Signal and Image Processing (236327)

Tutorials 2-3

Quantization
Approximate Representation of a Set of Values

Consider a set of $N$ scalar real values: \( \{x_1, x_2, \ldots, x_N \} \)

We would like to represent them using a single real value $\bar{x}$.

Recall that in the lecture we found that the average value of the set minimizes the mean squared error (MSE) of the representation.

Let us consider here the representation that minimizes the mean absolute deviation (MAD):

\[
\minimize_{\bar{x}} \frac{1}{N} \sum_{i=1}^{N} |x_i - \bar{x}|
\]

We optimize by demanding:

\[
\frac{\partial}{\partial \bar{x}} \left\{ \frac{1}{N} \sum_{i=1}^{N} |x_i - \bar{x}| \right\} = 0
\]
Approximate Representation of a Set of Values

\[
\frac{\partial}{\partial \bar{x}} \left\{ \frac{1}{N} \sum_{i=1}^{N} |x_i - \bar{x}| \right\} = \frac{1}{N} \sum_{i=1}^{N} \frac{\partial}{\partial \bar{x}} |x_i - \bar{x}| = -\frac{1}{N} \sum_{i=1}^{N} \text{sign}(x_i - \bar{x})
\]

where, we used the definitions:

\[
\frac{d}{dx} |x| = \text{sign}(x) \quad , \quad \text{sign}(x) \triangleq \begin{cases} 
1 & , x > 0 \\
0 & , x = 0 \\
-1 & , x < 0 
\end{cases}
\]

then, the optimality condition becomes

\[
\sum_{i=1}^{N} \text{sign}(x_i - \bar{x}) = 0 \quad \rightarrow \quad \sum_{i: x_i > \bar{x}} 1 = \sum_{i: x_i < \bar{x}} 1
\]

Optimality condition: The set contains equal amounts of values lower than \( \bar{x} \) and values higher than \( \bar{x} \).

meaning that the optimal \( \bar{x} \) is the median value of the set \( \{x_1, x_2, \ldots, x_N\} \).

For odd \( N \) and distinct values, the regular median is the solution.
In other cases, minor extensions of the median may be required.
Signal Digitization: The Quantization Role

• The signal **sampling** gives
  – A discrete signal of finite number of elements.
  – These elements are **real numbers**
    • hence, their **exact representation** would require an infinite amount of bits.

• **Quantization** is used for
  – **representing approximated-values** of the samples using a **finite number of bits**.
The Quantization Procedure

• The input $x$ is a value in the continuous range $[\varphi_L, \varphi_H]$.

• Representing values in $[\varphi_L, \varphi_H]$ using $b$ bits is formed of:
  - $J = 2^b$ discrete representation levels $\{r_i\}_1^J$.
  - Partitioning of $[\varphi_L, \varphi_H]$ using $J + 1$ decision levels $\{d_i\}_0^J$, creating $J$ decision regions (here, intervals).
  - The mapping function: $Q(x) = r_i$ for $x \in [d_{i-1}, d_i)$

Demonstration ($b = 3$):

\[
\begin{align*}
  &d_0 = \varphi_L & &d_1 & &d_2 & &d_3 & &d_4 & &d_5 & &d_6 & &d_7 & &d_8 = \varphi_H \\
  &r_1 & &r_2 & &r_3 & &r_4 & &r_5 & &r_6 & &r_7 & &r_8
\end{align*}
\]
The Quantization Error
for a Uniform Quantizer and a Uniform-Distributed Input

- Our input values are drawn from a uniform distribution over $[\phi_L, \phi_H]$: The probability density function is

$$p(x) = \begin{cases} \frac{1}{\phi_H - \phi_L} & \text{for } x \in [\phi_L, \phi_H) \\ 0 & \text{otherwise} \end{cases}$$
The Quantization Error for a Uniform Quantizer and a Uniform-Distributed Input

• Let us consider a uniform quantizer of \( b \) bits:
  
  ▪ There are \( J = 2^b \) quantization-intervals of equal size 
    \[ \Delta_q = \frac{\varphi_H - \varphi_L}{2^b}. \]
  
    • one of \( J, b \) or \( \Delta_q \) should be set, the other two are adjusted correspondingly.

  ▪ Decision levels: 
    \[ d_i = \varphi_L + i\Delta_q , \quad i = 0, 1, ..., J \]

  ▪ Representation levels: 
    \[ r_i = \varphi_L + (i - \frac{1}{2})\Delta_q , \quad i = 1, ..., J \]

  ▪ Mapping formula: 
    for \( x \in [\varphi_L, \varphi_H) \):
    \[ Q(x) = \varphi_L + \Delta_q \cdot \left[ \left\lfloor \frac{x - \varphi_L}{\Delta_q} \right\rfloor + \frac{1}{2} \right] \]

**Demonstration (\( b = 3 \)):**

\[ d_0 = \varphi_L \quad d_1 \quad d_2 \quad d_3 \quad d_4 \quad d_5 \quad d_6 \quad d_7 \quad d_8 = \varphi_H \]
The Quantization Error for a Uniform Quantizer and a Uniform-Distributed Input

The expected squared-error:

$$E\{\varepsilon_Q^2\} = E\{(x - Q(x))^2\} = \int_{\phi_L}^{\phi_H} (x - Q(x))^2 \, p(x) \, dx = \sum_{i=1}^{J} \int_{d_{i-1}}^{d_i} (x - r_i)^2 \, p(x) \, dx = (*)$$

setting

$$d_i = \phi_L + i\Delta \quad ; \quad d_{i-1} = \phi_L + (i - 1)\Delta \quad ; \quad r_i = d_{i-1} + 0.5\Delta$$

and \( p(x) = \frac{1}{\phi_H - \phi_L} \) for \( x \in [\phi_L, \phi_H] \) gives

$$(* \, ) = \frac{1}{\phi_H - \phi_L} \sum_{i=1}^{J} \int_{\phi_L + (i-1)\Delta}^{\phi_L + i\Delta} \left( x - \left( \phi_L + (i - 1)\Delta + 0.5\Delta \right) \right)^2 \, dx = (** \, )$$

integration-variable substitution: \( e_i \equiv x - \left( \phi_L + (i - 1)\Delta + 0.5\Delta \right) \) for each integral \( i = 1, \ldots, J \)

$$(** \, ) = \frac{1}{\phi_H - \phi_L} \sum_{i=1}^{J} \int_{-\Delta/2}^{\Delta/2} e_i^2 \, de_i = \frac{1}{\phi_H - \phi_L} J \int_{-\Delta/2}^{\Delta/2} z^2 \, dz = \frac{1}{\phi_H - \phi_L} J \frac{z^3}{3} \bigg|_{-\Delta/2}^{\Delta/2} = \frac{\phi_H - \phi_L}{\Delta}$$

$$= \frac{1}{3\Delta} \left( \frac{\Delta^3}{8} - \frac{\Delta^3}{8} \right) = \frac{\Delta^3}{\Delta \cdot 12} = \frac{\Delta^2}{12}$$

An important result that will return in various forms.
Optimal Scalar Quantization

Consider a scalar input value $x$, drawn from the distribution $p(x)$ over the continuous range $[\varphi_L, \varphi_H]$. The distribution is not necessarily uniform.

The quantizer is formed from:

- $J$ representation levels: $r_1, r_2, ..., r_J$
- $J + 1$ decision levels: $d_0, d_1, ..., d_J$
  creating $J$ decision regions $D_i = [d_{i-1}, d_i)$, $i = 1, ..., J$
- The mapping: $Q(x) = r_i$ for $x \in D_i$.

The expected squared error is:

$$E\{\varepsilon^2_Q\} = \sum_{i=1}^{J} \int_{d_{i-1}}^{d_i} (x - r_i)^2 p(x)dx$$

The quantizer optimization problem minimizing the expected squared error as a function of the representation and decision levels:

$$\text{minimize} \sum_{i=1}^{J} \int_{d_{i-1}}^{d_i} (x - r_i)^2 p(x)dx$$
Optimal Scalar Quantization

minimize $E\{\varepsilon_Q^2\}$
\[ \{d_j\}_{j=0}^J, \{r_j\}_{j=1}^J \]

Necessary conditions for optimality are given by:

\[ \frac{\partial}{\partial r_j} E\{\varepsilon_Q^2\} = 0 \quad \text{for} \quad j = 1, \ldots, J \]

\[ \frac{\partial}{\partial d_j} E\{\varepsilon_Q^2\} = 0 \quad \text{for} \quad j = 1, \ldots, J - 1 \]

Note that we set $d_0 = \varphi_L$ and $d_J = \varphi_H$. 
Optimal Scalar Quantization

Given the decision levels \( \{d_i\}_{i=0}^J \), let us develop the representation-level optimality conditions:

\[
\frac{\partial}{\partial r_j} E\{\varepsilon_Q^2\} = 0 \quad \text{for} \quad j = 1, \ldots, J
\]

Equivalently,

for \( j = 1, \ldots, J \):

\[
\frac{\partial}{\partial r_j} \sum_{i=1}^{d_i} \int_{d_{i-1}}^{d_i} (x - r_i)^2 p(x) \, dx = 0
\]

Developing the derivative expression:

\[
\frac{\partial}{\partial r_j} \sum_{i=1}^{d_i} \int_{d_{i-1}}^{d_i} (x - r_i)^2 p(x) \, dx = \frac{\partial}{\partial r_j} \int_{d_{j-1}}^{d_j} (x - r_j)^2 p(x) \, dx = \int_{d_{j-1}}^{d_j} \frac{\partial}{\partial r_j} (x - r_j)^2 p(x) \, dx = -2 \int_{d_{j-1}}^{d_j} (x - r_j) p(x) \, dx
\]
Optimal Scalar Quantization

We got that

for \( j = 1, ..., J \):

\[
\frac{\partial}{\partial r_j} E\{\varepsilon^2_0\} = 0
\]

implies

\[
-2 \int_{d_{j-1}}^{d_j} (x - r_j)p(x)dx = 0
\]

that is

\[
r_j = \frac{\int_{d_{j-1}}^{d_j} x p(x)dx}{\int_{d_{j-1}}^{d_j} p(x)dx}
\]

Given the decision regions,

the best representation levels are the "centers of mass" of the regions.
Optimal Scalar Quantization

Given the representation levels \( r_1 < r_2 < \cdots < r_J \), let us express the decision-level optimality conditions:

Recall that we fix \( d_0 = \varphi_L \) and \( d_J = \varphi_H \).

for \( j = 1, \ldots, J - 1 \):

\[
\frac{\partial}{\partial d_j} \sum_{i=1}^{J} \int_{d_{i-1}}^{d_i} (x - r_i)^2 p(x)dx = 0
\]

Demanding the derivatives equal to zero.

Each \( d_j \) participates in two intervals: \([d_{j-1}, d_j)\) and \([d_j, d_{j+1})\), hence

\[
\frac{\partial}{\partial d_j} \sum_{i=1}^{J} \int_{d_{i-1}}^{d_i} (x - r_i)^2 p(x)dx = \frac{\partial}{\partial d_j} \int_{d_{j-1}}^{d_j} (x - r_j)^2 p(x)dx + \frac{\partial}{\partial d_j} \int_{d_j}^{d_{j+1}} (x - r_{j+1})^2 p(x)dx
\]

\[
= (d_j - r_j)^2 p(d_j) - (d_j - r_{j+1})^2 p(d_j)
\]

Recall that \( \frac{\partial}{\partial x} \int_{y}^{x} f(t)dt = f(x) \) and \( \frac{\partial}{\partial x} \int_{x}^{y} f(t)dt = -f(x) \)
Optimal Scalar Quantization

for $j = 1, ..., J - 1$:

$$(d_j - r_j)^2 p(d_j) - (d_j - r_{j+1})^2 p(d_j) = 0$$

$$d_j - r_j = -(d_j - r_{j+1})$$

$$d_j - r_j = d_j - r_{j+1}$$

$$d_j = \frac{r_j + r_{j+1}}{2}$$

Irrelevant solution as we assume that $r_j < r_{j+1}$

Given the representation levels, the best decision levels are in the middle between the representation levels.

Given the representation levels we can find the decision regions, and given the decision regions we have a formula for the best representation levels.
Optimal Scalar Quantization

In the general case, it is difficult to directly optimize the decision and representation levels together. Hence, the quantizer-design optimization is addressed via alternating minimization (iterative optimization procedure):

\[
\text{minimize} \sum_{i=1}^{J} \int_{d_{i-1}}^{d_i} (x - r_i)^2 p(x) dx
\]

where \( k \) is the iteration number.

* The procedure terminates when a stopping criterion is satisfied.
* Initialization of the decision levels is required.
Optimal Scalar Quantization
Max-Lloyd Algorithm

The quantizer optimization procedure that we got is known as Max-Lloyd algorithm for quantizer design. It can be summarized as follows:

1. Initialization: set\guess decision levels \( \{d_i\} \)
2. Compute the optimal representation levels for \( \{d_i\} \) and set to \( \{r_i\} \)
3. Compute the optimal decision levels for \( \{r_i\} \) and set to \( \{d_i\} \)
4. If stopping criteria has not met, return to (2).

Remark: the roles of the representation and decision levels can be substituted.

Each step applies an optimization, hence the MSE will decrease or not changed. Convergence to a minima (however, not necessarily the global one) is guaranteed.
Max-Lloyd Quantizer

Exercise #1

The samples of a given signal have the following probability-density-function:

\[ p(x) = \begin{cases} 
2x, & 0 \leq x \leq 1 \\
0, & \text{else} 
\end{cases} \]

Find the Max-Lloyd quantizer for two representation levels (J=2).
Max-Lloyd Quantizer

Analytic Solution:

input range borders: $d_0 = 0$, $d_2 = 1$

optimal decision level:

(1) $d_1 = \frac{r_1 + r_2}{2}$

optimal representation levels:

(2) $r_1 = \frac{\int_{d_1}^{d_2} x \cdot p(x) \, dx}{\int_{d_1}^{d_2} p(x) \, dx} = \frac{\int_{d_1}^{d_2} x \cdot 2x \, dx}{\int_{0}^{d_1} 2x \, dx} = \frac{2}{3} d_1$

(3) $r_2 = \frac{\int_{d_1}^{d_2} x \cdot p(x) \, dx}{\int_{d_1}^{d_2} p(x) \, dx} = \frac{\int_{d_1}^{1} x \cdot 2x \, dx}{\int_{d_1}^{1} 2x \, dx} = \frac{2}{3} \frac{1 - d_1^3}{1 - d_1^2}$

substituting (2) and (3) into (1) yields: $d_1 = \frac{\frac{2}{3} d_1 + \frac{2}{3} \frac{1 - d_1^3}{1 - d_1^2}}{2} \Rightarrow d_1^3 - 2d_1 + 1 = 0 \Rightarrow d_1 = \sqrt[3]{0.618}, -1.618$

setting the optimal $d_1$ into (2) and (3) gives the optimal representation levels: $r_1 = 0.412$ and $r_2 = 0.824$
Max-Lloyd Quantizer

Iterative Solution:

(1) Initial guess: representation levels of a uniform quantizer in the range [0, 1]: \( r_1^{(0)} = 0.25 \), \( r_2^{(0)} = 0.75 \)

Iteration 1: \( d_1^{(1)} = \frac{r_1^{(0)} + r_2^{(0)}}{2} = \frac{1}{2} \)

\( r_1^{(1)} = \frac{2}{3} d_1^{(1)} = \frac{1}{3} \) \( r_2^{(1)} = \frac{2}{3} \cdot \frac{1 - (d_1^{(1)})^3}{1 - (d_1^{(1)})^2} = 0.778 \)

Iteration 2: \( d_1^{(2)} = \frac{r_1^{(1)} + r_2^{(1)}}{2} = 0.556 \)

\( r_1^{(2)} = \frac{2}{3} d_1^{(2)} = 0.37 \) \( r_2^{(2)} = \frac{2}{3} \cdot \frac{1 - (d_1^{(2)})^3}{1 - (d_1^{(2)})^2} = 0.799 \)

\( \cdots \)

Iteration 15: \( d_1^{(15)} = 0.618 \) \( r_1^{(15)} = 0.412 \) \( r_2^{(15)} = 0.824 \)
Vector Quantization

The extension of scalar quantization problem to the multidimensional case:

- The input is a real-valued $N$-dimensional vector: $\mathbf{x} \in \mathbb{R}^N$
- The $b$-bits quantizer mapping:
  \[ Q_v(\mathbf{x}) = \mathbf{r}_i \quad \text{for} \quad \mathbf{x} \in D_i \quad i = 1, \ldots, 2^b \]

where
\[ \{\mathbf{r}_i\}_{i=1}^J \] are the $N$-dimensional representation vectors.
\[ \{D_i\}_{i=1}^J \] are non-intersecting decision regions that segment $\mathbb{R}^N$.

The expected squared-error of the quantizer is:
\[ E \epsilon_Q^2 = \int_{\mathbb{R}^N} \|\mathbf{x} - Q_v(\mathbf{x})\|_2^2 \cdot p(\mathbf{x}) \, d\mathbf{x} \]

where $p(\mathbf{x})$ is the probability density function (recall that $\int_{\mathbb{R}^N} p(\mathbf{x}) \, d\mathbf{x} = 1$).
Vector Quantization

The expected squared error is

$$E \varepsilon_Q^2 = \sum_{i=1}^{J} \int_{x \in D_i} \|x - r_i \|^2 \cdot p(x) \, dx$$

- Optimality of representation vectors given the decision regions:

$$\frac{\partial}{\partial r_j} E \varepsilon_Q^2 = 0$$

Note that this is a derivative by a vector.
Vector Quantization

- The vector derivative of the $\ell_2$ norm:
  \[
  \frac{\partial}{\partial z} \|z\|_2^2 = \frac{\partial}{\partial z} z^T z = 2z
  \]

- **Optimality of representation vectors** given the decision regions:
  \[
  \frac{\partial}{\partial r_j} \sum_{i=1}^{J} \int_{x \in D_i} \|x - r_i\|_2^2 \cdot p(x) \, dx = \frac{\partial}{\partial r_j} \int_{x \in D_j} \|x - r_j\|_2^2 \cdot p(x) \, dx \\
  = \int_{x \in D_j} \frac{\partial}{\partial r_j} \|x - r_j\|_2^2 \cdot p(x) \, dx = -2 \int_{x \in D_j} (x - r_j) \cdot p(x) \, dx
  \]

Hence, the optimality condition is
\[
  r_j^{opt} = \frac{\int_{x \in D_j} x \cdot p(x) \, dx}{\int_{x \in D_j} p(x) \, dx}
\]
Vector Quantization

The optimal decision regions given the representation vectors:

- In contrast to the scalar case, we cannot use derivatives to optimize the multidimensional decision regions.

- The multidimensional decision region optimization:
  For $\mathbf{x} \in \mathbb{R}^N$, the optimal mapping to the a representation vector is the one minimizing the squared-error of the representation of $\mathbf{x}$. Accordingly, the optimal regions are defined as

$$D_j^{opt} = \left\{ \mathbf{x} \mid \|\mathbf{x} - \mathbf{r}_j\|_2^2 < \|\mathbf{x} - \mathbf{r}_i\|_2^2 \text{ for } i = 1, \ldots, J \right\}$$

(the regions boundaries, where equality holds, can be assigned to the region with the lower index).

The optimality conditions for representation vectors and decision regions can be applied in a Max-Lloyd procedure for Vector Quantization.