$$f_{Y,X_1}(y, x_1) = f_{X_2}(y - x_1) \cdot f_{X_1}(x_1)$$
$$f_Y(y) = \int_{-\infty}^{\infty} f_{X_2}(y - x_1) \cdot f_{X_1}(x_1)\, dx_1$$
$$= f_{X_1}(y) * f_{X_2}(y)$$

To find it exactly by using the distributions of X_1 and X_2 :

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X_2}(y - x) \cdot f_{X_1}(x)\, dx$$

$$= \begin{cases} \int_0^{y+8} \frac{1}{16} \cdot \frac{1}{8} dx, & -8 < y \le 0 \\ \int_y^{y+8} \frac{1}{16} \cdot \frac{1}{8} dx, & 0 < y \le 8 \\ \int_y^{16} \frac{1}{16} \cdot \frac{1}{8} dx, & 8 < y \le 16 \\ 0, & else \end{cases}$$

$$\boxed{= \begin{cases} \frac{(y+8)}{128}, & -8 < y \le 0 \\ \frac{1}{16}, & 0 < y \le 8 \\ \frac{(16-y)}{128}, & 8 < y \le 16 \\ 0, & else \end{cases}}$$

The plot of the pdf of Y can also examined from Figure 3 as it represents the formula calculated above.

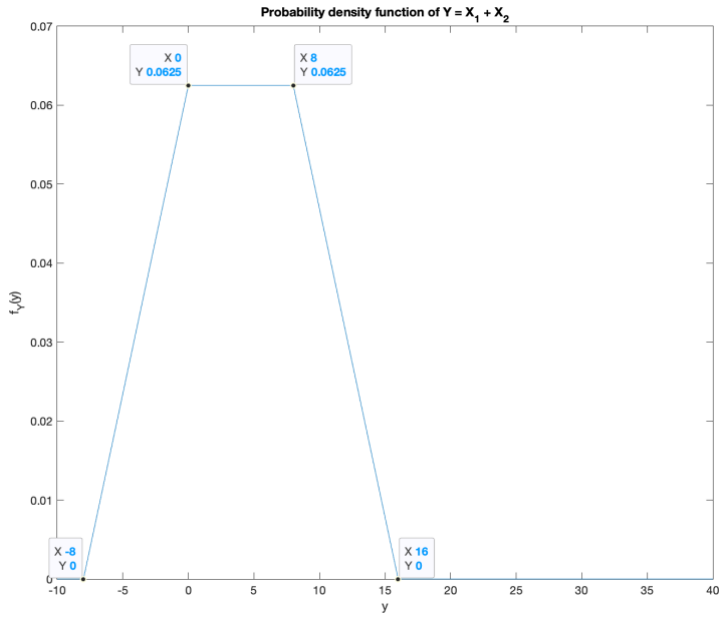

Fig.3 Probability density function of the r.v. Y.

Power of this random variable can also be calculated using a simple formula which is the following:

$$E[Y^2] = \int_{-\infty}^{\infty} y^2 f_Y(y) dy$$

$$= \int_{-8}^{0} \frac{y^2(y+8)}{128} dy + \int_{0}^{8} \frac{y^2}{16} dy + \int_{8}^{16} \frac{y^2(16-y)}{128} dy$$

$$= \left(\frac{y^4}{512} + \frac{8y^3}{384} \right) \Big|_{y=-8}^{y=0} + \frac{y^3}{48} \Big|_{y=0}^{y=8} + \left(\frac{y^3}{24} - \frac{y^4}{512} \right) \Big|_{y=8}^{y=16}$$

$$= \frac{8^4}{384} - \frac{8^4}{512} + \frac{8^3}{48} + \frac{16^3}{24} - \frac{16^4}{512} - \frac{8^3}{24} + \frac{8^4}{512} \cong \boxed{42.67}$$

This concludes the first task of the Part 1 of the MATLAB assignment.

After finding the pdf of Y. 1000000 samples has been generated from a source model that exhibit a pdf as Y. To accomplish this, I have created 1000000 samples for X_1 and 1000000 samples for X_2 then sum these two vectors up. Histogram of the resultant vector entries can be seen from Figure 4.

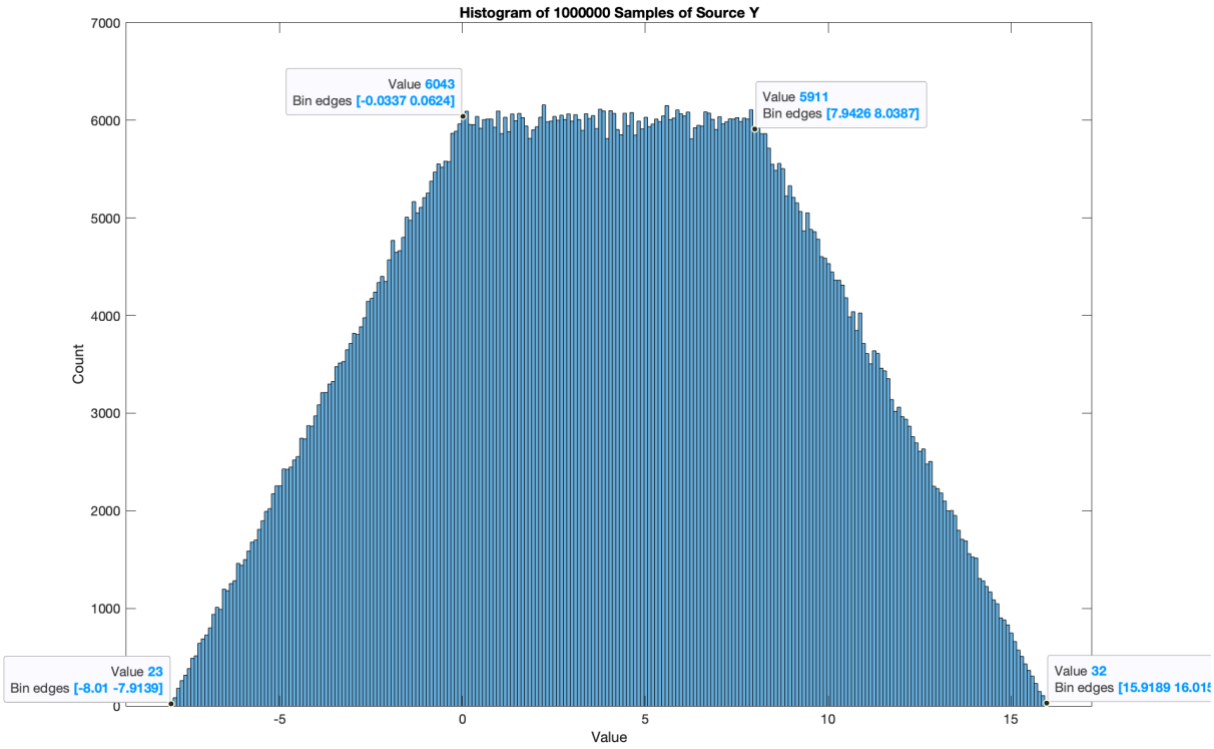

Fig.4 Histogram of 1000000 samples generated from the source Y.

As it can be seen the histogram has similar boundary points and a shape with the theoretical pdf of the random variable Y which is denoted in Figure 3. Only difference between two plots is that the histogram shows the number of times a value occurs between a given range for each bin therefore the values become like the pdf if the count value for each bin is divided to 1000000. Average power of the samples generated from Y can also be calculated with the following formula:

$$P = \frac{1}{1000000}\left(\sum_{i=1}^{1000000} Y[i]^2\right)$$

This value is being calculated with MATLAB for 5 times for 5 different trials of generating 1000000 from Y and the values can be seen in Table 1.

Trial	Power
1	42.7382
2	42.6284
3	42.6377
4	42.8342
5	42.7387

Table 1. Average power calculations of source Y for 5 trials

As it can be seen both 5 values are around the theoretically found power 42.67 which shows the resemblance between statistical and probabilistic results. This concludes the third task of the Part 1.

After generating the sequence Y, fourth task of Part 1 is to design a uniform quantizer with 4 quantization levels that will generate the quantized version of the source Y which is denoted as $\tilde{Y}$.

$$\tilde{Y} = \begin{cases} -5\,, & -8 < Y \leq -2 \\ 1\,, & -2 < Y \leq 4 \\ 7\,, & 4 < Y \leq 10 \\ 13\,, & 10 < Y \leq 16 \end{cases}$$

Quantization boundaries and levels as found above are shown in the Figure 5., on top of the pdf of Y.

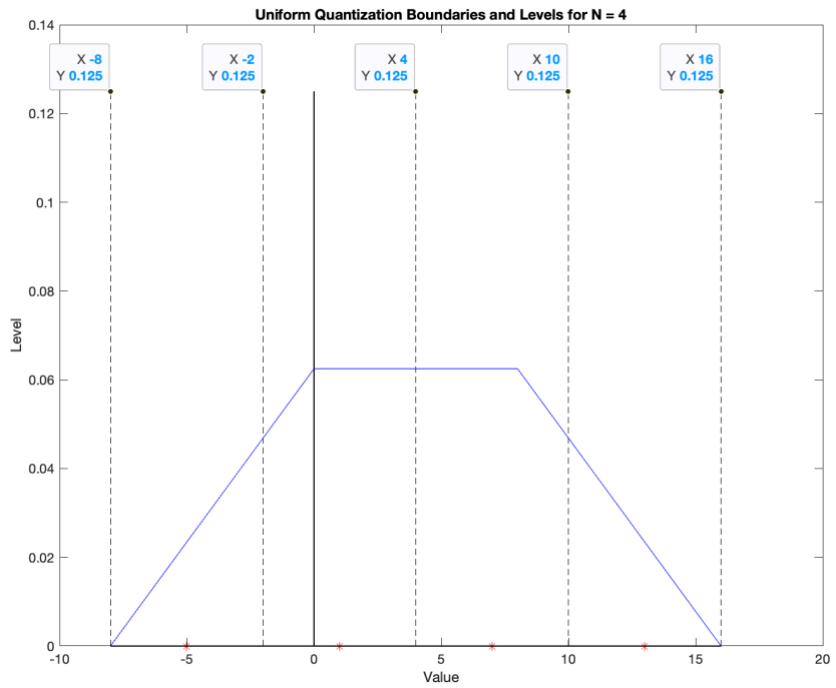

Fig.5 Quantization boundaries and levels of $\tilde{Y}$ denoted on top of the pdf of Y

After quantizing Y, the error is calculated for each sample between the original value and the quantized one by subtracting two vectors.

$$e = Y - \tilde{Y}$$

Then histogram of the error has been calculated and plotted which can be seen in Figure 6.

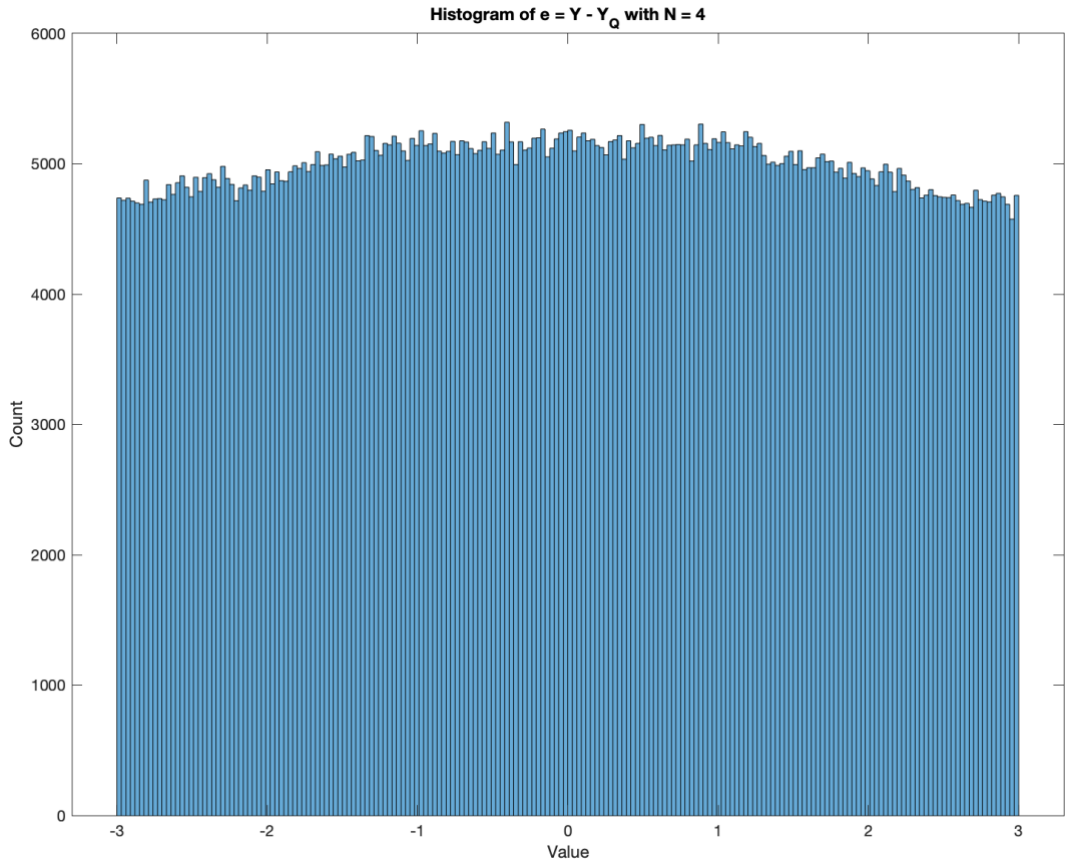

Fig.6 Histogram of the error between the samples of Y and $\tilde{Y}$.

The theoretical value of the pdf of error can be calculated with the following operations.

$$e|_{y\in[-8,-2]} = \left(Y - \tilde{Y}\right)|_{y\in[-8,-2]} = (Y+5)\,|_{y\in[-8,-2]}$$

$$\Longrightarrow f_{e,Y\in[-8,-2]}(x) = \begin{cases} \dfrac{(x+3)}{128}, & -3 \le x \le 3 \\ 0, & else \end{cases}$$

Similarly,

$$f_{e,Y\in[-2,4]}(x) = \begin{cases} \dfrac{(x+9)}{128}, & -3 \le x \le -1 \\ \dfrac{1}{16}, & -1 < x \le 3 \\ 0, & else \end{cases}$$

$$f_{e,Y\in[4,10]}(x) = \begin{cases} \dfrac{1}{16}, & -3 \le x \le 1 \\ \dfrac{(9-x)}{128}, & 1 < x \le 3 \\ 0, & else \end{cases}$$

$$f_{e,Y\in[10,16]}(x) = \begin{cases} \dfrac{(3-x)}{128}, & -3 < x \le 3 \\ 0, & else \end{cases}$$

As a result, the final formula for error can be found by the total probability theorem:

$$f_e(x) = \sum_{i=1}^{4} f_{e,Y\in[6i-14,6i-8]}(x)$$

$$= \begin{cases} \dfrac{(x+23)}{128}, & -3 < x \le -1 \\ \dfrac{11}{64}, & -1 < x \le 1 \\ \dfrac{(23-x)}{128}, & 1 < x \le 3 \end{cases}$$

This result explains why the histogram doesn't exhibit a perfectly uniform behavior, because the number of quantization intervals are not enough for eliminating the effect of nonconstant behavior of the pdf of the source. By increasing the quantization intervals, the error limits can be tightened, and the error histogram can have a more uniform shape.

After plotting the histogram of the quantization error, average power and SQNR can also be checked by using the generated samples. The formula for average power of quantization error is the following:

$$P_{Y-\tilde{Y}} = \frac{1}{1000000} \sum_{i=1}^{i=1000000} \left(Y[i] - \widetilde{Y}\,[i]\right)^2$$

Also, the SQNR of the quantization in 10log dB scale can be found by the following formula:

$$SQNR = 10\log\left(\sum_{i=1}^{i=1000000} Y[i]^2\right) - 10\log\left(\sum_{i=1}^{i=1000000} \left(Y[i] - \widetilde{Y}\,[i]\right)^2\right)$$

The simulation results using the generated samples can be seen in Figure 7.

```
Average power of quantization error where N = 4: 2.914300
SQNR in dB: 11.662864 dB
```

Fig.7 Average power of quantization error and SQNR on MATLAB terminal.

Now the after completing the operation for N = 4 quantization levels, the same procedure will be repeated for N = 32 quantization levels. Since there are 8 times more quantization levels than the initial quantization levels the results will be shown in more closed form because it will be hard to show theoretical calculations. The histogram of the calculated quantization error for N = 32 can be seen from Figure 8.

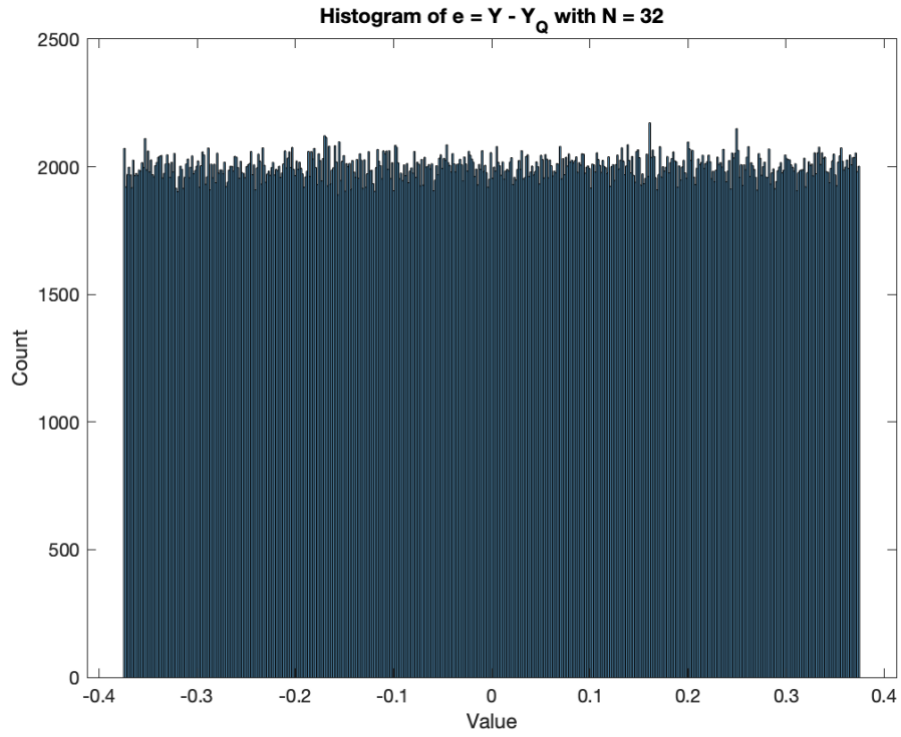

Fig.8 Histogram of quantization errors for N = 32

As it can be seen the maximum possible quantization error has decreased to 1/8 of the previous one which is equal to 0.375. Since the number of quantization levels has increased to 32, the histogram shows more uniform behavior compared to the histogram of quantization error with N = 4. This is because when intervals shrink, they also show more uniform behavior, it has mechanism like the intuition behind the Riemann integral.

Average quantization error power and the SQNR value in decibels can be seen from the Figure 9 which is calculated by using MATLAB.

```
Average power of quantization error where N = 32: 0.046855
SQNR in dB: 29.600626 dB
```

Fig.9 Average power of quantization error and SQNR for N = 32

As it can be seen the quantization error has significantly decreased and SQNR has significantly increased because the number of quantization levels are now 8 times bigger than the previous one which was an expected result of increasing number of quantization levels.

Both the quantization using N= 4 and N=32 were uniform quantization models where only the length between quantization levels and the first quantization level can be determined. The performance of quantization can also be increased by setting the location of the quantization boundaries and levels that minimizes the quantization error using Lloyd Max Algorithm. Since the pdf of the Y is time limited, b_0 and b_N does not have to be put to infinity and all 32 quantization levels can be put inside the pdf region. The pseudocode of Lloyd-Max quantization algorithm can be seen below.

BEGIN
FOR i = 1: N
INITALIZE $[\, b_{i-1}, b_i] = \left[\min(Y) + (i-1) \cdot \frac{(\max(Y) - \min(Y))}{N}, \min(Y) + i \cdot \frac{(\max(Y) - \min(Y))}{N}\right]$
INITALIZE $x_i = \frac{b_{i-1} + b_i}{2}$
WHILE $E[e^2] >$ threshold
    SET $x_i = E\left[x \mid x \in [b_{i-1}, b_i]\right]$
    SET $b_i = \frac{(x_i + x_{i+1})}{2}$
END

It is not easy to determine a threshold since it has a certain limit that can't be decreased anymore, the while loop has changed with a for loop that does the operation with given number of times.

For N = 4 the Lloyd – Max algorithm has been used and the following quantization regions and levels are obtained and stabilized after 100 iterations, those can be seen from Figure 10 where the levels and boundaries plotted on top of the pdf of Y.

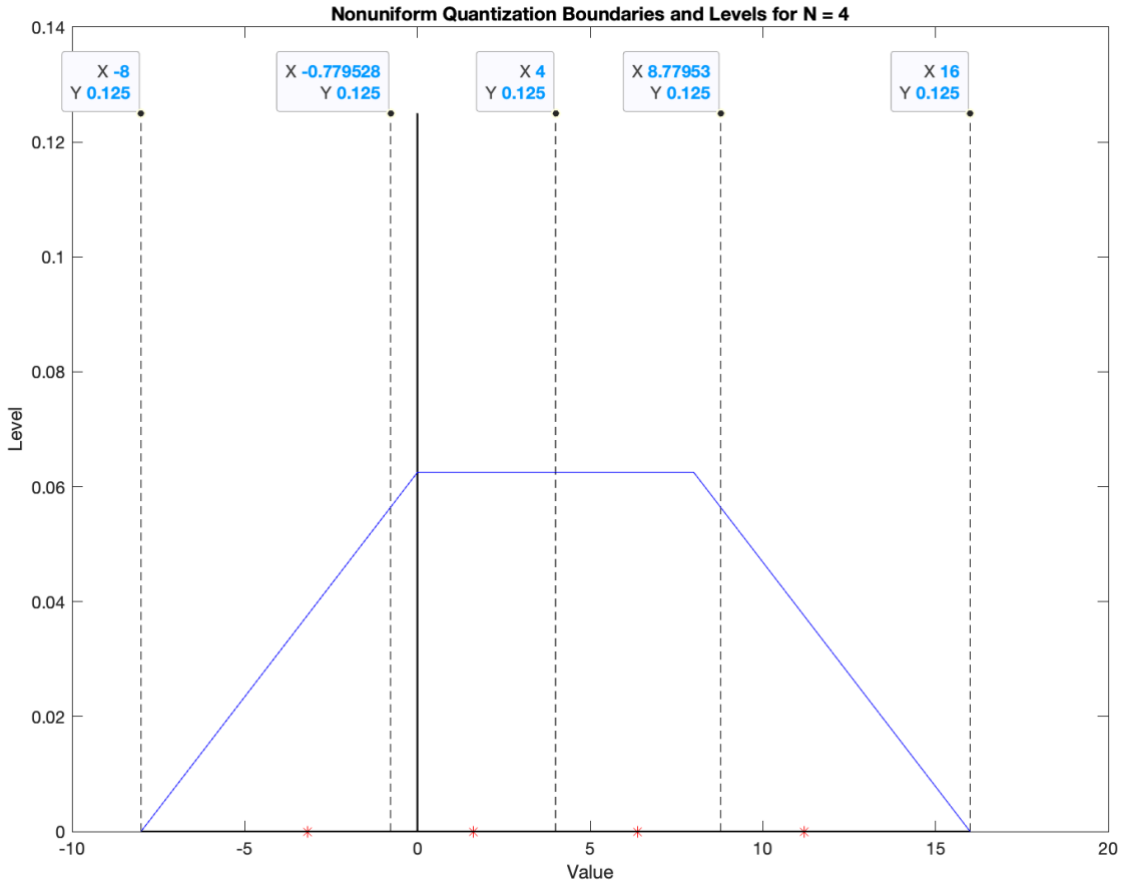

Fig.10 Quantization levels and boundaries obtained with Lloyd-Algorithm for N = 4

Compared to the quantization levels and boundaries of uniform quantization that can be seen from Figure 5, the levels and boundaries are shifted towards the region of the pdf with higher values. Histogram of the quantization errors has been plotted which can be seen in Figure 11.

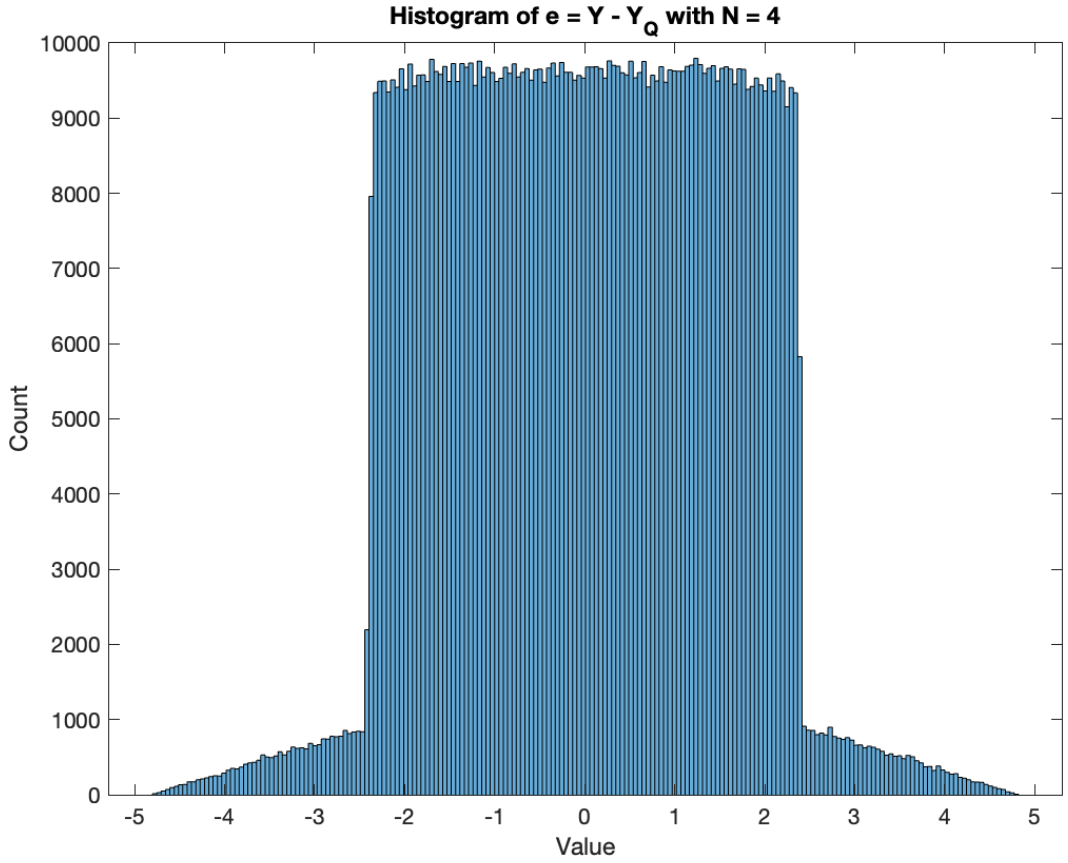

Fig.11 Histogram of the nonuniform quantization error for N = 4

Compared to the uniform quantization the histogram has a more discrete and nonuniform behavior since the regions does not have the same length and the quantization levels does not always locate at the middle of the quantization boundaries. Also, the average quantization error power and the SQNR were calculated using MATLAB which can be seen in Figure 12.

```
Average power of nonuniform quantization error where N = 4: 2.300416
SQNR in dB: 12.693147 dB
```

Fig.12 Average power of nonuniform quantization error and SQNR on MATLAB terminal.

Performance of the quantizer has increased compared to the results demonstrated in Figure 9 but compared to increasing the number of quantization levels the performance did not increase significantly since the optimization could improve it until a point.

For also comparing the result of the nonuniform quantizer for N = 32, the Lloyd-Max algorithm has also been applied for N = 32. The algorithm has again iterated 100 times to obtain the parameters. Resultant quantization regions and levels can be seen on top of the pdf of Y in Figure 13.

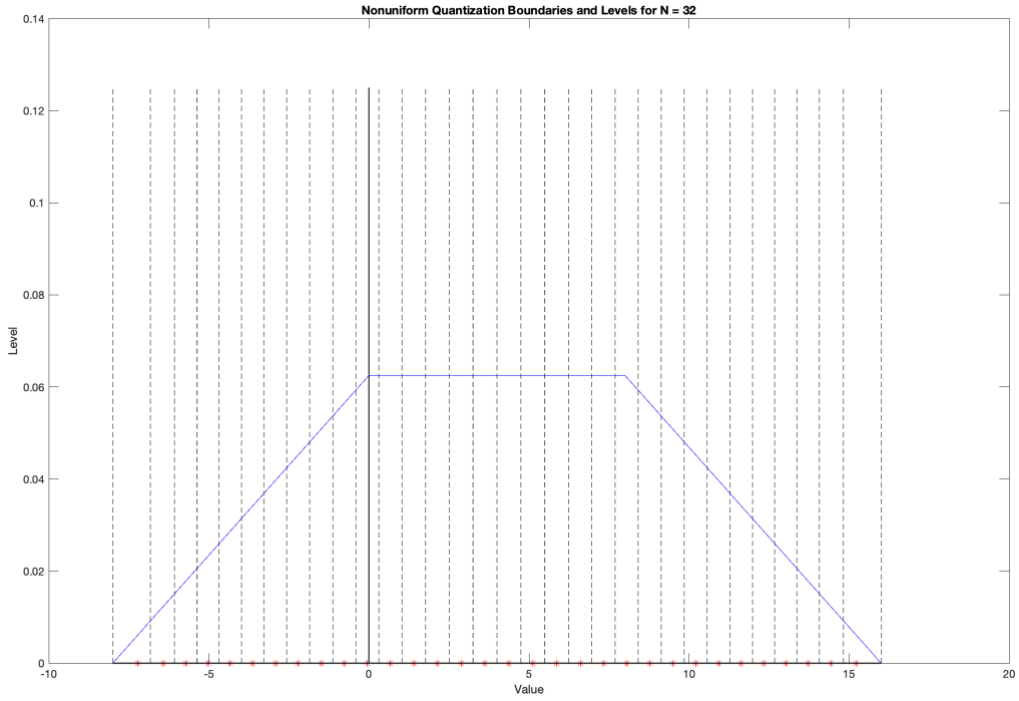

Fig.13 Quantization levels and boundaries of nonuniform quantization with N = 32

Compared to the nonuniform quantization with N = 4, the nonuniform quantization with higher number of quantization levels look more like the uniform quantization except the first and the last regions which look more nonuniform compared to the others. As a result, the histogram of quantization error for N = 32 look more like its uniform version. The histogram can be seen in Figure 14.

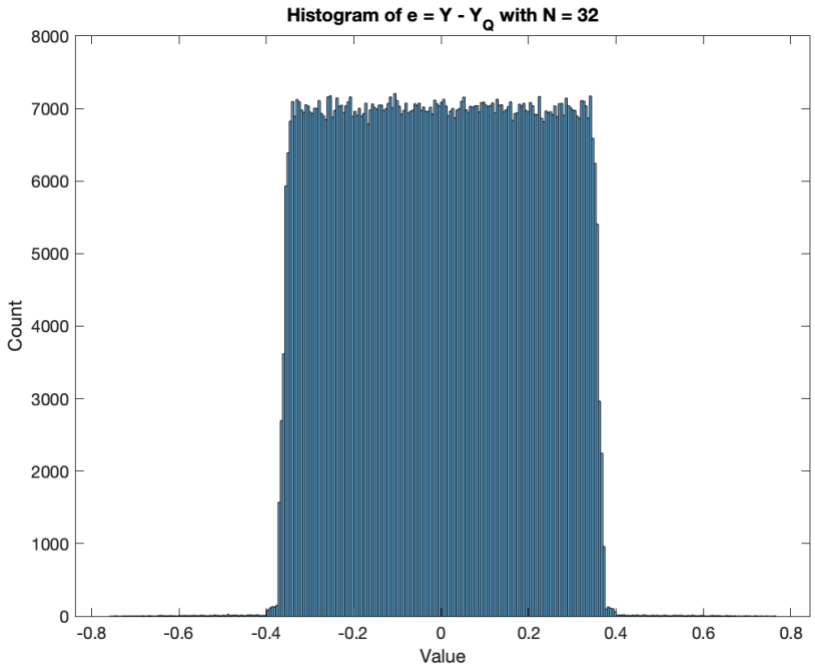

Fig.14 Histogram of nonuniform quantization error with N = 32

In Figure 14, there are tails at the end of the histogram which still indicate the nonuniform behavior of the quantizer possibly at the lower and higher ends of the pdf where the quantization level is not located at the midpoint of the region and the region is larger compared to others. Average quantization error power and SQNR were also calculated. The results can be seen from Figure 15.

```
Average power of nonuniform quantization error where N = 32: 0.044174
SQNR in dB: 29.859544 dB
```

Fig.15 Average power of nonuniform quantization error and SQNR for N = 32

Compared to uniform quantization with N = 32, the performance of the quantizer has increased but still not significantly.

After completing the nonuniform quantizer with N = 32 levels the quantization part of the digitization process of the given assignment has been completed. In the next section the second part of the assignment will be explained which will be about the LZW algorithm.

3) Part 2

For this part of the assignment, a classical text has been found from the Gutenberg project which is the “Republic” by Plato [1]. The used text can be seen in Figure 16.

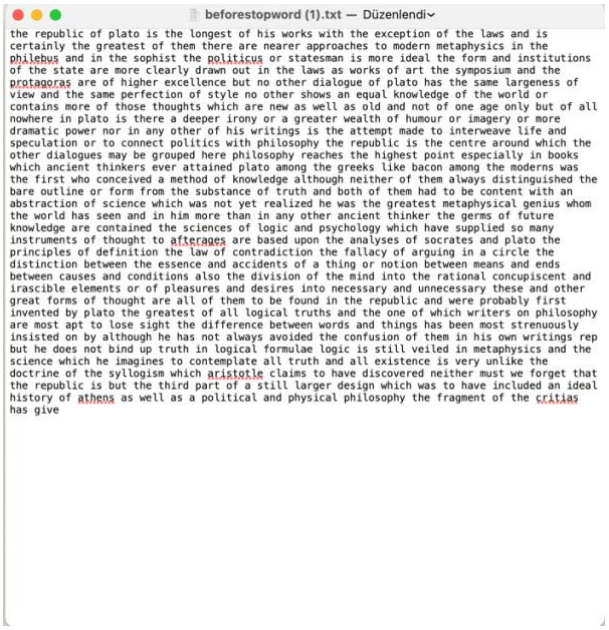

Fig.16 A brief part of the “Republic” by Plato [1]

For preprocessing, non-alphabetic characters have been removed from the text, all characters converted to uppercase and a termination sign “#” has added to the end of the text to indicate termination. After completing this step, the Lempel-Ziv-Welch algorithm has been used to encode the text which has created a binary encoded signal that can be send through a digital channel. The final dictionary contains 1259 elements which means 1231 new vocabulary items has been created for encoding. The average codeword length has been calculated by dividing the total number of bits of the encoded message to the number of letters in the original text including whitespaces. The result can be seen from Figure 17.

```
Average codeword length with no coding is: 5.000000 bits/sample!
Average codeword lenght after Lempel-Ziv Algorithm is: 4.075209 bits/sample!
```

Fig.17. Average codeword lengths of encoded and non-encoded texts

As it can be seen from the results the average codeword length of encoded text has decreased compared to the non-encoded text which is an expected result for encoding. Average number of bits for representing each character has decreased since the natural language has repetitions where the LZW algorithm can benefit from.

After encoding the text using the LZW algorithm the text is decoded using the same procedure. The decoded text can be seen in Figure 18.

THE REPUBLIC OF PLATO IS THE LONGEST OF HIS WORKS WITH THE EXCEPTION OF THE LAWS AND IS CERTAINLY THE GREATEST OF THEM THERE ARE NEARER APPROACHES TO MODERN METAPHYSICS IN THE PHILEBUS AND IN THE SOPHIST THE POLITICUS OR STATESMAN IS MORE IDEAL THE FORM AND INSTITUTIONS OF THE STATE ARE MORE CLEARLY DRAWN OUT IN THE LAWS AS WORKS OF ART THE SYMPOSIUM AND THE PROTAGORAS ARE OF HIGHER EXCELLENCE BUT NO OTHER DIALOGUE OF PLATO HAS THE SAME LARGENESS OF VIEW AND THE SAME PERFECTION OF STYLE NO OTHER SHOWS AN EQUAL KNOWLEDGE OF THE WORLD OR CONTAINS MORE OF THOSE THOUGHTS WHICH ARE NEW AS WELL AS OLD AND NOT OF ONE AGE ONLY BUT OF ALL NOWHERE IN PLATO IS THERE A DEEPER IRONY OR A GREATER WEALTH OF HUMOUR OR IMAGERY OR MORE DRAMATIC POWER NOR IN ANY OTHER OF HIS WRITINGS IS THE ATTEMPT MADE TO INTERWEAVE LIFE AND SPECULATION OR TO CONNECT POLITICS WITH PHILOSOPHY THE REPUBLIC IS THE CENTRE AROUND WHICH THE OTHER DIALOGUES MAY BE GROUPED HERE PHILOSOPHY REACHES THE HIGHEST POINT ESPECIALLY IN BOOKS WHICH ANCIENT THINKERS EVER ATTAINED PLATO AMONG THE GREEKS LIKE BACON AMONG THE MODERNS WAS THE FIRST WHO CONCEIVED A METHOD OF KNOWLEDGE ALTHOUGH NEITHER OF THEM ALWAYS DISTINGUISHED THE BARE OUTLINE OR FORM FROM THE SUBSTANCE OF TRUTH AND BOTH OF THEM HAD TO BE CONTENT WITH AN ABSTRACTION OF SCIENCE WHICH WAS NOT YET REALIZED HE WAS THE GREATEST METAPHYSICAL GENIUS WHOM THE WORLD HAS SEEN AND IN HIM MORE THAN IN ANY OTHER ANCIENT THINKER THE GERMS OF FUTURE KNOWLEDGE ARE CONTAINED THE SCIENCES OF LOGIC AND PSYCHOLOGY WHICH HAVE SUPPLIED SO MANY INSTRUMENTS OF THOUGHT TO AFTERAGES ARE BASED UPON THE ANALYSES OF SOCRATES AND PLATO THE PRINCIPLES OF DEFINITION THE LAW OF CONTRADICTION THE FALLACY OF ARGUING IN A CIRCLE THE DISTINCTION BETWEEN THE ESSENCE AND ACCIDENTS OF A THING OR NOTION BETWEEN MEANS AND ENDS BETWEEN CAUSES AND CONDITIONS ALSO THE DIVISION OF THE MIND INTO THE RATIONAL CONCUPISCENT AND IRASCIBLE ELEMENTS OR OF PLEASURES AND DESIRES INTO NECESSARY AND UNNECESSARY THESE AND OTHER GREAT FORMS OF THOUGHT ARE ALL OF THEM TO BE FOUND IN THE REPUBLIC AND WERE PROBABLY FIRST INVENTED BY PLATO THE GREATEST OF ALL LOGICAL TRUTHS AND THE ONE OF WHICH WRITERS ON PHILOSOPHY ARE MOST APT TO LOSE SIGHT THE DIFFERENCE BETWEEN WORDS AND THINGS HAS BEEN MOST STRENUOUSLY INSISTED ON BY ALTHOUGH HE HAS NOT ALWAYS AVOIDED THE CONFUSION OF THEM IN HIS OWN WRITINGS REP BUT HE DOES NOT BIND UP TRUTH IN LOGICAL FORMULAE LOGIC IS STILL VEILED IN METAPHYSICS AND THE SCIENCE WHICH HE IMAGINES TO CONTEMPLATE ALL TRUTH AND ALL EXISTENCE IS VERY UNLIKE THE DOCTRINE OF THE SYLLOGISM WHICH ARISTOTLE CLAIMS TO HAVE DISCOVERED NEITHER MUST WE FORGET THAT THE REPUBLIC IS BUT THE THIRD PART OF A STILL LARGER DESIGN WHICH WAS TO HAVE INCLUDED AN IDEAL HISTORY OF ATHENS AS WELL AS A POLITICAL AND PHYSICAL PHILOSOPHY THE FRAGMENT OF THE CRITIAS HAS GIVE#

Fig.18 The decoded text which was initially encoded by using the LZW algorithm.

After completing the encoding-decoding step, the count of each initial dictionary element in the text has been found and normalized by the total amount of characters in the text. This

has created a pmf if the text has been sent in terms of individual characters. The resulting pmf can be seen from the plot in Figure 19.

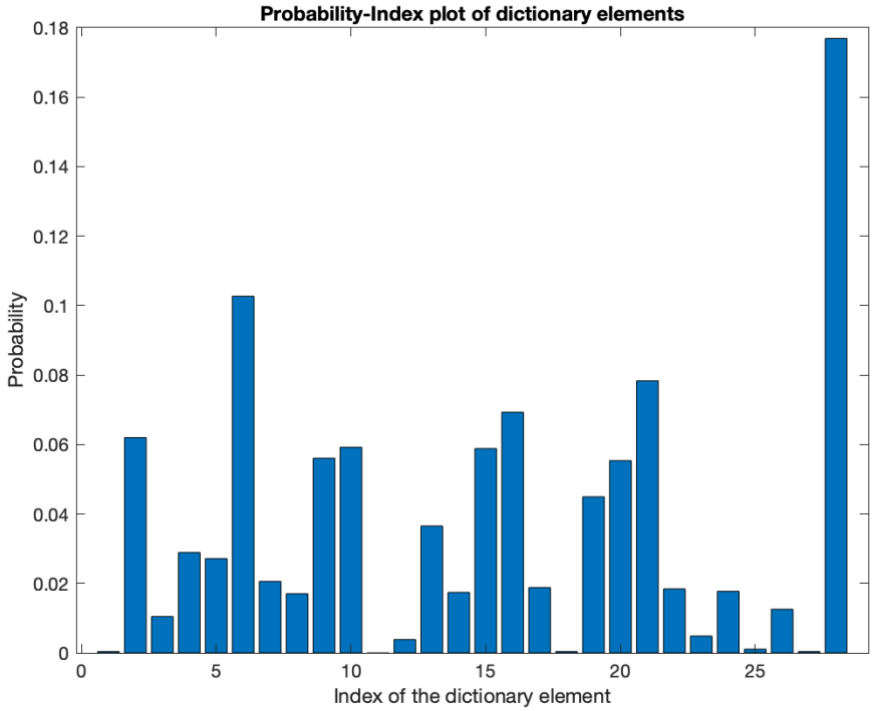

Fig.19 Probability-Index plot of initial dictionary elements created during the LZW algorithm.

Since the pmf is not perfectly uniform the entropy has appeared to be less than the maximum value of entropy of a uniform random variable with 28 discrete elements. The calculated entropy of the initial dictionary as a source can be seen in Fig.20.

```
Entropy of the transmitted text is: 4.066274 bits/sample!
```

Fig.20 The entropy of the dictionary created with LZW.

From the results the average codeword length obtained by LZW algorithm is nearly 0.09 bits/sample longer than the calculated entropy of the dictionary. This shows that the encoding algorithm has managed to decrease the average number of bits per each character by 1 which increased the efficiency of the transmission. If a larger text would be used, the average codeword length would possibly be closer to the value of entropy since the dictionary will contribute more to the transmission of repeated samples since language become more

repetitive with the increasing size of texts, this will happen even the words are generated randomly according to the Heap's Law [2].

4) References

[1] Plato, "The Republic," Project Gutenberg, 2023. [Online]. Available: https://www.gutenberg.org/ebooks/55201. (accessed October 28, 2023).

[2] "Heaps' law: Estimating the number of terms.", *Stanford NLP Group.* https://nlp.stanford.edu/IR-book/html/htmledition/heaps-law-estimating-the-number-of-terms-1.html (accessed October 21, 2023).