

# HW1-236200

Adir Rahamim 318382728  
Roi Tzur-Hilleli 207717414

November 25, 2019

## Question 1:

1. The function we want to minimize is the expected absolute-deviation:

$$E_{\epsilon_Q}^1 = \int_{\phi_L}^{\phi_H} |x - Q(x)|p(x)dx = (*)$$

As for the interval  $[\phi_L, \phi_H]$  we partitioning it using the given decision levels:  $\{d_i\}_{i=0}^K$  creating  $K$  decisions regions:  $D_i = [d_{i-1}, d_i), i = 1, \dots, K$  s.t:  $\phi_L = d_0 < d_1 < \dots < d_K = \phi_H$ , and for the mapping function, we know that:  $Q(x) = r_i | x \in D_i$  for representation levels  $\{r_i\}_{i=1}^J$  s.t  $r_1 < r_2 < \dots < r_J$ . and we get:

$$(*) = \sum_{i=1}^K \int_{d_{i-1}}^{d_i} |x - r_i|p(x)dx$$

and we get that the optimization problem we wish to solve is:

$$\underset{\{d_i\}_{i=0}^K \{r_i\}_{i=1}^J}{\text{minimize}} \sum_{i=1}^K \int_{d_{i-1}}^{d_i} |x - r_i|p(x)dx$$

2. Necessary condition for optimal representation level is given by:

$$\frac{\partial}{\partial r_j} E_{\epsilon_Q}^1 = 0 \text{ for } j = 1, \dots, J$$

$\Downarrow$

$$\begin{aligned} \frac{\partial}{\partial r_j} E_{\epsilon_Q}^1 &= \frac{\partial}{\partial r_j} \sum_{i=1}^K \int_{d_{i-1}}^{d_i} |x - r_j|p(x)dx = \frac{\partial}{\partial r_j} \int_{d_{j-1}}^{d_j} |x - r_j|p(x)dx = \\ &= \int_{d_{j-1}}^{d_j} \frac{\partial}{\partial r_j} |x - r_j|p(x)dx = - \int_{d_{j-1}}^{d_j} \text{sign}(x - r_j)p(x)dx = \end{aligned}$$

$$\begin{aligned}
&= \int_{d_{j-1}}^{r_j} p(x)dx - \int_{r_j}^{d_j} p(x)dx \stackrel{\text{demand}}{=} 0 \\
&\quad \Downarrow \\
&\int_{d_{j-1}}^{r_j} p(x)dx = \int_{r_j}^{d_j} p(x)dx
\end{aligned}$$

From the optimality condition we can see that the optimal  $r'_j$ 's divide the interval  $[d_{j-1}, d_j]$  into two equal probability regions, this can be interpreted as the probabilistic median of the interval for the given  $p(x)$ .

3. Necessary condition for optimal decision level is given by:

$$\frac{\partial}{\partial d_j} E_{\epsilon_Q}^1 = 0 \text{ for } j = 1, \dots, K-1$$

for fixed  $d_0 = \phi_L, d_K = \phi_H$

$$\begin{aligned}
&\Downarrow \\
&\frac{\partial}{\partial d_j} E_{\epsilon_Q}^1 = \frac{\partial}{\partial d_j} \sum_{i=1}^K \int_{d_{i-1}}^{d_i} |x - r_j| p(x) dx = \\
&= \frac{\partial}{\partial d_j} \int_{d_{j-1}}^{d_j} |x - r_j| p(x) dx + \frac{\partial}{\partial d_j} \int_{d_j}^{d_{j+1}} |x - r_{j+1}| p(x) dx = \\
&= |d_j - r_j| p(d_j) - |d_j - r_{j+1}| p(d_j) \stackrel{\text{demand}}{=} 0 \\
&\quad \Downarrow \\
&|d_j - r_j| = |d_j - r_{j+1}|
\end{aligned}$$

If we open the absolute value we get two possibilities:

- (a)  $d_j - r_j = d_j - r_{j+1} \Rightarrow r_j = r_{j+1}$  This answer is **impossible** as we know that for each  $j$ :  $r_j < r_{j+1}$ .
  - (b)  $d_j - r_j = -(d_j - r_{j+1}) \Rightarrow d_j = \frac{r_j + r_{j+1}}{2}$  So given the representation level we get that the optimal decision levels are the mean of the representation levels.
4. the Max-Lloyd procedure for designing a  $b$ -bit quantizer that minimizes the expected absolute-deviation for a given input PDF  $p(x)$  can be described as:
- (a) Initialization: Set arbitrary decisions levels  $\{d_i\}_{i=0}^K$ .
  - (b) Compute the optimal representation levels and set to  $\{r_i\}_{i=1}^J$  for the current decision levels  $\{d_i\}_{i=0}^K$ .
  - (c) Compute the optimal decisions levels and set to  $\{d_i\}_{i=0}^K$  for the current representation levels  $\{r_i\}_{i=1}^J$ .
  - (d) If stopping criteria has not met, return to (b).

## Question 2

1. We need to formulate an optimal quantizer design that uses  $b = 1$  bit

$\Downarrow$

we have  $J = 2^b = 2$  representation levels, labled  $r_1, r_2$ , and  $J + 1 = 3$  decision levels labled  $d_0 = -4a, d_1, d_2 = 4a$ .

- (a) As we learned in class we know that for given decision levels the optimal representation levles are:

$$r_j = \frac{\int_{d_{j-1}}^{d_j} xp(x)dx}{\int_{d_{j-1}}^{d_j} p(x)dx} \text{ for } j = 1, 2$$

and fr given representation levels the optimal decision levels are:

$$d_j = \frac{r_j + r_{j+1}}{2} \text{ for } j = 0, 1, 2$$

Now we guess the value of the decision level,  $d_1 = 0$ , and lets calculate the representation levels according to the current decision levels:

$$r_1 = \frac{\int_{-4a}^{-2a} \frac{x}{4a} dx}{\int_{-4a}^{-2a} \frac{1}{4a} dx} = \frac{\left. \frac{x^2}{8a} \right|_{-4a}^{-2a}}{\left. \frac{x}{4a} \right|_{-4a}^{-2a}} = -3a$$

$$r_2 = \frac{\int_{2a}^{3a} x \left( -\frac{1}{4a^2}x + \frac{7}{8a} \right) dx + \int_{3a}^{4a} x \left( \frac{1}{4a^2}x - \frac{5}{8a} \right) dx}{\int_{2a}^{3a} \left( -\frac{1}{4a^2}x + \frac{7}{8a} \right) dx + \int_{3a}^{4a} \left( \frac{1}{4a^2}x - \frac{5}{8a} \right) dx} = \dots = 3a$$

Now lets calculate the decision level according to the calculated representation levels:

$$d_1 = \frac{r_1 + r_2}{2} = 0$$

The result didnt changed, that mean that the values we got are the optimal.

- (b) The minimal squared error achived is:

$$E \{ \epsilon^2 \} = \int_{-4a}^{4a} p(x) (x - Q(x))^2 dx =$$

$$= \int_{-4a}^{-2a} \frac{1}{4a} (x - (-3a))^2 dx + \int_{2a}^{3a} \left( -\frac{1}{4a^2}x + \frac{7}{8a} \right) (x - 3a)^2 dx + \int_{3a}^{4a} \left( \frac{1}{4a^2}x - \frac{5}{8a} \right) (x - 3a)^2 dx =$$

$$= \dots = \frac{a^2}{6} + \frac{5a^2}{48} - \frac{a^2}{48} = \frac{a^2}{4}$$

2. We need to formulate an optimal quantizer design that uses  $b = 2$  bit

$\Downarrow$

we have  $J = 2^b = 4$  representation levels, labled  $r_1, r_2, r_3, r_4$ , and  $J + 1 = 5$  decision levels labled  $d_0, d_1, d_2, d_3, d_4$ .

- (a) We start with a guess for the decision levels:  $d_0 = -4a, d_1 = -2a, d_2 = 2a, d_3 = 3a, d_4 = 4a$  and calculate the matching representation levels:

$$r_1 = -3a, r_2 = 0, r_3 = 2\frac{5}{12}a, r_4 = 3\frac{7}{12}a$$

Now we update the decision levels according to the calculated representation levels:

$$d_0 = -4a, d_1 = -1.5a, d_2 = 1\frac{5}{24}a, d_3 = 3a, d_4 = 4a$$

Now we update the representation levels:

$$r_1 = -3a, r_2 = 0, r_3 = 2\frac{5}{12}a, r_4 = 3\frac{7}{12}a$$

We got the same result as the previous calculation, that means that the algorithm converged, and the final answer is:

$$d_0 = -4a, d_1 = -1.5a, d_2 = 1\frac{5}{24}a, d_3 = 3a, d_4 = 4a$$

$$r_1 = -3a, r_2 = 0, r_3 = 2\frac{5}{12}a, r_4 = 3\frac{7}{12}a$$

- (b) The minimal squared error achieved is:

$$\begin{aligned} E\{\epsilon^2\} &= \int_{-4a}^{4a} p(x) (x - Q(x))^2 dx = \\ &= \int_{-4a}^{-2a} \frac{1}{4a} (x - (-3a))^2 dx + \int_{-2a}^{3a} \left(-\frac{1}{4a^2}x + \frac{7}{8a}\right) \left(x - 2\frac{5}{12}a\right)^2 dx + \dots \\ &\dots + \int_{3a}^{4a} \left(\frac{1}{4a^2}x - \frac{5}{8a}\right) \left(x - 3\frac{7}{12}a\right)^2 dx = \dots = 0.20466a^2 \end{aligned}$$

### Question 3

$$\begin{aligned}
 1. \quad \int_{t \in \Delta_i} (t - t_i)^k dt &= \int_{\frac{i-1}{N}}^{\frac{i}{N}} (t - t_i)^k dt = \left. \frac{(t - t_i)^{k+1}}{k+1} \right|_{\frac{i-1}{N}}^{\frac{i}{N}} = \\
 &= \frac{\left(\frac{i}{N} - t_i\right)^{k+1}}{k+1} - \frac{\left(\frac{i-1}{N} - t_i\right)^{k+1}}{k+1} = \frac{\left(\frac{|\Delta_i|}{2}\right)^{k+1} - \left(\frac{-|\Delta_i|}{2}\right)^{k+1}}{k+1} = \\
 &= \frac{(1 + (-1)^{k+2}) |\Delta_i|^{k+1}}{2^{k+1}(k+1)} = \begin{cases} 0, k - \text{odd} \\ \frac{|\Delta_i|^{k+1}}{2^k(k+1)}, k - \text{even} \end{cases}
 \end{aligned}$$

2. As we studied in class:

$$MSE = \int_0^1 \left( \phi(t) - \hat{\phi}(t) \right)^2 dt = \sum_{i=1}^N \overbrace{\int_{\frac{i-1}{N}}^{\frac{i}{N}} (\phi(t) - a_i(t - t_i) - c_i)^2 dt}^{MSE_i}$$

Find the optimal  $a_i, c_i$  by requiring:

$$\frac{\partial}{\partial a_i} MSE_i = 0$$

$$\frac{\partial}{\partial c_i} MSE_i = 0$$

For  $a_i$ :

$$\begin{aligned}
 \frac{\partial}{\partial a_i} MSE_i &= \frac{\partial}{\partial a_i} \int_{\frac{i-1}{N}}^{\frac{i}{N}} (\phi(t) - a_i(t - t_i) - c_i)^2 dt = \\
 &= \int_{\frac{i-1}{N}}^{\frac{i}{N}} \frac{\partial}{\partial a_i} (\phi(t) - a_i(t - t_i) - c_i)^2 dt = -2 \int_{\frac{i-1}{N}}^{\frac{i}{N}} (t - t_i) (\phi(t) - a_i(t - t_i) - c_i) dt = \\
 &= -2 \int_{\frac{i-1}{N}}^{\frac{i}{N}} \phi(t) (t - t_i) dt + 2a_i \int_{\frac{i-1}{N}}^{\frac{i}{N}} (t - t_i)^2 dt + 2c_i \int_{\frac{i-1}{N}}^{\frac{i}{N}} (t - t_i) dt
 \end{aligned}$$

We calculate separately the first and second integral, based on the answer from previous section:

$$2a_i \int_{\frac{i-1}{N}}^{\frac{i}{N}} (t - t_i)^2 dt = 2a_i \frac{|\Delta_i|^3}{2^2 3} = \frac{a_i |\Delta_i|^3}{6}$$

$$2c_i \int_{\frac{i-1}{N}}^{\frac{i}{N}} (t - t_i) dt = 0$$

By demending that  $\frac{\partial}{\partial a_i} MSE_i = 0$  we get the equation:

$$-2 \int_{\frac{i-1}{N}}^{\frac{i}{N}} \phi(t) (t - t_i) dt + \frac{a_i |\Delta_i|^3}{6} = 0$$

$\Downarrow$

$$a_i = \frac{12}{|\Delta_i|^3} \int_{\frac{i-1}{N}}^{\frac{i}{N}} \phi(t) (t - t_i) dt$$

Now we will optimize  $c_i$ :

$$\begin{aligned} \frac{\partial}{\partial c_i} MSE_i &= \frac{\partial}{\partial c_i} \int_{\frac{i-1}{N}}^{\frac{i}{N}} (\phi(t) - a_i(t - t_i) - c_i)^2 dt = \\ &= \int_{\frac{i-1}{N}}^{\frac{i}{N}} \frac{\partial}{\partial c_i} (\phi(t) - a_i(t - t_i) - c_i)^2 dt = -2 \int_{\frac{i-1}{N}}^{\frac{i}{N}} (\phi(t) - a_i(t - t_i) - c_i) dt = \\ &= -2 \int_{\frac{i-1}{N}}^{\frac{i}{N}} \phi(t) dt + 2a_i \int_{\frac{i-1}{N}}^{\frac{i}{N}} (t - t_i) dt + 2c_i \int_{\frac{i-1}{N}}^{\frac{i}{N}} 1 dt \end{aligned}$$

Based on the answer from previous section we get that the middle integral equals zero, and:

$$2c_i \int_{\frac{i-1}{N}}^{\frac{i}{N}} 1 dt = 2c_i |\Delta_i|$$

By demanding that  $\frac{\partial}{\partial c_i} MSE_i = 0$  we get that:

$$-2 \int_{\frac{i-1}{N}}^{\frac{i}{N}} \phi(t) dt + 2c_i |\Delta_i| = 0$$

$\Downarrow$

$$c_i = \frac{1}{|\Delta_i|} \int_{\frac{i-1}{N}}^{\frac{i}{N}} \phi(t) dt$$

3. The optimal  $MSE$  is calculated by using the optimal  $a_i, c_i$  (that give us the optimal  $\hat{\phi}(t)$ ):

$$MSE^{opt} = \int_0^1 \left( \phi(t) - \hat{\phi}(t) \right)^2 dt = \sum_{i=1}^N \int_{\frac{i-1}{N}}^{\frac{i}{N}} \left( \phi(t) - a_i^{opt} (t - t_i) - c_i^{opt} \right)^2 dt =$$

$$\begin{aligned}
&= \sum_{i=1}^N \int_{\frac{i-1}{N}}^{\frac{i}{N}} \left( \phi(t)^2 - 2a_i^{opt} \phi(t) (t - t_i) - 2c_i^{opt} \phi(t) + (a_i^{opt})^2 (t - t_i)^2 + 2a_i^{opt} c_i^{opt} (t - t_i) \right. \\
&\quad \left. + (c_i^{opt})^2 \right) dt = \\
&= \int_0^1 (\phi(t))^2 dt - 2 \sum_{i=1}^N a_i^{opt} \int_{\frac{i-1}{N}}^{\frac{i}{N}} \phi(t) (t - t_i) dt - 2 \sum_{i=1}^N c_i^{opt} \int_{\frac{i-1}{N}}^{\frac{i}{N}} \phi(t) dt + \\
&+ \sum_{i=1}^N (a_i^{opt})^2 \int_{\frac{i-1}{N}}^{\frac{i}{N}} (t - t_i)^2 dt + 2 \sum_{i=1}^N a_i^{opt} c_i^{opt} \int_{\frac{i-1}{N}}^{\frac{i}{N}} (t - t_i) dt + |\Delta| \sum_{i=1}^N (c_i^{opt})^2 = (**)
\end{aligned}$$

Using the outcomes from section 1 we get that:

$$\sum_{i=1}^N (a_i^{opt})^2 \int_{\frac{i-1}{N}}^{\frac{i}{N}} (t - t_i)^2 dt = \sum_{i=1}^N (a_i^{opt})^2 \frac{|\Delta_i|^3}{2^2 3} \stackrel{(*)}{=} \frac{|\Delta|^3}{12} \sum_{i=1}^N (a_i^{opt})^2$$

(\*): All the intervals are in the same size:  $|\Delta_i| = |\Delta| = \frac{1}{N}$ .

$$2 \sum_{i=1}^N a_i^{opt} c_i^{opt} \int_{\frac{i-1}{N}}^{\frac{i}{N}} (t - t_i) dt = 0$$

Using the expressions we got for optimal  $a_i, c_i$  we get:

$$\begin{aligned}
&-2 \sum_{i=1}^N a_i^{opt} \int_{\frac{i-1}{N}}^{\frac{i}{N}} \phi(t) (t - t_i) dt = -2 \sum_{i=1}^N (a_i^{opt})^2 \frac{|\Delta_i|^3}{12} = \\
&= -\frac{|\Delta|^3}{6} \sum_{i=1}^N (a_i^{opt})^2 \\
&-2 \sum_{i=1}^N c_i^{opt} \int_{\frac{i-1}{N}}^{\frac{i}{N}} \phi(t) dt = -2 \sum_{i=1}^N c_i^{opt} |\Delta_i| c_i^{opt} = -2 |\Delta| \sum_{i=1}^N (c_i^{opt})^2
\end{aligned}$$

And finally we get:

$$\begin{aligned}
(**) &= \int_0^1 (\phi(t))^2 dt - \frac{|\Delta|^3}{6} \sum_{i=1}^N (a_i^{opt})^2 - 2 |\Delta| \sum_{i=1}^N (c_i^{opt})^2 + \frac{|\Delta|^3}{12} \sum_{i=1}^N (a_i^{opt})^2 + |\Delta| \sum_{i=1}^N (c_i^{opt})^2 = \\
&= \int_0^1 (\phi(t))^2 dt - \frac{|\Delta|^3}{12} \sum_{i=1}^N (a_i^{opt})^2 - |\Delta| \sum_{i=1}^N (c_i^{opt})^2 =
\end{aligned}$$

$$|\Delta| = \frac{1}{N}$$

$$= \int_0^1 (\phi(t))^2 dt - \frac{1}{12N^3} \sum_{i=1}^N (a_i^{opt})^2 - \frac{1}{N} \sum_{i=1}^N (c_i^{opt})^2$$

4. As we saw in class the minimal MSE for picewise-constant approximation is:

$$MSE_{const} = \int_0^1 \phi^2(t) dt - \frac{1}{N} \sum_{i=1}^N (c_i^{opt})^2$$

If we calculate the difference  $MSE_{lin} - MSE_{const}$  we get:

$$MSE_{lin} - MSE_{const} = -\frac{1}{12N^3} \sum_{i=1}^N (a_i^{opt})^2$$

And because  $\forall i, (a_i)^2 \geq 0$  we get that:  $-\frac{1}{12N^3} \sum_{i=1}^N (a_i^{opt})^2 \leq 0$

$\Downarrow$

$$MSE_{lin} \leq MSE_{const}$$

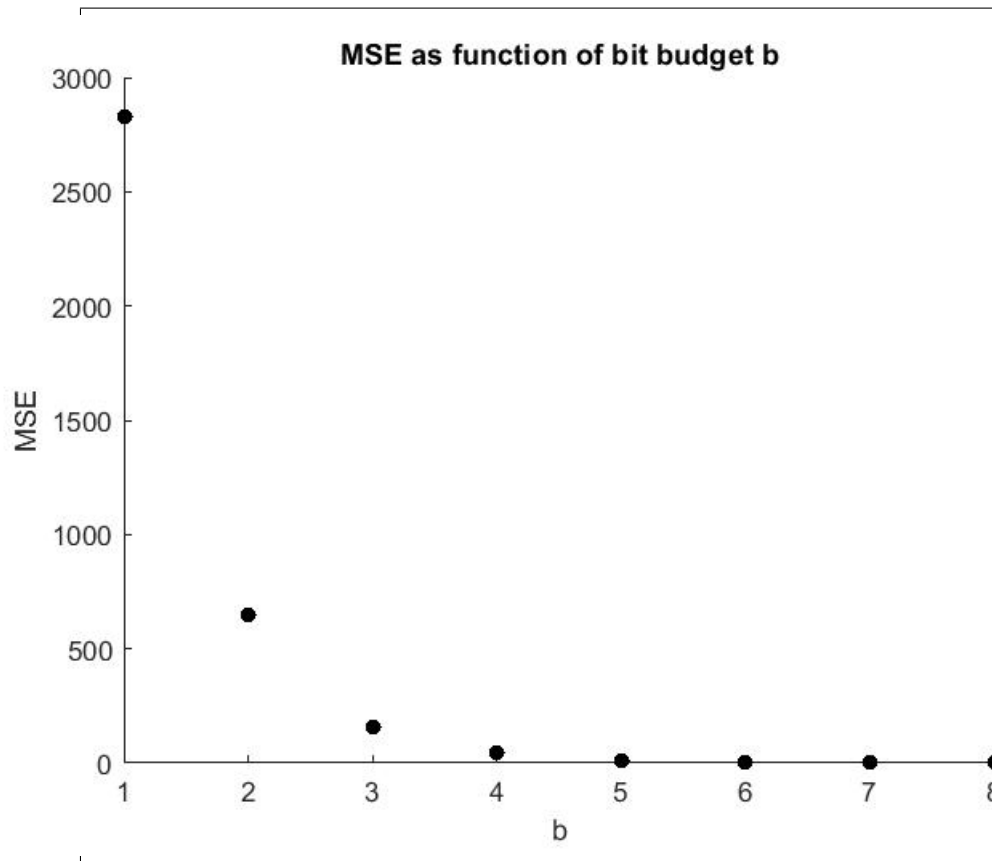
Therefore the MSE using picewise linear approximation is lower(or equal) to the MSE achieved by using picewise-constant approximation.



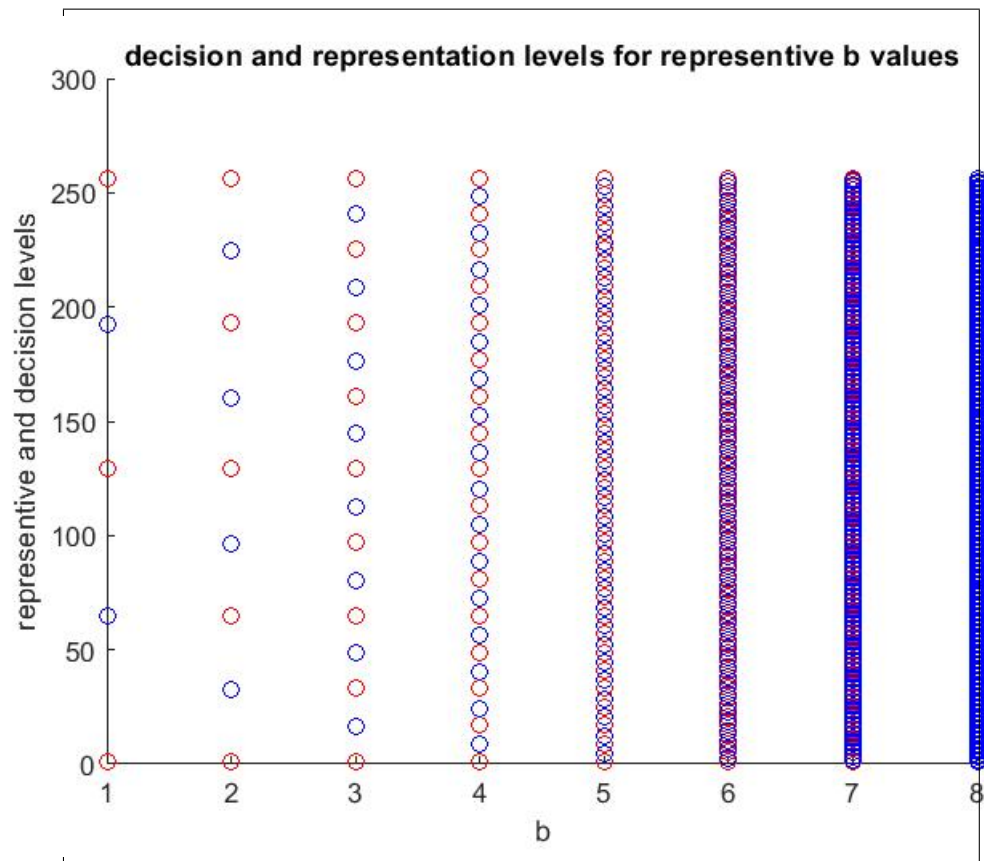
## Part 2 - Matlab

### Question 2 - uniform quantizer:

1. MSE as a function of the bit budget for  $b=1,\dots,8$ :



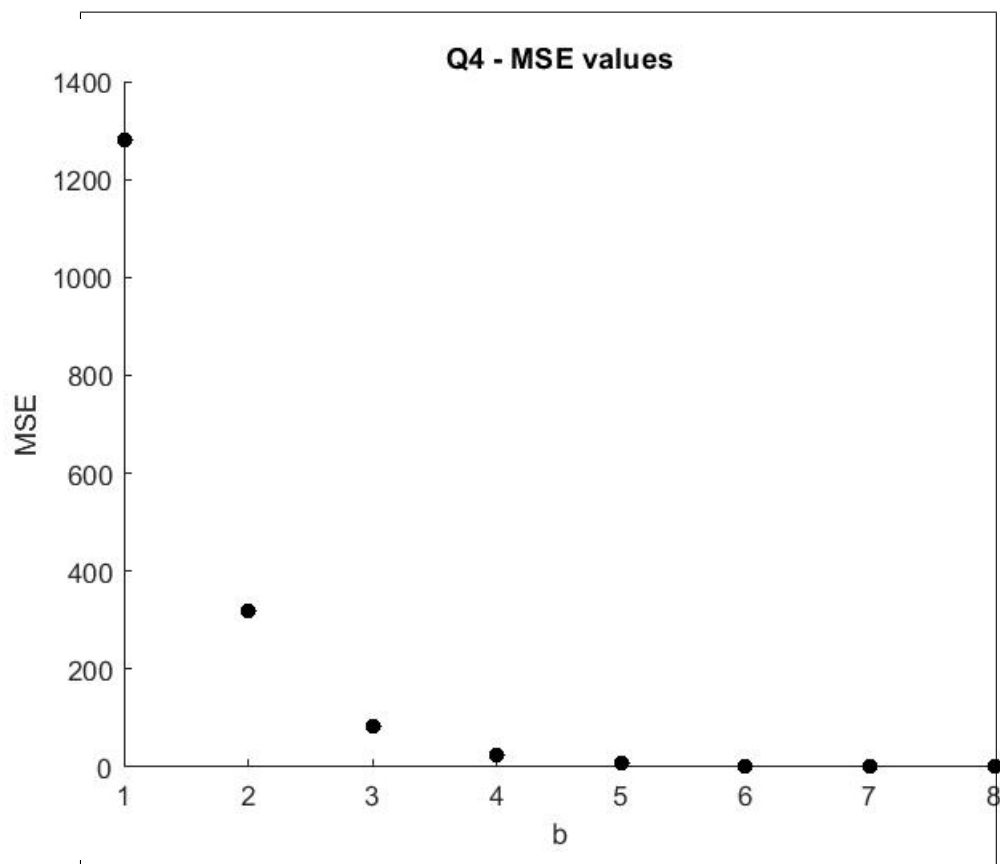
2. The decision and representation levels for representative b values:



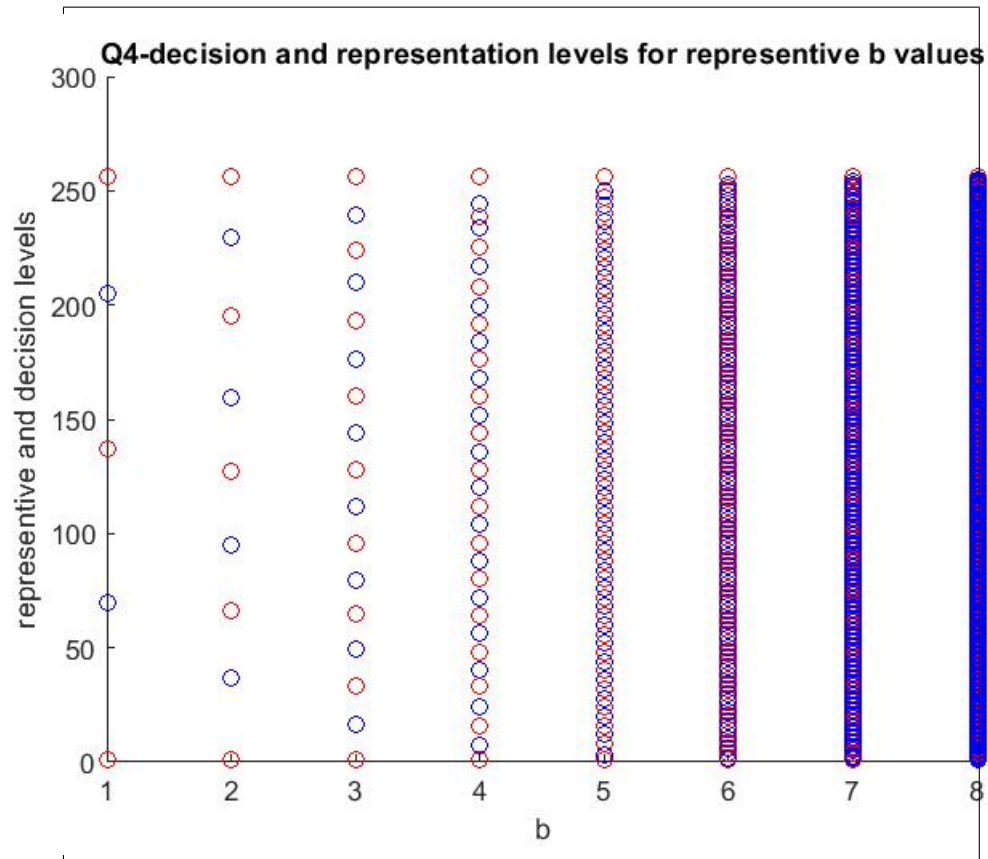
red-decision blue-representation 0.1: Figure

**Question 4 - Max-Lloyd unifrom quantization:**

1. MSE as a function of the bit budget for  $b=1, \dots, 8$ :



2. The decision and representation levels for representative  $b$  values:

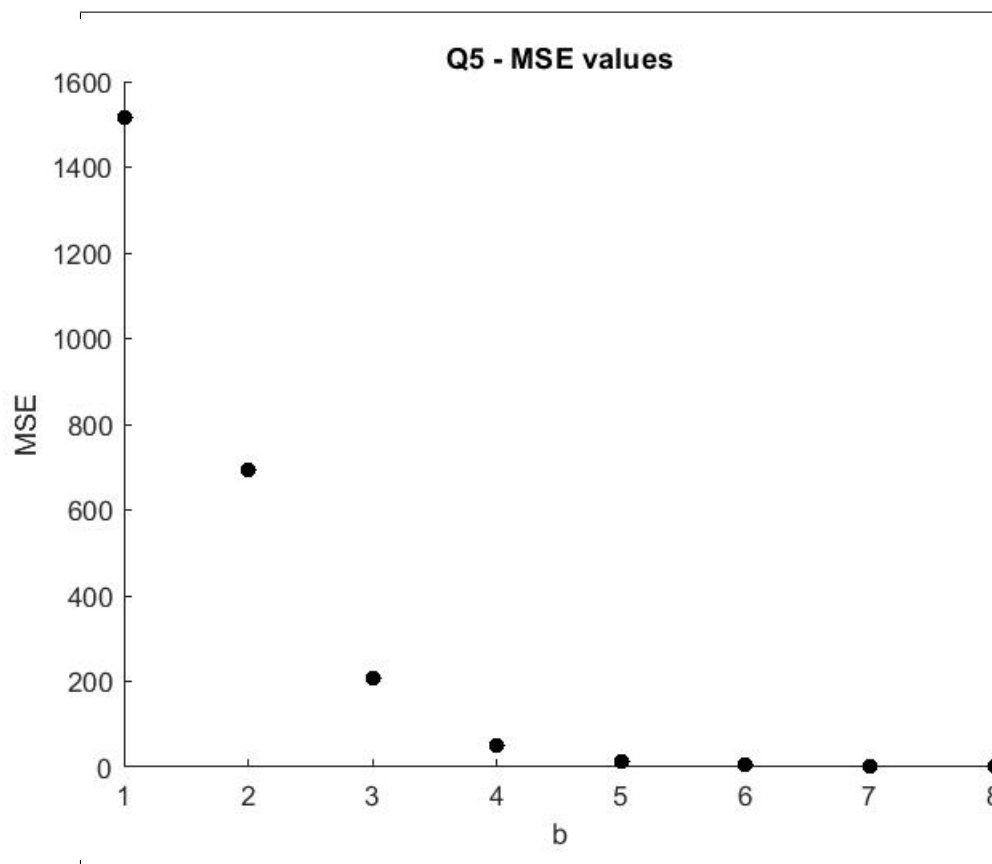


red-decision blue-representation 0.2: Figure

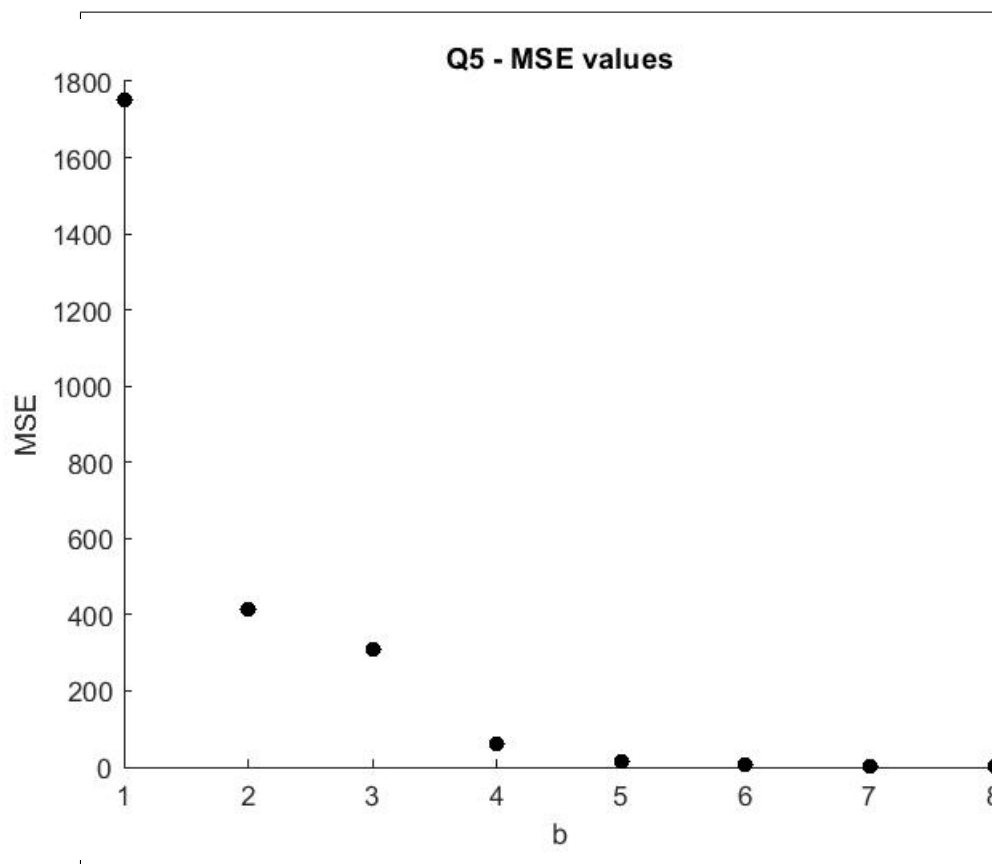
3. As we could expected, the MSE values obtained using Max-Lloyd are lower than the MSE values obtained by the unifrom quantization, as the Max-Lloyd algorithm should converge to the best representation and decision levels. And as we can see, the decision levels in the uniform quntization remain uniform, where as in the Max-Lloyd algorithm they do not.

**Question 5-Max-Lloyd random quantization:**

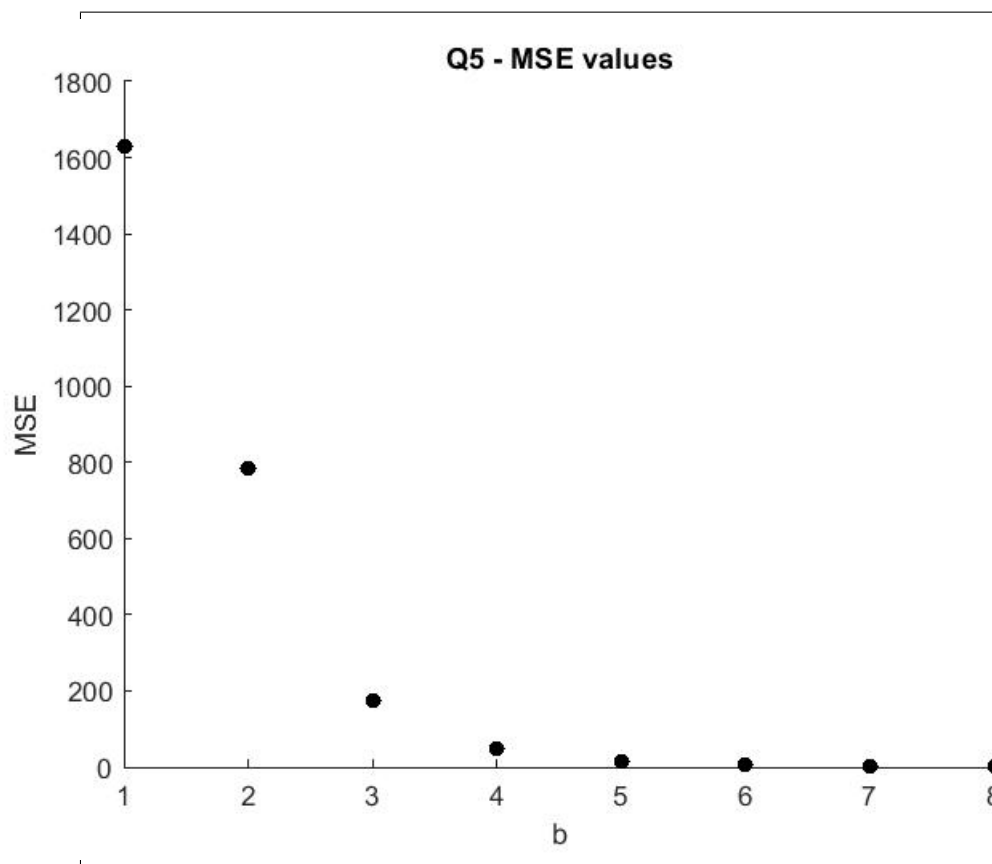
1. MSE as a function of the bit budget for  $b=1, \dots, 8$  for 5 random quantizations:



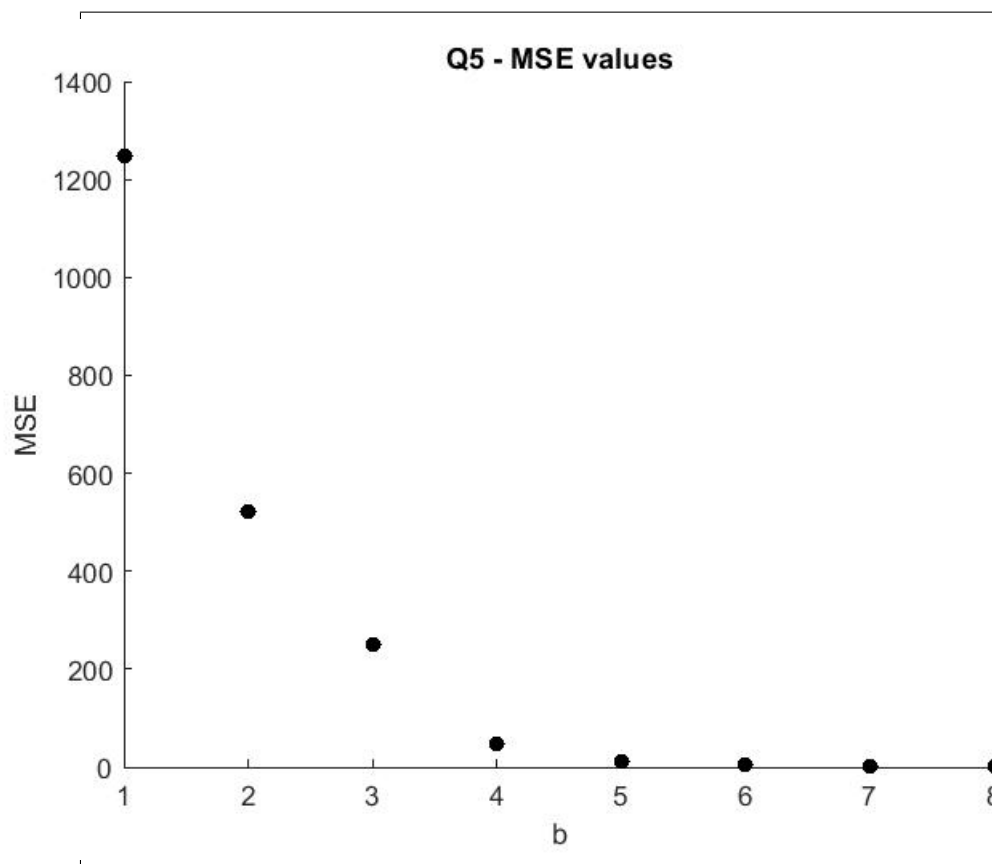
run no.1 0.3: Figure



run no.2 0.4: Figure

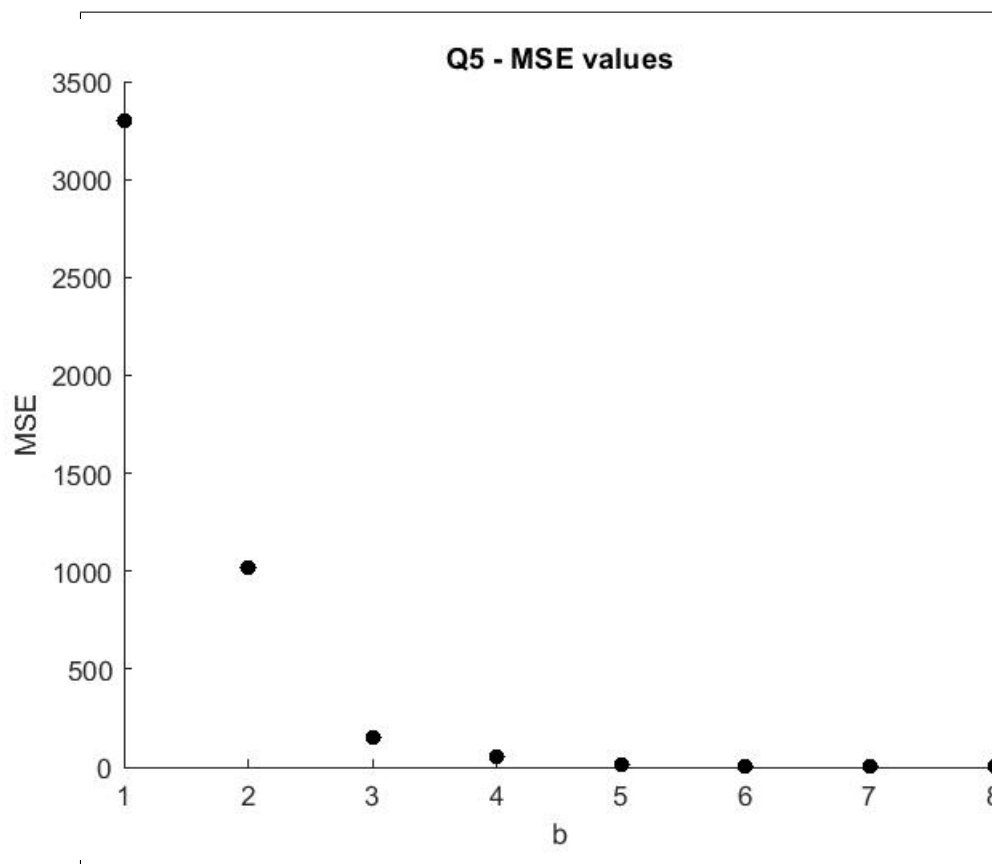


run no.3 0.5: Figure



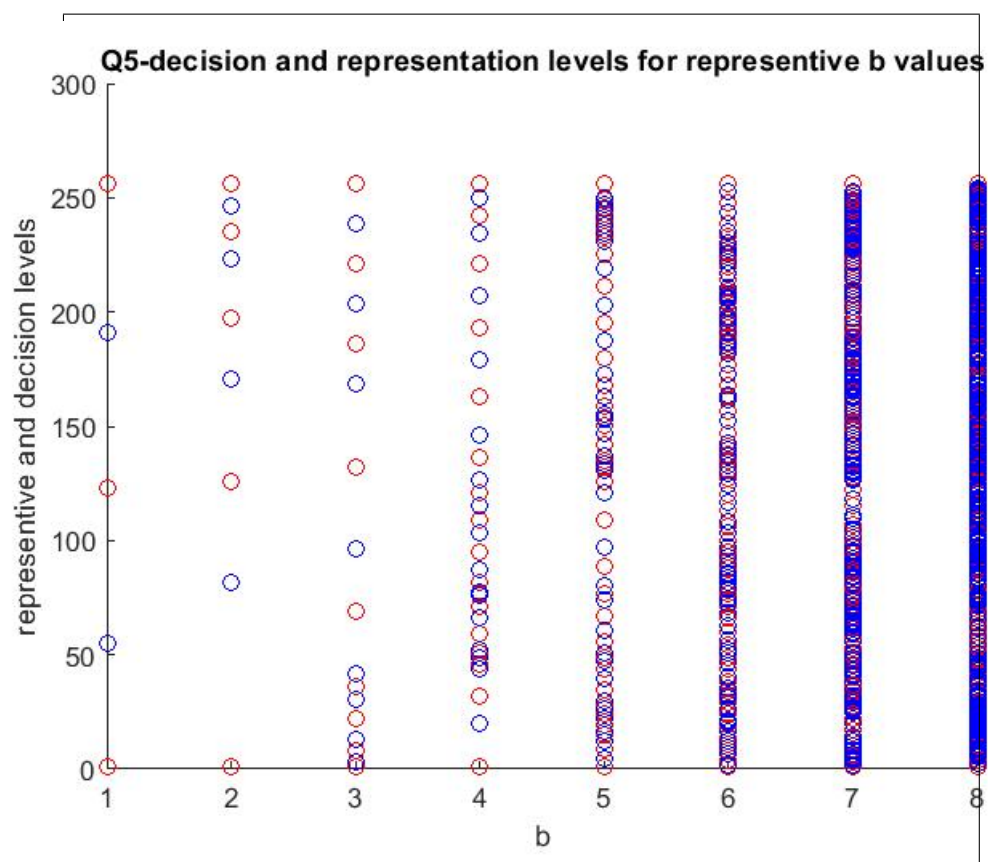
run no.4 0.6: Figure



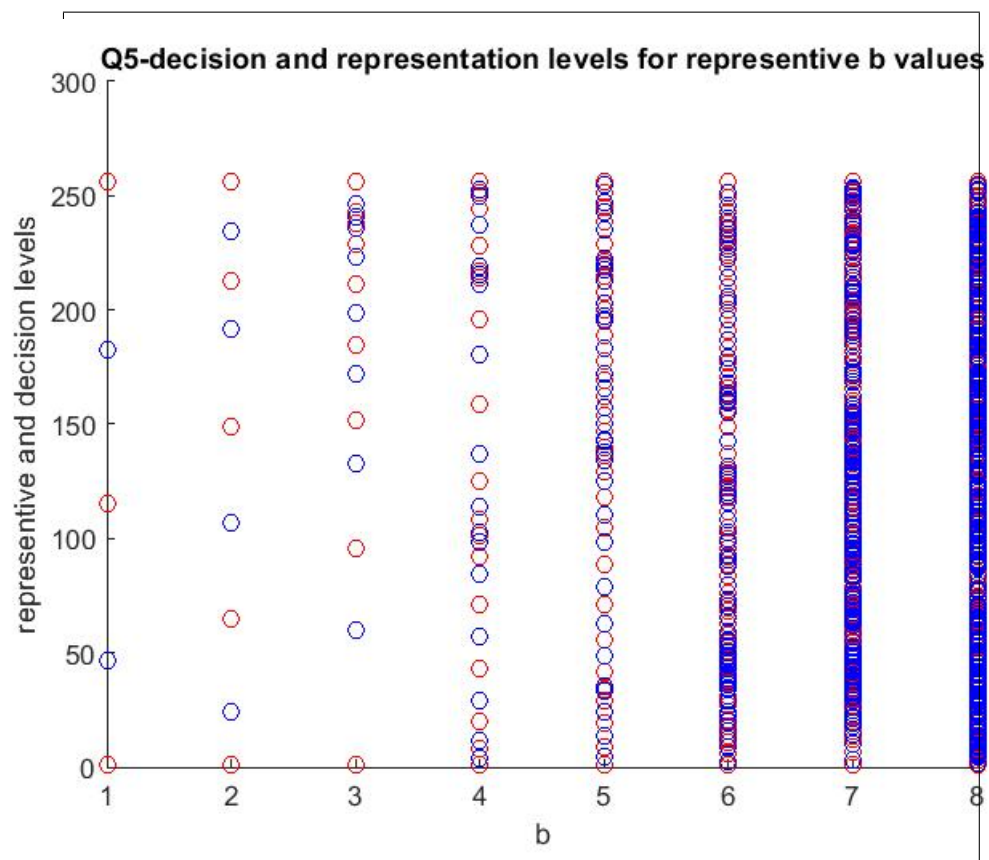


run no.5 0.7: Figure

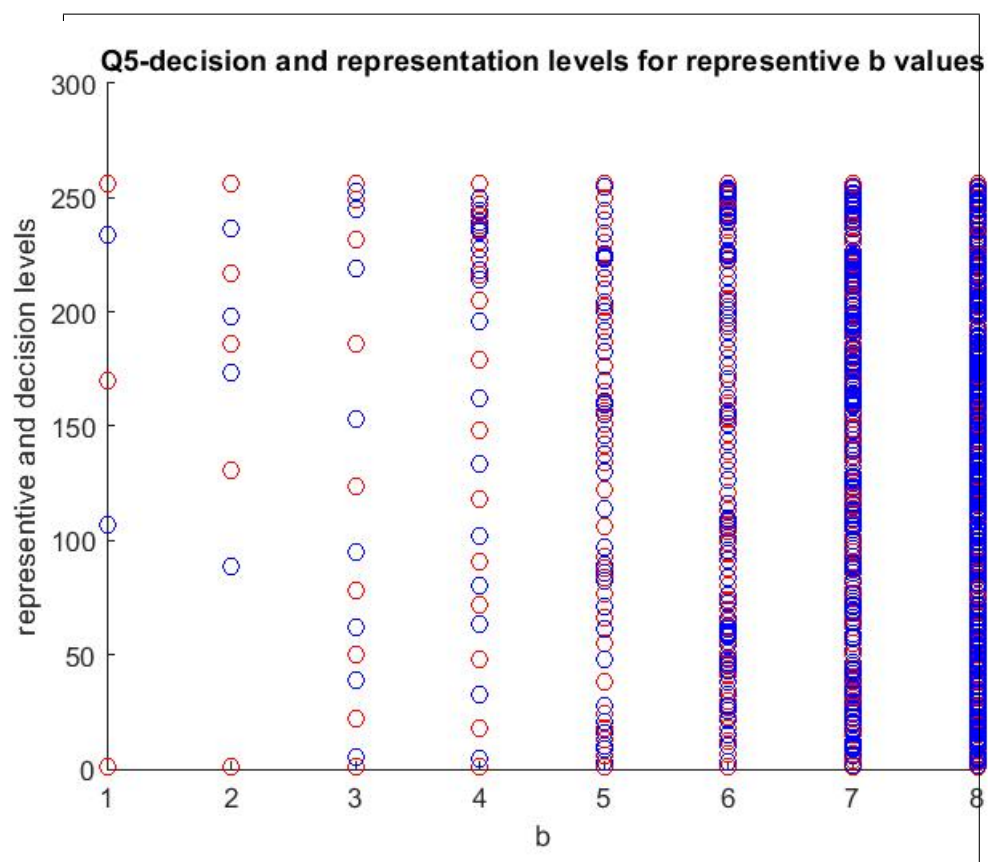
2. The decision and representation levels for representative  $b$  values, for 5 random quantizations:



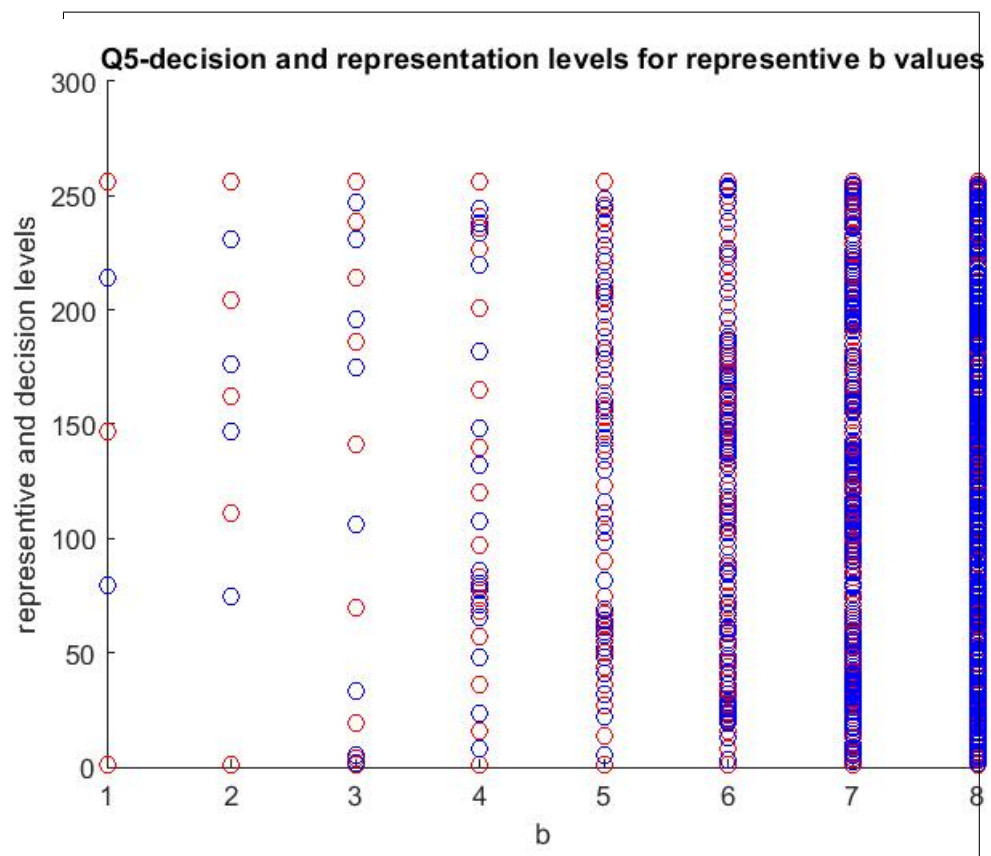
run no.1 0.8: Figure



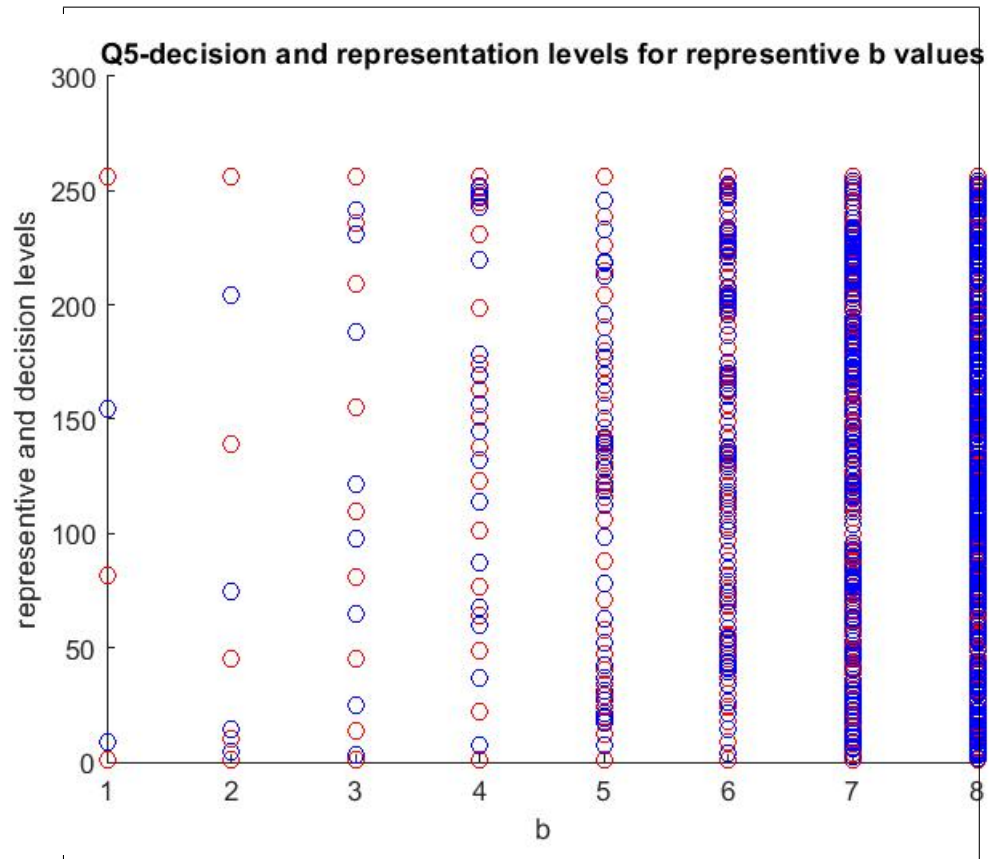
run no.2 0.9: Figure



run no.3 0.10: Figure



run no.4 0.11: Figure



run no.5 0.12: Figure

3. As we can see, the Max-Lloyd algorithm using uniform quantization obtained better results rather than starting with random quantization, as the MSE converged to zero in better performance.

### Question 4

- a. At step  $n$ , after computing the  $\{d\}_{i=0}^k$  values, the Max-Lloyd algorithm will compute the optimal representation levels based on these decision levels

according to the following formula (as seen in class):  $r_i^n = \frac{\int_{d_{i-1}^n}^{d_i^n} xp(x)dx}{\int_{d_{i-1}^n}^{d_i^n} p(x)dx}$ .

In the case of a uniform  $p(x)$ , we can simplify the expression:

$$\begin{aligned} r_i^n &= \frac{\int_{d_{i-1}^n}^{d_i^n} xp(x)dx}{\int_{d_{i-1}^n}^{d_i^n} p(x)dx} = \frac{\frac{1}{b-a} \int_{d_{i-1}^n}^{d_i^n} xdx}{\frac{1}{b-a} \int_{d_{i-1}^n}^{d_i^n} dx} = \frac{\frac{1}{2}x^2 \Big|_{d_{i-1}^n}^{d_i^n}}{d_i^n - d_{i-1}^n} = \frac{1}{2} \frac{(d_i^n)^2 - (d_{i-1}^n)^2}{d_i^n - d_{i-1}^n} \\ &= \frac{1}{2} \frac{(d_i^n + d_{i-1}^n)(d_i^n - d_{i-1}^n)}{d_i^n - d_{i-1}^n} = \frac{d_i^n + d_{i-1}^n}{2} \end{aligned}$$

For  $i = 0$ , we take  $r_0^n = a$ , and for  $i = k$  we take  $r_k^n = b$ .

We got, as expected, that every representation level is the average of the 2 decision levels next to it.

Now, the algorithm computes the  $d^{n+1}$  values:

$$d_i^{n+1} = \frac{r_i^n + r_{i+1}^n}{2} = \frac{\frac{d_{i-1}^n + d_i^n}{2} + \frac{d_i^n + d_{i+1}^n}{2}}{2} = \frac{d_{i-1}^n + 2d_i^n + d_{i+1}^n}{4}.$$

To finish step  $(n + 1)$ , the algorithm computes the  $r^{n+1}$  values: (we saw before that the integral formula becomes just an average)

$$\begin{aligned} r_i^{n+1} &= \frac{d_i^{n+1} + d_{i-1}^{n+1}}{2} = \frac{1}{2} \left( \frac{r_i^n + r_{i+1}^n}{2} + \frac{r_{i-1}^n + r_i^n}{2} \right) = \frac{r_{i-1}^n + 2r_i^n + r_{i+1}^n}{4} \\ &= \frac{1}{4} \left( \frac{d_{i-1}^n + d_{i-2}^n}{2} + 2 \frac{d_i^n + d_{i-1}^n}{2} + \frac{d_{i+1}^n + d_i^n}{2} \right) \\ &= \frac{d_{i-2}^n + 3d_{i-1}^n + 3d_i^n + d_{i+1}^n}{8} \end{aligned}$$

- b. The uniform quantization will be invariant through the Max-Lloyd algorithm:

$$\begin{aligned} d_i^n &= a + i \frac{(b-a)}{k}, \forall 0 \leq i \leq k \\ r_i^n &= a + \left(i - \frac{1}{2}\right) \frac{(b-a)}{k}, \forall 1 \leq i \leq k \end{aligned}$$

We will show that this quantization satisfies  $r_i^n = r_i^{n+1}$ ,  $d_j^n = d_j^{n+1}$ .

$$\begin{aligned} r_i^{n+1} &= \frac{d_{i-1}^{n+1} + d_i^{n+1}}{2} = \frac{1}{2} \left( \left( a + (i-1) \frac{b-a}{k} \right) + \left( a + i \frac{b-a}{k} \right) \right) \\ &= \frac{1}{2} \left( a + i \frac{b-a}{k} - \frac{b-a}{k} + a + i \frac{b-a}{k} \right) \\ &= \frac{1}{2} \left( 2a + (2i-1) \frac{b-a}{k} \right) = a + \left(i - \frac{1}{2}\right) \frac{b-a}{k} = r_i^n \end{aligned}$$

$$\begin{aligned}
d_i^{n+1} &= \frac{r_i^n + r_{i+1}^n}{2} = \frac{1}{2} \left( a + \left( i - \frac{1}{2} \right) \frac{b-a}{k} + a + \left( i + 1 - \frac{1}{2} \right) \frac{b-a}{k} \right) \\
&= \frac{1}{2} \left( 2a + \left( i + i + 1 - \frac{1}{2} - \frac{1}{2} \right) \frac{b-a}{k} \right) = \frac{1}{2} \left( 2a + 2i \frac{b-a}{k} \right) \\
&= a + i \frac{b-a}{k} = d_i^n
\end{aligned}$$

- c. We will formulate  $d_i^{n+1} = \frac{d_{i-1}^n + 2d_i^n + d_{i+1}^n}{4} = \frac{1}{4}d_{i-1}^n + \frac{1}{2}d_i^n + \frac{1}{4}d_{i+1}^n$  into matrix form:

$$d^{n+1} = Ad^n \rightarrow \begin{pmatrix} d_0^{n+1} \\ \vdots \\ d_k^{n+1} \end{pmatrix} = A \begin{pmatrix} d_0^n \\ \vdots \\ d_k^n \end{pmatrix}$$

The matrix  $A$  will be of size  $(k+1) \times (k+1)$ .

For every  $0 \leq i \leq k$ , we want  $d_i^{n+1} = A_i \begin{pmatrix} d_0^n \\ \vdots \\ d_k^n \end{pmatrix}$ , where  $A_i$  is the  $i$ 'th row of  $A$ .

$$\rightarrow \frac{1}{4}d_{i-1}^n + \frac{1}{2}d_i^n + \frac{1}{4}d_{i+1}^n = \sum_{j=1}^{k+1} A_{ij}d_j^n \rightarrow A_{i,j-1} = \frac{1}{4}, A_{i,j} = \frac{1}{2}, A_{i,j+1} = \frac{1}{4}$$

We will now look at the edge cases ( $i = 0$  or  $i = k$ ):

$$i = 0 \rightarrow d_0^n = a \rightarrow d_0^n = d_0^{n-1} \rightarrow A_{i,0} = 1$$

$$i = k \rightarrow d_k^n = b \rightarrow d_k^n = d_k^{n-1} \rightarrow A_{i,k} = 1$$

This result finds the constant  $c_1 = 1$ .

So, we know that  $A$  looks as follows: 
$$\begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & 0 & \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

Therefore, we also found the constant  $c_2 = \frac{1}{4}$ .

Also, we found that  $A$  can be written as  $A = \begin{pmatrix} 1 & 0 & \dots & 0 & 0 \\ \frac{1}{4} & & & & \vdots \\ 0 & & & & \vdots \\ \vdots & \tilde{B} & & & \vdots \\ 0 & & & & 0 \\ \vdots & & & & \frac{1}{4} \\ 0 & 0 & \dots & 0 & 1 \end{pmatrix}$

Where  $\tilde{B}$  is a tridiagonal Toeplitz matrix:  $\tilde{B} = \begin{pmatrix} \frac{1}{2} & \frac{1}{4} & \dots & \mathbf{0} & \mathbf{0} \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} & & \vdots \\ \vdots & & \ddots & & \vdots \\ \mathbf{0} & & \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ \mathbf{0} & \mathbf{0} & \dots & \frac{1}{4} & \frac{1}{2} \end{pmatrix}$

- d. We will expand the expression we found for  $d^n$ :



$$d^n = Ad^{n-1} = A(Ad^{n-2}) = A^2d^{n-2} = \dots = A^i d^{n-i} = \dots = A^n d^0 \\ \rightarrow d^n = A^n d^0$$

- e. Since  $\tilde{B}$  has the same values in the diagonals above and below its main diagonal, it is symmetric, and therefore diagonalizable (every real symmetric matrix is diagonalizable).
- f. We will start by showing that  $x_0$  and  $x_1$  are eigenvectors of  $A$ , associated to the eigenvalue 1:

$$Ax_0 = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & 0 & \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ 2 \\ \vdots \\ 2(k-1) \\ 2k \end{pmatrix} \\ = \begin{pmatrix} 1 \cdot 0 \\ \frac{1}{4} \cdot 0 + \frac{1}{2} \cdot 2 + \frac{1}{4} \cdot 4 \\ \vdots \\ \frac{1}{4} \cdot 2(j-2) + \frac{1}{2} \cdot 2(j-1) + \frac{1}{4} \cdot 2j \\ \vdots \\ \frac{1}{4} \cdot 2(k-2) + \frac{1}{2} \cdot 2(k-1) + \frac{1}{4} \cdot 2k \\ 1 \cdot 2k \end{pmatrix} = \begin{pmatrix} 0 \\ 2 \\ \vdots \\ 2(j-1) \\ \vdots \\ 2(k-1) \\ 2k \end{pmatrix} \\ = 1 \cdot x_0$$

$$Ax_1 = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & 0 & \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 2k \\ 2(k-1) \\ \vdots \\ 2 \\ 0 \end{pmatrix} = \\ \begin{pmatrix} 1 \cdot 2k \\ \frac{1}{4} \cdot 2k + \frac{1}{2} \cdot 2(k-1) + \frac{1}{4} \cdot 2(k-2) \\ \vdots \\ \frac{1}{4} \cdot 2(k-j+2) + \frac{1}{2} \cdot 2(k-j+1) + \frac{1}{4} \cdot 2(k-j) \\ \vdots \\ \frac{1}{4} \cdot 4 + \frac{1}{2} \cdot 2 + \frac{1}{4} \cdot 0 \\ 1 \cdot 0 \end{pmatrix} = \begin{pmatrix} 2k \\ 2(k-1) \\ \vdots \\ 2(k-j+1) \\ \vdots \\ 2 \\ 0 \end{pmatrix} = \\ 1 \cdot x_1$$

We will now show that  $x_0$  and  $x_1$  are linearly independent:

Let  $a, b$  be constants such that  $ax_0 + bx_1 = \vec{0}$ :

$$\rightarrow a \begin{pmatrix} 0 \\ 2 \\ \vdots \\ 2(k-1) \\ 2k \end{pmatrix} + b \begin{pmatrix} 2k \\ 2(k-1) \\ \vdots \\ 2 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix}$$

$$\rightarrow \begin{pmatrix} 0 \cdot a + 2bk \\ 2a + 2b(k-1) \\ \vdots \\ 2a(k-1) + 2b \\ 2ak + 0 \cdot b \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix}$$

From the first row we get:  $2bk = 0 \rightarrow b = 0$ .

From the last row we get:  $2ak = 0 \rightarrow a = 0$ .

We found that the only  $a, b$  that satisfy  $ax_0 + bx_1 = \vec{0}$  are  $a = b = 0$ , and so  $x_0$  and  $x_1$  are linearly independent.

g. We will use the following observation:

If  $\tilde{B}e_{\lambda_1} = \lambda_i e_{\lambda_1}$  then

$$A \begin{pmatrix} v1 \\ | \\ e_{\lambda_1} \\ | \\ v_{k+1} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ 1 & 1 & 1 & \dots & 0 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & 0 & \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} v1 \\ | \\ e_{\lambda_1} \\ | \\ v_{k+1} \end{pmatrix}$$

$$= \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ \frac{1}{4} & & & & \vdots \\ 0 & & & & \vdots \\ \vdots & & \tilde{B} & & \vdots \\ 0 & & & & 0 \\ \vdots & & & & \frac{1}{4} \\ 0 & 0 & \dots & 0 & 1 \end{pmatrix} \begin{pmatrix} v1 \\ | \\ e_{\lambda_1} \\ | \\ v_{k+1} \end{pmatrix} =$$

$$\left( \begin{pmatrix} 1 & 0 & \dots & 0 & 0 \\ \frac{1}{4} & & & & \vdots \\ 0 & & & & \vdots \\ \vdots & & 0 & & \vdots \\ 0 & & & 0 & \frac{1}{4} \\ \vdots & & & & 1 \end{pmatrix} + \begin{pmatrix} 0 & 0 & \dots & 0 & 0 \\ 0 & & & & \vdots \\ 0 & & & & \vdots \\ \vdots & & \tilde{B} & & \vdots \\ 0 & & & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 \end{pmatrix} \right) \begin{pmatrix} v1 \\ | \\ e_{\lambda_1} \\ | \\ v_{k+1} \end{pmatrix} =$$

$$\begin{pmatrix} 1 & 0 & \dots & 0 & 0 \\ 1 & & & & \vdots \\ \frac{1}{4} & & & & \vdots \\ 0 & & & & \vdots \\ \vdots & 0 & & & 0 \\ 0 & & & & \frac{1}{4} \\ \vdots & & & & \vdots \\ 0 & 0 & \dots & 0 & 1 \end{pmatrix} \begin{pmatrix} v_1 \\ | \\ e_{\lambda_1} \\ | \\ v_{k+1} \end{pmatrix} + \begin{pmatrix} 0 & 0 & \dots & 0 & 0 \\ 0 & & & & \vdots \\ 0 & & \tilde{B} & & \vdots \\ \vdots & & & & 0 \\ 0 & 0 & \dots & 0 & 0 \end{pmatrix} \begin{pmatrix} v_1 \\ | \\ e_{\lambda_1} \\ | \\ v_{k+1} \end{pmatrix} = \begin{pmatrix} v_1 \\ \frac{1}{4}v_1 \\ 0 \\ \vdots \\ 0 \\ \frac{1}{4}v_{k+1} \\ v_{k+1} \end{pmatrix} + \begin{pmatrix} 0 \\ | \\ \tilde{B}e_{\lambda_1} \\ | \\ 0 \end{pmatrix} = \begin{pmatrix} v_1 \\ \frac{1}{4}v_1 \\ 0 \\ \vdots \\ 0 \\ \frac{1}{4}v_{k+1} \\ v_{k+1} \end{pmatrix} + \begin{pmatrix} 0 \\ | \\ \lambda_i e_{\lambda_1} \\ | \\ 0 \end{pmatrix}$$

So, if we pick a vector with  $v_1 = v_{k+1} = 0$ , then for every eigenvector  $e_{\lambda_i}$  of

$\tilde{B}$ , the vector  $\begin{pmatrix} 0 \\ | \\ e_{\lambda_i} \\ | \\ 0 \end{pmatrix}$  is an eigenvector of  $A$  with the eigenvalue  $\lambda_i$ :

$$A \begin{pmatrix} 0 \\ | \\ e_{\lambda_i} \\ | \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ \frac{1}{4} \cdot 0 \\ 0 \\ \vdots \\ 0 \\ \frac{1}{4} \cdot 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ | \\ \lambda_i e_{\lambda_1} \\ | \\ 0 \end{pmatrix} = \lambda_i \begin{pmatrix} 0 \\ | \\ e_{\lambda_1} \\ | \\ 0 \end{pmatrix}$$

In this way we found  $(k - 1)$  eigenvectors of  $A$ , where every the multiplicity

of the eigenvector  $\begin{pmatrix} 0 \\ | \\ e_{\lambda_i} \\ | \\ 0 \end{pmatrix}$  is the same as the multiplicity of  $e_{\lambda_i}$  as an

eigenvector of  $\tilde{B}$ . We need to find 2 more eigenvectors. Fortunately, we

found these 2 eigenvectors in section (f), which are not in the form  $\begin{pmatrix} 0 \\ | \\ \bar{v} \\ | \\ 0 \end{pmatrix}$ , so

we know that these are the last 2 eigenvectors, and they share the eigenvalue of 1.

$\tilde{B}$  is diagonalizable and so the algebraic multiplicity of its eigenvalues equals to their geometric multiplicity. Let's have a look at the characteristic polynomial of  $A$ :

$$\begin{aligned}
 p_A(\lambda) = \det(\lambda I - A) &= \det \begin{pmatrix} \lambda - 1 & 0 & \cdots & 0 & 0 \\ \frac{1}{4} & & & & \vdots \\ 0 & & & & \vdots \\ \vdots & & \lambda I - \tilde{B} & & \vdots \\ 0 & & & & 0 \\ \vdots & & & & \frac{1}{4} \\ 0 & 0 & \cdots & 0 & \lambda - 1 \end{pmatrix} = \\
 &= (\lambda - 1) \cdot \det \begin{pmatrix} & & & \vdots \\ & & & \vdots \\ & & \lambda I - \tilde{B} & \vdots \\ & & & 0 \\ & & & \frac{1}{4} \\ 0 & \cdots & 0 & 1 \end{pmatrix} = \\
 &= \pm(\lambda - 1)(\lambda - 1) \cdot \det \begin{pmatrix} & & & \vdots \\ & & & \vdots \\ & & \lambda I - \tilde{B} & \vdots \\ & & & 0 \\ & & & \frac{1}{4} \\ 0 & \cdots & 0 & 1 \end{pmatrix} = \pm(\lambda - 1)^2 \cdot p_{\tilde{B}}(\lambda)
 \end{aligned}$$

From this polynomial we see that  $\lambda = 1$  is an eigenvalue of algebraic multiplicity of 2, and we previously saw that it has 2 corresponding linearly independent eigenvectors, so its geometric multiplicity is also 2.

To conclude, we found out that the algebraic multiplicities of the eigenvalues of  $A$  equal to their geometric multiplicities, and so  $A$  is diagonalizable.

$$\text{h. } B = 4\tilde{B} = 4 \begin{pmatrix} \frac{1}{2} & \frac{1}{4} & \cdots & 0 & 0 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} & & \vdots \\ \vdots & & \ddots & & \vdots \\ 0 & & \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ 0 & 0 & \cdots & \frac{1}{4} & \frac{1}{2} \end{pmatrix} = \begin{pmatrix} 2 & 1 & \cdots & 0 & 0 \\ 1 & 2 & 1 & & \vdots \\ \vdots & & \ddots & & \vdots \\ 0 & & 1 & 2 & 1 \\ 0 & 0 & \cdots & 1 & 2 \end{pmatrix}.$$

The eigenvectors of  $B$  are the same as those of  $\tilde{B}$ :  $e_{\lambda_1}, \dots, e_{\lambda_{k-1}}$ .

However the eigenvalues are slightly different: for every eigenvector  $e_{\lambda_i}$ , the eigenvalue of  $B$  is  $4\lambda_i$ :

$$B e_{\lambda_i} = (4\tilde{B}) e_{\lambda_i} = 4 \cdot (\tilde{B} e_{\lambda_i}) = 4 \cdot (\lambda_i e_{\lambda_i}) \rightarrow B e_{\lambda_i} = (4\lambda_i) e_{\lambda_i}$$

$$\text{i. } \chi_{B_{k-1}}(X) = \det(B_{k-1} - XI) =$$

$$\det(B - XI) = \det \begin{pmatrix} 2 - X & 1 & \cdots & 0 & 0 \\ 1 & 2 - X & 1 & & \vdots \\ \vdots & & \ddots & & \vdots \\ 0 & & 1 & 2 - X & 1 \\ 0 & 0 & \cdots & 1 & 2 - X \end{pmatrix}_{(k-1) \times (k-1)} =$$

$$\begin{aligned}
& (2-X) \det \begin{pmatrix} 2-X & 1 & \cdots & 0 \\ \vdots & \ddots & & \vdots \\ 0 & \cdots & 1 & 2-X \end{pmatrix}_{(k-2) \times (k-2)} \\
& - \det \begin{pmatrix} 1 & 1 & 0 & \cdots & 0 \\ 0 & 2-X & 1 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 2-X & 1 \\ 0 & 0 & \cdots & 1 & 2-X \end{pmatrix}_{(k-2) \times (k-2)} = \\
& (2-X) \cdot \chi B_{k-2}(X) - 1 \cdot \det \begin{pmatrix} 2-X & 1 & \cdots & 0 \\ \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 2-X & 1 \\ 0 & \cdots & 1 & 2-X \end{pmatrix}_{(k-3) \times (k-3)} = \\
& (2-X) \cdot \chi B_{k-2}(X) - \chi B_{k-3}(X)
\end{aligned}$$

Now we will calculate  $\chi B_0(X)$  and  $\chi B_1(X)$ :

$$\chi B_0(X) = \det(B_0 - XI) = \det\left(\begin{pmatrix} & \\ & \end{pmatrix} - XI_{0 \times 0}\right) = \det\left(\begin{pmatrix} & \\ & \end{pmatrix}\right) = 1$$

$$\chi B_1(X) = \det(B_1 - XI) = \det((2) - XI) = \det((2-X)) = 2-X$$

- j. We can rewrite the recursive formula from the last section as:  
(for  $i \in \{2, \dots, k-1\}$ )

$$\chi B_i(X) = (2-X)\chi B_{i-1}(X) - \chi B_{i-2}(X)$$

So, if we will define  $2Y = 2 - 2X \rightarrow 2 - 2Y = 2X$ , then we can get that

$$\begin{aligned}
\chi B_i(2-2X) &= \chi B_i(2Y) = (2-2Y)\chi B_{i-1}(2Y) - \chi B_{i-2}(2Y) = \\
&= 2X \cdot \chi B_{i-1}(2-2X) - \chi B_{i-2}(2-2X)
\end{aligned}$$

Therefore, if we will define  $Q_i(X) = \chi B_i(2-2X)$ , we will get the wanted recursive relation: (for every  $i \in \{2, \dots, k-1\}$ )

$$\begin{aligned}
Q_i(X) &= \chi B_i(2-2X) = 2X \cdot \chi B_{i-1}(2-2X) - \chi B_{i-2}(2-2X) = \\
&= 2XQ_{i-1}(X) - Q_{i-2}(X)
\end{aligned}$$

$Q_i(X)$  and  $\chi B_i(X)$  are easily expressed by one another:

$$- Q_i(X) = \chi B_i(2-2X)$$

$$- 2-2X = Y \rightarrow 2X = 2-Y \rightarrow X = 1 - \frac{Y}{2} \rightarrow \chi B_i(X) = Q_i\left(1 - \frac{X}{2}\right)$$

We will now find  $Q_0(X)$  and  $Q_1(X)$  using the relation we found:

$$Q_0(X) = \chi B_0(2-2X) = 1 \Big|_{2-2X} = 1$$

$$Q_1(X) = \chi B_1(2-2X) = (2-X) \Big|_{2-2X} = 2 - (2-2X) = 2X$$

- k. Let's calculate the next few  $Q_i$ 's:

$$Q_2(X) = 2X \cdot Q_1(X) - Q_0(X) = 2X \cdot 2X - 1 = 4X^2 - 1$$

$$Q_3(X) = 2X \cdot Q_2(X) - Q_1(X) = 2X \cdot (4X^2 - 1) - 2X = 8X^3 - 4X$$

The  $Q_i(X)$  series is the **Chebyshev polynomials of the second kind**.

- l. For every  $i \in \{0, \dots, k-1\}$ , the degree of  $Q_i(X)$  is  $i$ . We will prove that by induction on  $i$ :

Base: For  $i = 0$ , we know that  $Q_0(X) = 1$ , which is of degree  $i = 0$ .

For  $i = 1$ , we know that  $Q_1(X) = 2X$ , which is of degree  $i = 1$ .

Step: Let  $i$  be in the range of  $[2, k - 1]$ . We will assume that the degree of  $Q_{i-1}(X)$  is  $(i - 1)$ , and that the degree of  $Q_{i-2}(X)$  is  $(i - 2)$ , and then we will show that the degree of  $Q_i(X)$  is  $i$ .

The degree of  $2X \cdot Q_{i-1}(X)$  is then necessarily  $i$ , which is strictly larger than  $(i - 2)$ . We will use the recursive definition of  $Q_i(X)$ :

$Q_i(X) = 2X \cdot Q_{i-1}(X) - Q_{i-2}(X)$ , and therefore the degree of  $Q_i(X)$  is the same as  $2X \cdot Q_{i-1}(X)$ , which is  $i$ .

According to a theorem from linear algebra, we know that since  $Q_i(X)$  is of degree  $i$ , it has at most  $i$  roots.

m. We will prove that  $Q_i(\cos(\theta)) = \frac{\sin((i+1)\theta)}{\sin(\theta)}$  by induction on  $i$ :

Base: For  $i = 0$ , we know that  $Q_0(\cos(\theta)) = 1 = \frac{\sin(\theta)}{\sin(\theta)} = \frac{\sin((0+1)\theta)}{\sin(\theta)}$ .

For  $i = 1$ , we know that  $Q_1(\cos(\theta)) = 2 \cos(\theta) = \frac{\sin(2\theta)}{\sin(\theta)} = \frac{\sin((1+1)\theta)}{\sin(\theta)}$ .

(We used the following identity:

$$\sin(2\theta) = 2 \sin(\theta) \cos(\theta) \rightarrow 2 \cos(\theta) = \frac{\sin(2\theta)}{\sin(\theta)}$$

Step: We assume that  $Q_{i-1}(\cos(\theta)) = \frac{\sin(((i-1)+1)\theta)}{\sin(\theta)} = \frac{\sin(i\theta)}{\sin(\theta)}$ , and that

$Q_{i-2}(\cos(\theta)) = \frac{\sin(((i-2)+1)\theta)}{\sin(\theta)} = \frac{\sin((i-1)\theta)}{\sin(\theta)}$ . Now we will show that the statement holds for  $Q_i$ :

$$\begin{aligned} Q_i(\cos(\theta)) &= 2 \cos(\theta) \cdot Q_{i-1}(\cos(\theta)) - Q_{i-2}(\cos(\theta)) = \\ 2 \cos(\theta) \cdot \frac{\sin(i\theta)}{\sin(\theta)} - \frac{\sin((i-1)\theta)}{\sin(\theta)} &= \frac{2 \cos(\theta) \sin(i\theta) - \sin((i-1)\theta)}{\sin(\theta)} = \\ \frac{2 \cos(\theta) \sin(i\theta) - \sin(i\theta - \theta)}{\sin(\theta)} &= \\ = \frac{2 \cos(\theta) \sin(i\theta) - (\sin(i\theta) \cos(\theta) - \cos(i\theta) \sin(\theta))}{\sin(\theta)} &= \\ = \frac{2 \sin(i\theta) \cos(\theta) - \sin(i\theta) \cos(\theta) + \sin(\theta) \cos(i\theta)}{\sin(\theta)} &= \\ \frac{\sin(i\theta) \cos(\theta) + \sin(\theta) \cos(i\theta)}{\sin(\theta)} &= \frac{\sin(i\theta + \theta)}{\sin(\theta)} = \frac{\sin((i+1)\theta)}{\sin(\theta)} \end{aligned}$$

n. We will use the equivalent expression from the last section in order to find roots of  $Q_i(X)$  in the range  $[-1, 1]$ :  $\frac{\sin((i+1)\theta)}{\sin(\theta)} = 0$

The denominator can't be 0, so we can rule out  $\theta = 0$  and  $\theta = \pi$ .

Now we are left with  $\sin((i+1)\theta) = 0 \rightarrow (i+1)\theta = \pi m$  (for  $m \in \mathbb{N}$ )

$$\rightarrow \theta = \frac{m}{i+1} \pi$$

For every  $m \in \{1, 2, \dots, i\}$  we get a valid solution for  $\theta$  which yields a unique  $X = \cos(\theta)$  solution to  $Q_i(X) = 0$ . The values of  $\cos(\theta)$  are unique for these

solutions because the  $\theta$  solutions are unique, and are in the range

$[\frac{1}{i+1}\pi, \frac{i}{i+1}\pi]$ , and  $\cos(\theta)$  is strictly decreasing in that range.

We found  $i$  unique roots of  $Q_i(X)$ , which is the upper bound to the number of roots, and so we know that there are no other roots.

To conclude, for every  $m \in \{1, 2, \dots, i\}$ ,  $X = \cos(\frac{m}{i+1}\pi)$  is a root of multiplicity 1 of the polynomial  $Q_i(X)$ .

From the relation  $\chi_{B_i}(X) = Q_i(1 - \frac{X}{2})$ , we know that for every root

$X = \cos(\frac{m}{i+1}\pi)$  of  $Q_i(X)$ ,  $1 - \frac{X}{2} = 1 - \frac{1}{2}\cos(\frac{m}{i+1}\pi)$  is a root of  $\chi_{B_i}$ .

Since  $f(X) = 1 - \frac{X}{2}$  is a one-to-one function, we get that there are  $i$  unique roots of  $\chi_{B_i}$ , each of multiplicity 1:  $1 - \frac{1}{2}\cos(\frac{m}{i+1}\pi)$  for every  $m \in \{1, \dots, i\}$ .

- o. The eigenvalues of  $B$  are the roots of the characteristic polynomial  $\chi_{B_{k-1}}$ .

So, the eigenvalues of  $B$  are  $(1 - \frac{1}{2}\cos(\frac{m}{k}\pi))$  for every  $m \in \{1, \dots, k-1\}$ .

In section (h) we showed that for every eigenvalue  $\lambda$  of  $\tilde{B}$ ,  $4\lambda$  is an eigenvalue of  $B$ , therefore, for every eigenvalue  $\lambda$  of  $B$ ,  $\frac{1}{4}\lambda$  is an eigenvalue of  $\tilde{B}$ :

For every  $m \in \{1, \dots, k-1\}$ ,  $\frac{1}{4}(1 - \frac{1}{2}\cos(\frac{m}{k}\pi)) = \frac{1}{4} - \frac{1}{8}\cos(\frac{m}{k}\pi)$  is an eigenvalue of  $\tilde{B}$ .

In section (g) we saw that the eigenvalues of  $A$  are 1, with multiplicity 2, and all the eigenvalues of  $\tilde{B}$ , with multiplicity 1.

- p. If we remember the definition of  $A$ , it is the matrix transformation of  $d^{n+1} = Ad^n$ , and in section (d) we reshaped it to  $d^{n+1} = A^n d^0$ .

In section (g) we proved that  $A$  is diagonalizable, and so we can express it by its eigendecomposition:  $A = QSQ^{-1}$ .

$$\text{Where } S = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & \frac{1}{4} - \frac{1}{8}\cos(\frac{1}{k}\pi) & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \frac{1}{4} - \frac{1}{8}\cos(\frac{k-1}{k}\pi) \end{pmatrix}$$

(a diagonal matrix with eigenvalues at the diagonal).

$$\text{And } Q = \begin{pmatrix} 0 & 2k & & & \\ 2 & 2(k-1) & & & \\ \vdots & \vdots & & & \\ 2(k-1) & 2 & & & \\ 2k & 0 & & & \end{pmatrix} \begin{array}{c} | \\ | \\ v_1 \dots v_{k-1} \\ | \end{array} \text{ where } v_i$$

( $i \in \{1, \dots, k-1\}$ ) is the eigenvector of the eigenvalue  $\frac{1}{4} - \frac{1}{8}\cos(\frac{i}{k}\pi)$ .

As said, we can express  $A$  as  $A = QSQ^{-1} \rightarrow A^n = QS^nQ^{-1}$ .

$S$  is a diagonal matrix and therefore  $S^n$  is:

$$S^n = \begin{pmatrix} 1^n & 0 & 0 & \dots & 0 \\ 0 & 1^n & 0 & \dots & 0 \\ 0 & 0 & \left(\frac{1}{4} - \frac{1}{8} \cos\left(\frac{1}{k}\pi\right)\right)^n & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \left(\frac{1}{4} - \frac{1}{8} \cos\left(\frac{k-1}{k}\pi\right)\right)^n \end{pmatrix}$$

We will look at every eigenvalue of the form  $\left(\frac{1}{4} - \frac{1}{8} \cos\left(\frac{i}{k}\pi\right)\right)$ , since

$|\cos(?)| \leq 1$ , we know that  $\frac{1}{4} - \frac{1}{8} \cos\left(\frac{i}{k}\pi\right) \in \left[\frac{1}{4} - \frac{1}{8}, \frac{1}{4} + \frac{1}{8}\right]$  and in particular

$\left|\frac{1}{4} - \frac{1}{8} \cos\left(\frac{i}{k}\pi\right)\right| < 1$ , and so we know that  $\lim_{n \rightarrow \infty} \left(\frac{1}{4} - \frac{1}{8} \cos\left(\frac{i}{k}\pi\right)\right)^n = 0$ .

$$\text{Therefore we know that } \lim_{n \rightarrow \infty} S^n = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 0 \end{pmatrix}.$$

$$\text{Let's denote } Q^{-1} = \begin{pmatrix} - & r_1 & - \\ - & r_2 & - \\ & \vdots & \\ - & r_{k+1} & - \end{pmatrix}.$$

$$\begin{aligned} \rightarrow \lim_{n \rightarrow \infty} A^n &= Q \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 0 \end{pmatrix} Q^{-1} \\ &= \left( Q \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 0 \end{pmatrix} \right) \left( \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 0 \end{pmatrix} Q^{-1} \right) \\ &= \begin{pmatrix} 0 & 2k & 0 & \dots & 0 \\ 2 & 2(k-1) & 0 & \dots & 0 \\ \vdots & \vdots & 0 & \dots & 0 \\ 2(k-1) & 2 & \vdots & \ddots & \vdots \\ 2k & 0 & 0 & \dots & 0 \end{pmatrix} \begin{pmatrix} - & r_1 & - \\ - & r_2 & - \\ - & 0 & - \\ & \vdots & \\ - & 0 & - \end{pmatrix} \\ &= \begin{pmatrix} - & 2k \cdot r_2 & - \\ - & 2r_1 + 2(k-1) \cdot r_2 & - \\ & \vdots & \\ - & 2(k-1) \cdot r_1 + 2 \cdot r_2 & - \\ - & 2k \cdot r_1 & - \end{pmatrix} \end{aligned}$$

We don't know the exact values of  $r_1$  and  $r_2$ , but we proved that  $A^n$  converges as  $n \rightarrow \infty$ .



To conclude,

$$\lim_{n \rightarrow \infty} d^n = \lim_{n \rightarrow \infty} A^n d^0 = \begin{pmatrix} - & 2k \cdot r_2 & - \\ - & 2r_1 + 2(k-1) \cdot r_2 & - \\ & \vdots & \\ - & 2(k-1) \cdot r_1 + 2 \cdot r_2 & - \\ - & 2k \cdot r_1 & - \end{pmatrix} d^0.$$

q. We are looking for vectors such that  $vA = v \rightarrow A^t v^t = v^t$ .

$$\text{Let's examine } A^t: A^t = \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 \\ \frac{1}{4} & & & & \vdots \\ 0 & & & & \vdots \\ \vdots & \tilde{B} & & & \vdots \\ 0 & & & 0 & \frac{1}{4} \\ \vdots & & & & \vdots \\ 0 & 0 & \cdots & 0 & 1 \end{pmatrix}^t = \begin{pmatrix} 1 & \frac{1}{4} & \cdots & 0 & 0 \\ 0 & & & & \vdots \\ 0 & & & & \vdots \\ \vdots & \tilde{B}^t & & & \vdots \\ 0 & & & 0 & 0 \\ \vdots & & & & 0 \\ 0 & 0 & \cdots & \frac{1}{4} & 1 \end{pmatrix}$$

$$\text{Since } \tilde{B} \text{ is symmetric, we get that } A^t = \begin{pmatrix} 1 & \frac{1}{4} & \cdots & 0 & 0 \\ 0 & & & & \vdots \\ 0 & & & & \vdots \\ \vdots & \tilde{B} & & & \vdots \\ 0 & & & 0 & 0 \\ \vdots & & & & 0 \\ 0 & 0 & \cdots & \frac{1}{4} & 1 \end{pmatrix}.$$

We will denote a wanted eigenvector as  $v^t = (v_1 \ v_2 \ \cdots \ v_k \ v_{k+1})^t$ .

$$\rightarrow A^t v^t = v^t \rightarrow \begin{pmatrix} 1 & \frac{1}{4} & \cdots & 0 & 0 \\ 0 & & & & \vdots \\ 0 & & & & \vdots \\ \vdots & \tilde{B}^t & & & \vdots \\ 0 & & & 0 & 0 \\ \vdots & & & & 0 \\ 0 & 0 & \cdots & \frac{1}{4} & 1 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_k \\ v_{k+1} \end{pmatrix} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_k \\ v_{k+1} \end{pmatrix}$$

$$\rightarrow \begin{pmatrix} v_1 + \frac{1}{4}v_2 \\ 0 \\ \vdots \\ 0 \\ \frac{1}{4}v_k + v_{k+1} \end{pmatrix} + \begin{pmatrix} 0 \\ v_2 \\ v_3 \\ \vdots \\ v_k \\ 0 \end{pmatrix} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_k \\ v_{k+1} \end{pmatrix}$$

We can subtract  $\begin{pmatrix} v_1 \\ 0 \\ \vdots \\ 0 \\ v_{k+1} \end{pmatrix}$  from both sides and get:

$$\begin{pmatrix} \frac{1}{4}v_2 \\ 0 \\ \vdots \\ 0 \\ \frac{1}{4}v_k \end{pmatrix} + \begin{pmatrix} 0 \\ \tilde{B} \begin{pmatrix} v_2 \\ v_3 \\ \vdots \\ v_k \end{pmatrix} \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ v_2 \\ \vdots \\ v_k \\ 0 \end{pmatrix}$$

Now we know that  $v_2 = v_k = 0$  (from elementwise equality).

So we can substitute these values and get:

$$\begin{pmatrix} \frac{1}{4} \cdot 0 \\ 0 \\ \vdots \\ 0 \\ \frac{1}{4} \cdot 0 \end{pmatrix} + \begin{pmatrix} 0 \\ \tilde{B} \begin{pmatrix} 0 \\ v_3 \\ \vdots \\ v_{k-1} \\ 0 \end{pmatrix} \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ \tilde{B} \begin{pmatrix} 0 \\ v_3 \\ \vdots \\ v_{k-1} \\ 0 \end{pmatrix} \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ v_3 \\ \vdots \\ v_{k-1} \\ 0 \\ 0 \end{pmatrix}$$

Again, from elementwise equality we can deduce that  $\tilde{B} \begin{pmatrix} 0 \\ v_3 \\ \vdots \\ v_{k-1} \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ v_3 \\ \vdots \\ v_{k-1} \\ 0 \\ 0 \end{pmatrix}$

We can look at it as looking for an eigenvector of  $\tilde{B}$  with eigenvalue of 1.

We already know that 1 isn't an eigenvalue of  $\tilde{B}$ , so the only solution is

$$\begin{pmatrix} 0 \\ v_3 \\ \vdots \\ v_{k-1} \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix}. \text{ So, the wanted eigenvector is of the shape } \begin{pmatrix} v_1 \\ 0 \\ \vdots \\ 0 \\ v_{k+1} \end{pmatrix}, \text{ where}$$

$v_1$  and  $v_{k+1}$  are degrees of freedom. We will pick them such that we get 2

$$\text{linearly independent eigenvectors: } \overline{v_1} = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix}, \overline{v_2} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}.$$

- r. We have the eigendecomposition  $A = U\Sigma U^{-1}$  where the columns of  $U$  are the right eigenvectors of  $A$ , and the rows of  $U^{-1}$  are the left eigenvectors of  $A$ , and  $\Sigma$  is a diagonal matrix with the elements sorted in decreasing order. From the structure of the eigendecomposition, we know that we can assume that the values of  $\Sigma$  are the eigenvalues of  $A$ , sorted in decreasing order.

From our deductions in section (p), we know that:

$$\lim_{n \rightarrow \infty} \Sigma^n = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix}. \text{ Therefore:}$$

$$\begin{aligned}
\lim_{n \rightarrow \infty} A^n &= \lim_{n \rightarrow \infty} U \Sigma^n U^{-1} = U \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix} U^{-1} = \\
&\begin{pmatrix} | & | & & | \\ R_1 & R_2 & \cdots & R_{k+1} \\ | & | & & | \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix} \begin{pmatrix} - & L_1 & - \\ - & L_2 & - \\ & \vdots & \\ - & L_{k+1} & - \end{pmatrix} = \\
&\begin{pmatrix} | & | & & | \\ R_1 & R_2 & \cdots & R_{k+1} \\ | & | & & | \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix} \begin{pmatrix} - & L_1 & - \\ - & L_2 & - \\ & \vdots & \\ - & L_{k+1} & - \end{pmatrix} \\
&= \begin{pmatrix} | & | & | & \cdots & | \\ R_1 & R_2 & 0 & \cdots & 0 \\ | & | & | & & | \end{pmatrix} \begin{pmatrix} - & L_1 & - \\ - & L_2 & - \\ - & 0 & - \\ & \vdots & \\ - & 0 & - \end{pmatrix} = \\
&\begin{pmatrix} | \\ R_1 \\ | \end{pmatrix} \begin{pmatrix} - & L_1 & - \end{pmatrix} + \begin{pmatrix} | \\ R_2 \\ | \end{pmatrix} \begin{pmatrix} - & L_2 & - \end{pmatrix} = R_1 L_1 + R_2 L_2 \\
&\rightarrow \lim_{n \rightarrow \infty} d^n = \lim_{n \rightarrow \infty} A^n d^0 = \lim_{n \rightarrow \infty} A^n d^0 = U \left( \lim_{n \rightarrow \infty} \Sigma^n \right) U^{-1} d^0 \\
&= (R_1 L_1 + R_2 L_2) d^0
\end{aligned}$$

From the structure of  $U$  and  $U^{-1}$  in the eigendecomposition, we know that  $R_1, R_2, L_1, L_2$  are the eigenvectors associated with the 2 largest eigenvalues, which are both 1. We have already found left and right eigenvectors associated with the eigenvalue 1, so we know  $R_1, R_2, L_1, L_2$  (the order between each pair doesn't matter since they are orthogonal), up to some constant (on one the matrices  $U$  or  $U^{-1}$ ):

$$\begin{aligned}
R_1 &= c_1 \begin{pmatrix} 0 \\ 2 \\ \vdots \\ 2(k-1) \\ 2k \end{pmatrix}, R_2 = c_1 \begin{pmatrix} 2k \\ 2(k-1) \\ \vdots \\ 2 \\ 0 \end{pmatrix}, L_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix}^t, L_2 = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}^t \\
\rightarrow \lim_{n \rightarrow \infty} d^n &= (R_1 L_1 + R_2 L_2) d^0 = c_1 \begin{pmatrix} 0 & 0 & \cdots & 0 & 2k \\ 2 & 0 & \cdots & 0 & 2(k-1) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 2(k-1) & 0 & \cdots & 0 & 2 \\ 2k & 0 & \cdots & 0 & 0 \end{pmatrix} d^0
\end{aligned}$$

We can denote  $d^0 = \begin{pmatrix} d_0 \\ d_1 \\ \vdots \\ d_{k-1} \\ d_k \end{pmatrix}$  and get the following expression:

$$\lim_{n \rightarrow \infty} d^n = c_1 \begin{pmatrix} 0 & 0 & \dots & 0 & 2k \\ 2 & 0 & \dots & 0 & 2(k-1) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 2(k-1) & 0 & \dots & 0 & 2 \\ 2k & 0 & \dots & 0 & 0 \end{pmatrix} \begin{pmatrix} d_0 \\ d_1 \\ \vdots \\ d_{k-1} \\ d_k \end{pmatrix} =$$

$$c_1 \begin{pmatrix} 2k \cdot d_k \\ 2 \cdot d_0 + 2(k-1) \cdot d_k \\ \vdots \\ 2(k-1) \cdot d_0 + 2 \cdot d_k \\ 2k \cdot d_0 \end{pmatrix}$$

The  $d^n$  vector is the vector of decision levels for our quantization, so we know that for every  $n$ , the first and last values are the boundaries:

$$d_0 = a, d_k = b.$$

$$\rightarrow d = c_1 \begin{pmatrix} 2k \cdot d_k \\ 2 \cdot d_0 + 2(k-1) \cdot d_k \\ \vdots \\ 2(k-1) \cdot d_0 + 2 \cdot d_k \\ 2k \cdot d_0 \end{pmatrix} = c_1 \begin{pmatrix} 2kb \\ 2a + 2(k-1)b \\ \vdots \\ 2(k-1)a + 2b \\ 2ka \end{pmatrix}$$

We can see that  $d$  is only dependent on  $a, b, k$  ( $c_1$  is a constant that we can find), and in particular  $d$  is independent from the choice of  $d^0$ .

- s. What we are looking for is the value of  $c_1$ . We already said that for every  $d^n$ , the first and last values are the bounds, and that also include  $d$ . Therefore we can find  $c_1$  by requiring the first and last values to be  $a$  and  $b$ :

$$c_1 \cdot 2kb = b \rightarrow c_1 = \frac{1}{2k}$$

$$c_1 \cdot 2ka = a \rightarrow c_1 = \frac{1}{2k}$$

$$\rightarrow d = \frac{1}{2k} \begin{pmatrix} 2kb \\ 2a + 2(k-1)b \\ \vdots \\ 2(k-1)a + 2b \\ 2ka \end{pmatrix} = \begin{pmatrix} \frac{b}{k} \\ \frac{a + (k-1)b}{k} \\ \vdots \\ \frac{(k-1)a + b}{k} \\ a \end{pmatrix}$$

$$\rightarrow \forall i \in \{0, \dots, k\}: d_i = \frac{(k-i) \cdot a + i \cdot b}{k} = \frac{ka - ia + ib}{k} = a + i \frac{b-a}{k}$$

Also, we know from section (p) that the diagonal values of  $\Sigma$  (other than the first 2) are less than 1 in their absolute value, and therefore they converge to 0 when raised to a power, and that is why the algorithm will converge in exponential time.

- t. What we have shown so far is that the Max-Lloyd converges to the uniform decision levels. We learned in class that this algorithm converges to a local minimum, but since we found that for every initial  $d^0$  it converges to the same point, we know that all local minima are the same point, and so it must be a single global minimum of the loss function. After we have the optimal decision levels, the representation levels are chosen optimally as the mid-point in every decision level (we saw in the lectures).