Intro to Signal Processing – HW1

Yinon Goldshtein , Tomer Kalmanzon
[redacted]

Submitted 4 Dec 2022

1 Question 1

1.1 a

We can write the expected error in a different form:

$$\min_{\hat{f}} \varepsilon^p(f,\hat{f}) = \min_{\hat{f}} \int_0^1 |f(x)-\hat{f}(x)|^p \cdot w(x)dx = \min_{\hat{f}} \sum_{i=1}^{N} \int_{I_i} |f(x)-\hat{f}(x)|^p \cdot w(x)dx \underbrace{=}_{w \text{ constant}} \min_{\hat{f}} w \cdot \sum_{i=1}^{N} \int_{I_i} |f(x)-\hat{f}(x)|^p dx$$

For $p = 1$ For $p = 1$, this is the MAD error, and we'll need to find the "median" value of each interval, as this is the optimal $\hat{f}$ value in each interval.

Denoting f^i_{min} as the minimal value of $f(x)$ in interval I_i, and f^i_{max} as the max value of $f(x)$ in I_i, and assuming we have a method to compute integrals, the median value θ_i for value of $I_i = [d_i, d_{i+1}]$ is:

$$\mathbb{P}(f^i_{min} \leq x \leq \theta_i) = \mathbb{P}(\theta_i \leq x \leq f^i_{max})$$

Given that we found such θ_i for each interval I_i, the optimal value of $\hat{f}_p(x)$ is:

$$\hat{f}_p(x) = \sum_{i=1}^{N} \theta_i \cdot 1^{I_i}(x)$$

Where $1^{I_i}(x)$ is a function returning 1 if $x \in I_i$ and 0 otherwise.

For $p = 2$ For $p = 2$, this is the MSE error, and the optimal value of $\hat{f}_p$ for each interval would be the "average" value of $f(x)$ in it.

Therefore:

$$\hat{f}_p(x) = \sum_{i}^{N} \frac{1}{I_i} \int_{I_i} f(x) \cdot 1^{I_i}(x)dx$$

1.2 b

As shown in the first lecture, with the example of MSE on a discrete set and with a constant $\hat{f}_p$, the value of $\hat{f}_p(x)$ for each interval should be $\frac{\sum_{i=1}^{N} w_i x_i}{\sum_{i=1}^{N} w_i}$.

Therefore, the analogy for a continous function with a continous weight function on an interval I_i, would be: $\frac{1}{\int_{I_i} w(x)} \int_{I_i} f(x) \cdot w(x)dx$

Hence:

$$\hat{f}_p(x) = \sum_{i}^{N} \frac{1}{\int_{I_i} w(x)} \int_{I_i} f(x) \cdot w(x) 1^{I_i}(x)dx$$

1.3 c

With $p = 1$, the error will be:

$$\varepsilon(f, \hat{f}) = \sum_{i=1}^{N} \int_{I_i} |f(x) - \hat{f}(x)| \cdot w(x)dx = \sum_{i=1}^{N} \int_{I_i} w(x) \cdot (f(x) - \hat{f}(x)) \cdot sign(f(x) - \hat{f}(x))$$

We'll differentiate to get the min point in each interval I_i:

$$\frac{\partial \varepsilon(f, \hat{f})}{\partial \hat{f}}_{I_i} = \int_{I_i} -w(x) \cdot sign(f(x) - \hat{f})dx$$

The $\hat{f}$ value that minimizes this expression will be the value of $\hat{f}(x)$ in interval I_i.
Hence:

$$\hat{f}(x) = \sum_{i=1}^{N} \int_{I_i} -w(x) \cdot sign(f(x) - \hat{f})dx \cdot 1^{I_i}(x)$$

Where $1^{I_i}(x)$ is a function returning 1 if $x \in I_i$ and 0 otherwise.

1.4 d

Since we use the function $1^{I_i}(x)$, which return 1 if $x \in I_i$ and 0 otherwise, for every interval I_i the optimal solution is given by $\varepsilon_i^p(f_i, \hat{f}_i)$.

From the additive property of this optimization problem, we get the optimal solution $\varepsilon^p(f, \hat{f})$ by solving the same problem for each interval I_i.

$$\varepsilon^p(f, \hat{f}) = \sum_{i=1}^{N} \varepsilon_i^p(f_i, \hat{f}_i)$$

1.5 e

i We can divide the equation by $(f_i(x) - \hat{f}_i(x))^2$, as $f_i(x) \neq \hat{f}_i(x)$, to get:

$$w_{f_i, \hat{f}_i} = |f_i(x) - \hat{f}_i(x)|^{p-2}$$

ii Using the expression we found in section i, $w'(f, \hat{f})$ we can plug it in the main problem:

$$\min_{\hat{f}} \varepsilon^p(f, \hat{f}) = \min_{\hat{f}} \int_0^1 |f(x) - \hat{f}(x)|^p \cdot w(x)dx = \min_{\hat{f}} \int_0^1 |f(x) - \hat{f}(x)|^2 \cdot w'_{f, \hat{f}}(x) \cdot w(x)dx$$

iii It would be much simpler if the weight function was independent of $\hat{f}_i$, as the optimization problem would become a simple L^2 optimization problem.

If the weight function is dependent on $\hat{f}_i$, when we would diferentiate the loss by $\hat{f}_i$, we'd get an optimiaztion problem with chain rule derivatives, as $\hat{f}_i$ is a function of x, complicating the optimization process.

iv We prefer to have a maximal value of $\frac{1}{\epsilon}$ for our effective weight function for 2 main reasons:

1. In case of an interval I_i where we have $|f(x) - \hat{f}_i(x)| > 1$ and $p \gg 1$, this expression will "explode" to a large number, effectively dismissing the other interval losses we have.
2. In case of high variance in the loss in different intervals (for example - f is constant in one interval and in another interval $|f(x) - \hat{f}(x)| \gg 1$), we would like to have a ceiling for possible losses, so one major loss from one interval would not dismiss the effect of other intervals on the optimization scheme.

v We assume we have a function $w(f_i(x), \hat{f}_i(x))$, that its output is the function $w'_i = w_{f_i, \hat{f}_i}(x)$.

Algorithm 1 Pseudo-code for $\hat{f}_i^{next}$ approximation

Input: f_i, initial approximation $\hat{f}_i$
Output: $\hat{f}_i$ approximation for f_i
Initialization: $w'_i \leftarrow w(f_i(x), \hat{f}_i(x))$
Iterate as much times as needed:

1. $\hat{f}_i^{next} \leftarrow \arg\min_{\hat{f}_i} \int_{I_i} \left(f_i(x) - \hat{f}_i(x)\right)^2 w'_i(x)dx$, notice that w'_i is independent of $\hat{f}_i$
2. $\hat{f}_i \leftarrow \hat{f}_i^{next}$
3. $w'_i \leftarrow w'_i(f_i, \hat{f}_i)$

return $\hat{f}_i$

Algorithm 2 Solving weighted L^p using only L^2

Inputs:
$f(x)$: original function
$w(x)$: weight function
$\{\Delta_i\}_{i=1}^N$: convergence factor (Δ_i is convergence factor for interval I_i)
Output:
$\hat{f}$: $f(x)$ approximation
Iterate: For each interval I_i:

1. $\hat{f}_i(x) \leftarrow 0$
2. $w'_i \leftarrow \min\left\{\frac{1}{\epsilon}, |f_i(x) - \hat{f}_i(x)|^{p-2} \cdot w(x)\right\}$
3. While $\min_{\hat{f}} \int_0^1 |f(x) - \hat{f}(x)|^2 \cdot w'_i(x) \cdot w(x)dx \geq \Delta_i$:
   (a) $\delta_i \leftarrow \frac{\int_{I_I} f(x)w'_i(x)dx}{\int_{I_i} w'_i(x)dx}$
   (b) $\hat{f}_i \leftarrow \delta_i \cdot 1^{I_i}(x)$
   (c) $w'_i \leftarrow \min\left\{\frac{1}{\epsilon}, |f_i(x) - \hat{f}_i(x)|^{p-2} \cdot w(x)\right\}$

return $\hat{f}$

1.6 f

We assume we have a partition $\{I_i\}_{i=1}^N$ of $[0, 1]$ we want to solve the L^p problem for, and where $\epsilon > 0$ is a small fixed number

1.7 g

The name of the algorithm is "Iteratively reweighted least squares"

1 Question 2

1.1 a

We can write the expected error in a different form:

$$\min_{\hat{f}} \varepsilon^p(f,\hat{f}) = \min_{\hat{f}} \int_0^1 |f(x)-\hat{f}(x)|^p \cdot w(x)dx = \min_{\hat{f}} \sum_{i=1}^{N} \int_{I_i} |f(x)-\hat{f}(x)|^p \cdot w(x)dx \underbrace{=}_{w \text{ constant}} \min_{\hat{f}} w \cdot \sum_{i=1}^{N} \int_{I_i} |f(x)-\hat{f}(x)|^p dx$$

For $p = 1$ For $p = 1$, this is the MAD error, and we'll need to find the "median" value of each interval, as this is the optimal $\hat{f}$ value in each interval.

Denoting f^i_{min} as the minimal value of $f(x)$ in interval I_i, and f^i_{max} as the max value of $f(x)$ in I_i, and assuming we have a method to compute integrals, the median value θ_i for value of $I_i = [d_i, d_{i+1}]$ is:

$$\mathbb{P}(f^i_{min} \leq x \leq \theta_i) = \mathbb{P}(\theta_i \leq x \leq f^i_{max})$$

Given that we found such θ_i for each interval I_i, the optimal value of $\hat{f}_p(x)$ is:

$$\hat{f}_p(x) = \sum_{i=1}^{N} \theta_i \cdot 1^{I_i}(x)$$

Where $1^{I_i}(x)$ is a function returning 1 if $x \in I_i$ and 0 otherwise.

For $p = 2$ For $p = 2$, this is the MSE error, and the optimal value of $\hat{f}_p$ for each interval would be the "average" value of $f(x)$ in it.

Therefore:

$$\hat{f}_p(x) = \sum_{i}^{N} \frac{1}{I_i} \int_{I_i} f(x) \cdot 1^{I_i}(x)dx$$

1.2 b

As shown in the first lecture, with the example of MSE on a discrete set and with a constant $\hat{f}_p$, the value of $\hat{f}_p(x)$ for each interval should be $\frac{\sum_{i=1}^{N} w_i x_i}{\sum_{i=1}^{N} w_i}$.

Therefore, the analogy for a continous function with a continous weight function on an interval I_i, would be: $\frac{1}{\int_{I_i} w(x)} \int_{I_i} f(x) \cdot w(x)dx$

Hence:

$$\hat{f}_p(x) = \sum_{i}^{N} \frac{1}{\int_{I_i} w(x)} \int_{I_i} f(x) \cdot w(x) 1^{I_i}(x)dx$$

1.3 c

With $p = 1$, the error will be:

$$\varepsilon(f,\hat{f}) = \sum_{i=1}^{N} \int_{I_i} |f(x) - \hat{f}(x)| \cdot w(x)dx = \sum_{i=1}^{N} \int_{I_i} w(x) \cdot (f(x) - \hat{f}(x)) \cdot sign(f(x) - \hat{f}(x))$$

We'll differentiate to get the min point in each interval I_i:

$$\frac{\partial \varepsilon(f,\hat{f})}{\partial \hat{f}}_{I_i} = \int_{I_i} -w(x) \cdot sign(f(x) - \hat{f})dx$$

The $\hat{f}$ value that minimizes this expression will be the value of $\hat{f}(x)$ in interval I_i.

Hence:

$$\hat{f}(x)=\sum_{i=1}^{N}\int_{I_i}-w(x)\cdot sign(f(x)-\hat{f})dx\cdot 1^{I_i}(x)$$

Where $1^{I_i}(x)$ is a function returning 1 if $x\in I_i$ and 0 otherwise.

d

Since we use the function $1^{I_i}(x)$, which return 1 if $x\in I_i$ and 0 otherwise, for every interval I_i the optimal solution is given by $\varepsilon_i^p(f_i,\hat{f}_i)$.

From the additive property of this optimization problem, we get the optimal solution $\varepsilon^p(f,\hat{f})$ by solving the same problem for each interval I_i.

$$\varepsilon^p(f,\hat{f})=\sum_{i=1}^{N}\varepsilon_i^p(f_i,\hat{f}_i)$$

1.4 e

i We can divide the equation by $(f_i(x)-\hat{f}_i(x))^2$, as $f_i(x)\neq\hat{f}_i(x)$, to get:

$$w_{f_i,\hat{f}_i}=|f_i(x)-\hat{f}_i(x)|^{p-2}$$

ii Using the expression we found in section i, $w'(f,\hat{f})$ we can plug it in the main problem:

$$\min_{\hat{f}}\varepsilon^p(f,\hat{f})=\min_{\hat{f}}\int_0^1|f(x)-\hat{f}(x)|^p\cdot w(x)dx=\min_{\hat{f}}\int_0^1|f(x)-\hat{f}(x)|^2\cdot w'_{f,\hat{f}}(x)\cdot w(x)dx$$

iii It would be much simpler if the weight function was independent of $\hat{f}_i$, as the optimization problem would become a simple L^2 optimization problem.

If the weight function is dependent on $\hat{f}_i$, when we would diferentiate the loss by $\hat{f}_i$, we'd get an optimiaztion problem with chain rule derivatives, as $\hat{f}_i$ is a function of x, complicating the optimization process.

iv We prefer to have a maximal value of $\frac{1}{\epsilon}$ for our effective weight function for 2 main reasons:

1. In case of an interval I_i where we have $|f(x)-\hat{f}_i(x)|>1$ and $p\gg 1$, this expression will "explode" to a large number, effectively dismissing the other interval losses we have.
2. In case of high variance in the loss in different intervals (for example - f is constant in one interval and in another interval $|f(x)-\hat{f}(x)|\gg 1$), we would like to have a ceiling for possible losses, so one major loss from one interval would not dismiss the effect of other intervals on the optimization scheme.

v

2 Question 2

2.1 let us define $\Delta_i=[\frac{i-1}{N},\frac{i}{N}), i\in[N]$ such that $\phi(t)_i=a_i\hat{(t-t_i)}+c_i$as $t\in\Delta_i$and that $a_i,c_i\in\mathbb{R}$. moreover it is defined that $t_i$is the mid point of $\Delta_i$hence $t_i=\frac{2i-1}{2N}$:

$$\int_{t\in\Delta_i}(t-t_i)^k dt=\frac{(t-t_i)^{k+1}}{k+1}|_{\frac{i-1}{N}}^{\frac{i}{N}}=\frac{(\frac{i}{N}-\frac{2i-1}{2N})^{k+1}-(\frac{i-1}{N}-\frac{2i-1}{2N})^{k+1}}{k+1}=\frac{(\frac{2i-2i+1}{2N})^{k+1}-(\frac{2i-2-2i+1}{2N})^{k+1}}{k+1}=$$

$$\frac{|\Delta_i|^{k+1}[(1)^{k+1}-(-1)^{k+1}]}{2^{k+1}(k+1)}=$$

$$=\begin{cases}0 & k(mod2)=1\\ \frac{|\Delta_i|^{k+1}}{2^k(k+1)} & k(mod2)=0\end{cases}$$

2.2 b

since $\hat{\phi}_i$is defined upon each interval $\Delta_i$then we can define that $\hat{}(t) = \sum_{i=1}^{n} \hat{\phi}(t)_i$, hence the MSE of $\hat{\phi}(t)$ is given by:

$$\Psi(\phi) = \int_0^1 (\phi(t) - \hat{}(t))^2 dt = \int_0^1 (\phi(t) - \sum_{i=1}^{n} \hat{\phi}(t)_i 1_{\Delta_i}(t))dt = \int_0^1 (\phi(t) - \Sigma_{i=1}^{n}(a_i(t-t_i)+c_i)1_{\Delta_i}(t))^2 dt$$

$$= \Sigma_{i=1}^{n} \int_0^1 (\phi(t) - (a_i(t-t_i)+c_i)1_{\Delta_i}(t))^2 dt = \Sigma_{i=1}^{n} \int_{\Delta_i} (\phi(t) - \hat{\phi}_i(t))^2 dt$$

as a cosequence of $\Psi(\phi(t)) = \Sigma_{i=1}^{n} \int_{\Delta_i} (\phi(t) - \hat{\phi}_i(t))^2 dt$ in order to calculate the minimum value of the MSE we must clculate the minimum value of $\Psi(\phi_i(t))$ for every interval $i \in [N]$. so in order to defrentiate $\Psi(\phi(t))$ and equate to zero we must show that $\Psi(\phi(t))$ is convex.

we know that a function is convex iff $\frac{\partial^2 \Psi}{\partial a_i} > 0$:

$$\frac{\partial^2 (\int_{\Delta_i} (\phi(t) - \hat{\phi}_i(t))^2 dt)}{\partial a_i} = \frac{\partial^2 (\int_{\Delta_i} \phi^2(t) + a_i^2(t-t_i)^2 + c_i^2 - 2\phi(t)a_i(t-t_i) - 2\phi(t)c_i - 2a_i(t-t_i)c_i dt)}{\partial a_i} =$$

$$= \int_{\Delta_i} 2(t-t_i)^2 dt = \frac{|\Delta_i|^3}{2*3} > 0$$

hence:

$$0 = \frac{\partial (\int_{\Delta_i} (\phi(t) - \hat{\phi}_i(t))^2 dt)}{\partial a_i} = \frac{\partial (\int_{\Delta_i} \phi^2(t) + a_i^2(t-t_i)^2 + c_i^2 - 2\phi(t)a_i(t-t_i) - 2\phi(t)c_i - 2a_i(t-t_i)c_i dt)}{\partial a_i}$$

$$0 = \int_{\Delta_i} 2a(t-t_i)^2 - \phi(t)(t-t_i) - (t-t_i)c_i dt \iff a^* = \frac{12 \int \phi(t)(t-t_i)dt}{|\Delta_i|^3}$$

as of c_i:

$$\frac{\partial^2 (\int_{\Delta_i} (\phi(t) - \hat{\phi}_i(t))^2 dt)}{\partial c_i} = \int_{\Delta_i} 2dt = \frac{2}{N} > 0$$

hence:

$$0 = \frac{\partial (\int_{\Delta_i} (\phi(t) - \hat{\phi}_i(t))^2 dt)}{\partial c_i} = \int_{\Delta_i} 2c_i - 2\phi(t) - 2a_i(t-t_i)dt \Rightarrow \int_{i\Delta_i} c_i - \phi(t)dt = a_i \int_{\Delta_i} t - t_i dt$$

$$\iff \int_{i\Delta_i} c_i^* dt = \int_{i\Delta_i} \phi(t)dt \Rightarrow c_i^* = \frac{\int_{i\Delta_i} \phi(t)dt}{|\Delta_i|}$$

2.3 c

As a result of the last section we can represent $\Psi_{MSE}(\hat{\phi}) = \Sigma_{i=1}^{n} \int_{\Delta_i} (\phi(t) - \hat{\phi}_i(t))^2 dt = \Sigma_{i=1}^{n} \int_0^1 (\phi(t) - (a_i(t - t_i) + c_i)1_{\Delta_i}(t))^2 dt$ in the following way:

$$\Psi_{MSE}(\hat{\phi}) = \Sigma_{i=1}^{n} \int_{\Delta_i} (\phi(t) - a_i^*(t-t_i) - c_i^*)^2 dt = \Sigma_{i=1}^{n} \int_{\Delta_i} (\phi(t) - \frac{12 \int \phi(t)(t-t_i)^2}{|\Delta_i^3|} - \frac{\int_{i\Delta_i} \phi(t)}{|\Delta_i|})^2 dt$$

2.4 d

As we have seen in class the MSE for Piecewise-constant approximation is given by:

$$\Psi_{MSE}(\hat{\phi}) = \Sigma_{i=1}^{N} \int_{\Delta_i} (\phi(t) - \gamma_i^* 1_{\Delta_i}(x))^2 dt, s.t : \gamma_i^* = \frac{\int_{\Delta_i} \phi(t)dt}{|\Delta_i|}$$

$$\Psi_{MSE}(\hat{\phi}) = \Sigma_{i=1}^{N} \int_{\Delta_i} (\phi(t) - \frac{\int_{\Delta_i} \phi(t)dt}{|\Delta_i|})^2 dt$$

based on the work we have done in the last section we can conclude that the MSE for Piecewise-constant approximation is a private case of a linear approximation where $\forall i \in [N] : a_i = 0$, and since $\Psi_{MSE}(\hat{\phi})$ was shown to be convex w.r.t $\hat{\phi}$we can conclude that the coeficiants a_i^*, c_i^* that were calculated earlier are indeed the optimal coeficiants of te solution of the linear approximation we were given. and as a consequence we can conclude that $\forall a_i, c_i \in \mathbb{R} : \Psi_{MSE}(\phi^*) \leq \Psi_{MSE}(\hat{\phi}^{PC}$

Part II

Implementation

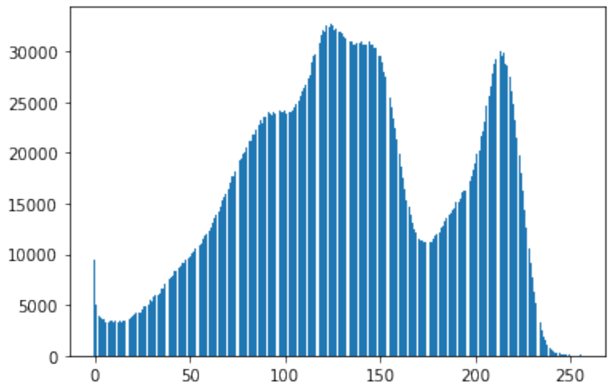
All the code implemented is in the file sol1.ipynb

1 Q1 - Quantization

1.1 Picking an image

Our image is:

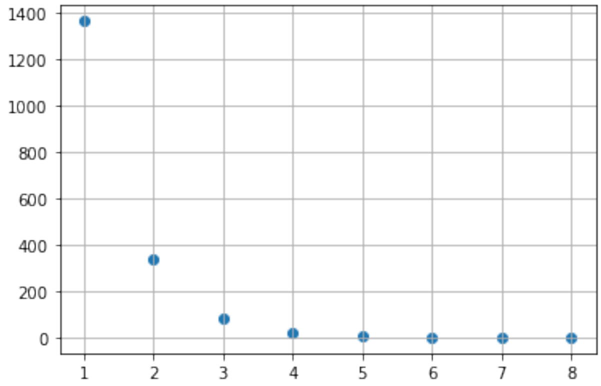
A 2048×2048 picture in grayscale of a kitten
Its histogram is:



Where 256 is a white and 0 is a black
The histogram is far enough from a uniform distribution

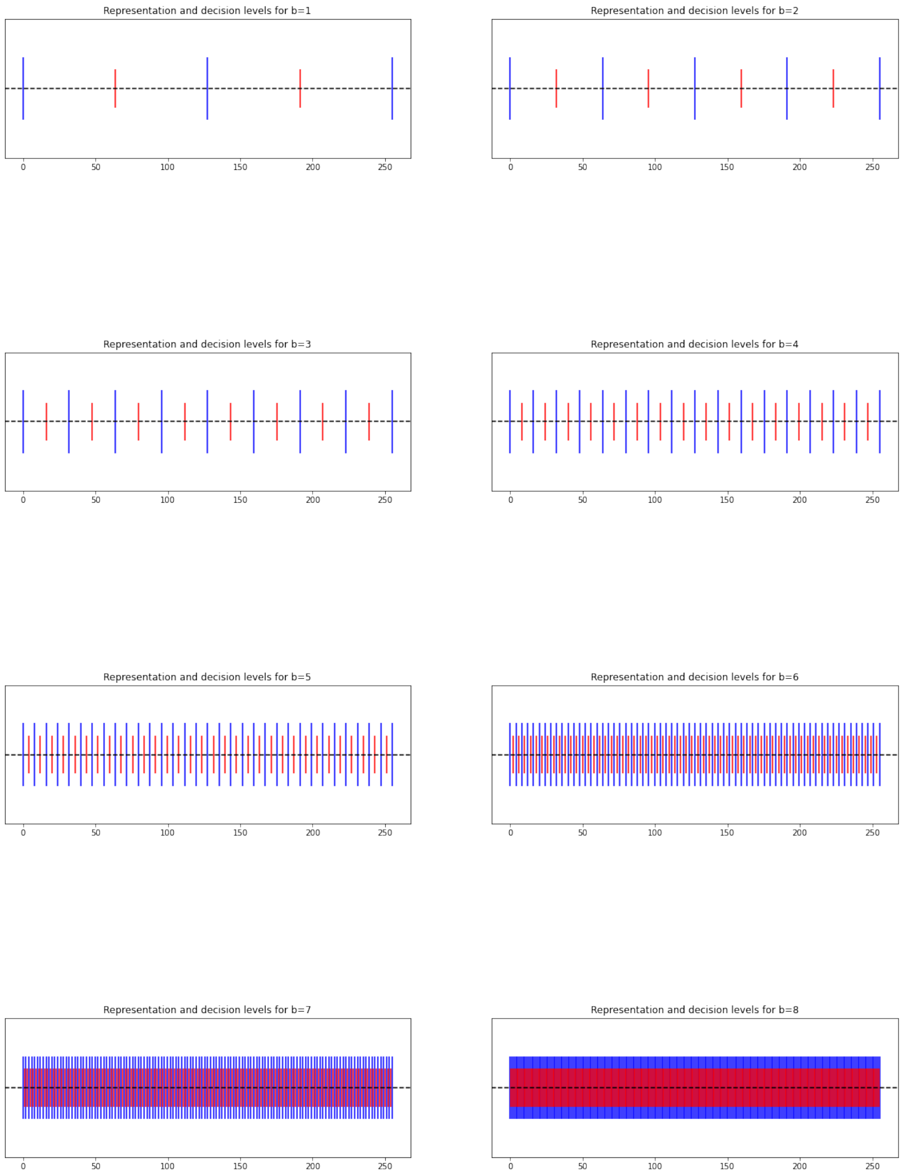
1.2 Uniform quantization

a The MSE as a function of b:



We can clearly see that the MSE error decreases as the bit budget grows, as expected.

b The decision and representation levels:

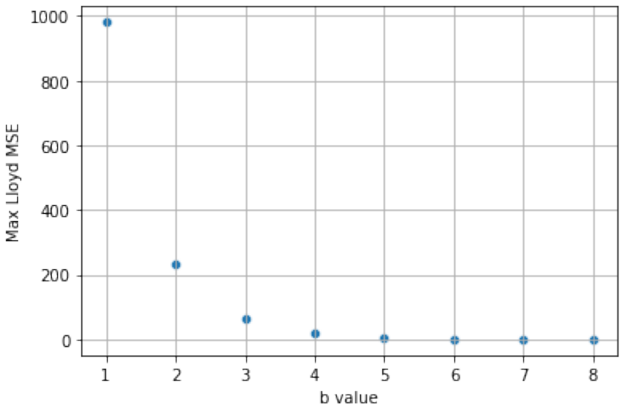


1.3 Max-Lloyd algorithm

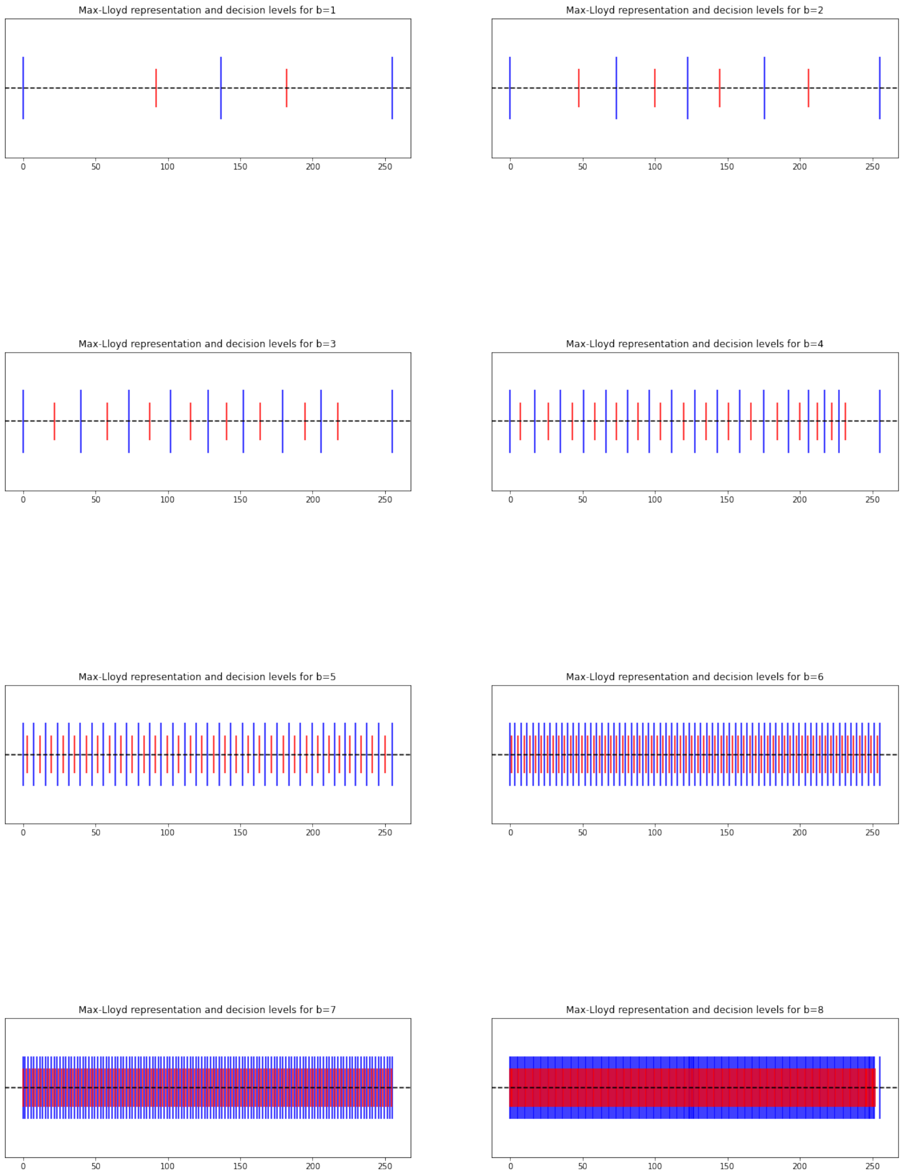
Pseudo-code and implementation is in sol1.ipynb

1.4 Applying the Max-Lloyd quantizer

a MSE as a function of the bit budget for $b \in \{1, \dots, 8\}$:



b Decision and representation levels for b values:



c As we can see from the plots, the Max-Lloyd quantizer is better than a uniform quantizer, especially at lower b values.

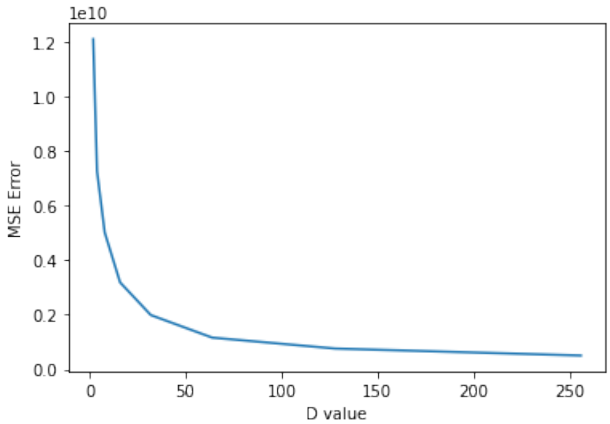
That is expected – as Max-Lloyd takes the actual distribution into account when choosing levels. As b get bigger, the difference is getting smaller until it's undistinguishable at high values of b.

2 Subsampling and Reconstruction

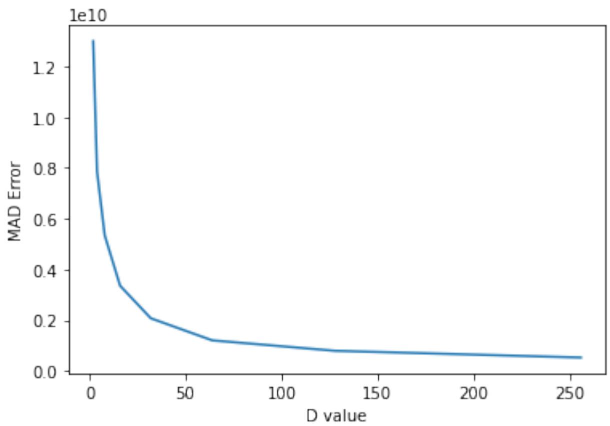
2.1 Sampling errors

We use $D = 2^b$, where b is the bit budget $b \in \{1, \dots 8\}$

a The sampling error in the MSE sense as a function of D:

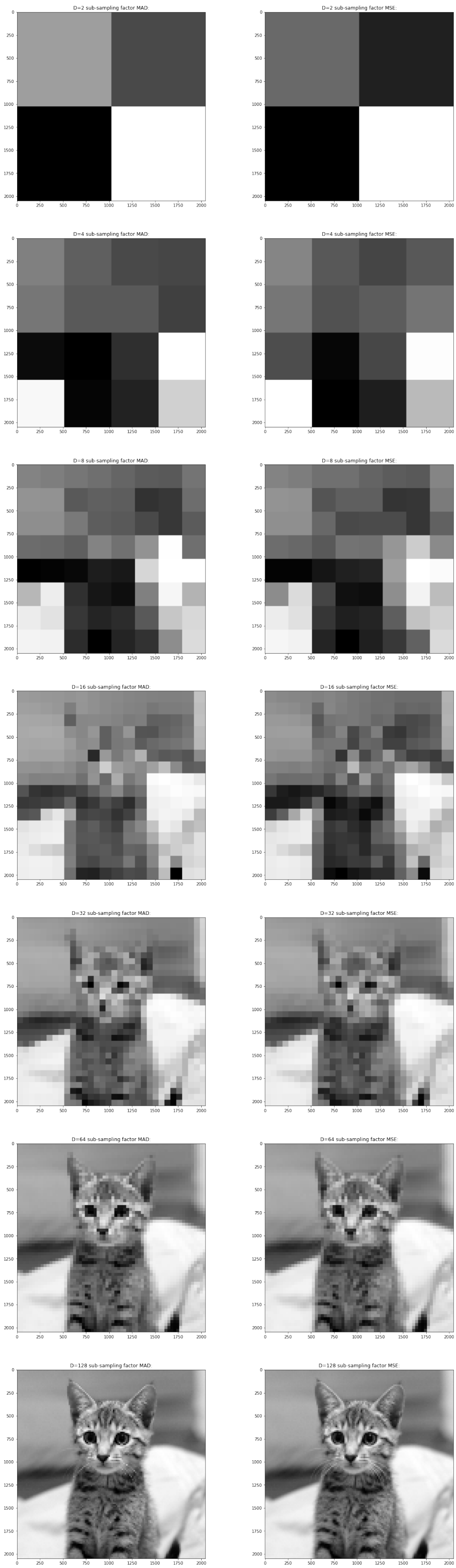


b The sampling error in the MAD sense as a function of D:



2.2 Reconstruction of the photo for MAD and MSE

The recontructed photos:


D=2 sub-sampling factor MAD
D=2 sub-sampling factor MSE:
D=4 sub-sampling factor MAD
D=4 sub-sampling factor MSE:
D=8 sub-sampling factor MAD
D=8 sub-sampling factor MSE:
D=16 sub-sampling factor MAD
D=16 sub-sampling factor MSE:
D=32 sub-sampling factor MAD
D=32 sub-sampling factor MSE:
D=64 sub-sampling factor MAD
D=64 sub-sampling factor MSE:
D=128 sub-sampling factor MAD
D=128 sub-sampling factor MSE:

2.3 Discussion

As we can see, low sampling factor means that we cannot understand what the subject of the image is.

As D is getting bigger, and more pixels appear in the photo we can see it more clearly.

In a photographer point of view - using MSE gives us a "blur" of the original image, whilst using MAD gives us a "high contrast" version of the image.

3 Solving L^p using L^2 solution

3.1 Pseudo code for L^p solver using L^2 solutions

We mark Δ_i as the rectangle $[i', i'+1] \times [j', j'+1]$ – one from the $D \times D$ domains.

Algorithm 1 L^p solver using L^2

Inputs
$f(x, y)$ – function to approximate
δ – convergence constant
Initialization
For each Δ_i we have:
$\hat{f}_i(x, y) \leftarrow 128 \cdot 1_{\Delta_i}(x, y)$
$w'_i(x, y) \leftarrow \min\left\{\frac{1}{\epsilon}, |f_i(x, y) - \hat{f}_i(x, y)|^{p-2} w(x, y)\right\}$, where $\epsilon > 0$ is a small number
Iterate (until the error L^p changes by less than constant δ)

1. $\delta_i \leftarrow \frac{\int_{\Delta_i} f(x,y) w'_i(x,y) dxdy}{\int_{\Delta_i} w'_i(x,y) dxdy}$
2. $\hat{f}_i \leftarrow \delta_i \cdot 1^{\Delta_i}(x, y)$
3. $w'_i(x, y) \leftarrow \min\left\{\frac{1}{\epsilon}, |f_i(x, y) - \hat{f}_i(x, y)|^{p-2} w(x, y)\right\}$

3.2 Implemetation of L^p solver

Implementation in sol1.ipynb

3.3 Empirical L^1 solver

Implementation in sol1.ipynb

3.4 Comparison between the results

D	Exact L^1 error	ϵ	L^p error for $p = 1$
4	43.01	1	47.89
		0.1	47.90
		0.01	48.01
		0.001	47.60
		0.0001	48.05
		0.00001	47.90
16	31.01	1	32.88
		0.1	32.95
		0.01	33.58
		0.001	33.72
		0.0001	33.21
		0.00001	33.25
64	17.57	1	18.26
		0.1	18.87
		0.01	18.69
		0.001	18.76
		0.0001	18.86
		0.00001	18.87

As expected, we can see that the approximation for L^1 wasn't as accurate as the exact L^1 calculation.

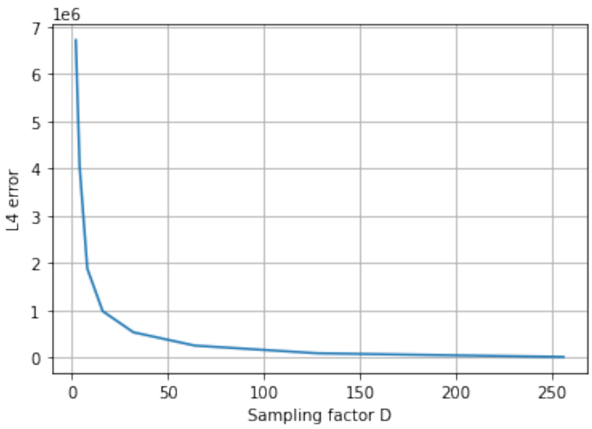
Higher subsampling factors got us better results, and we can see a small trend where smaller ϵ got us a higher iterative L^p error.

This can happen from the algorithm's tendency to pick more of the gray values used in each subsample we look at, rather than defaulting to a fixed small number, causing higher error to be generated.

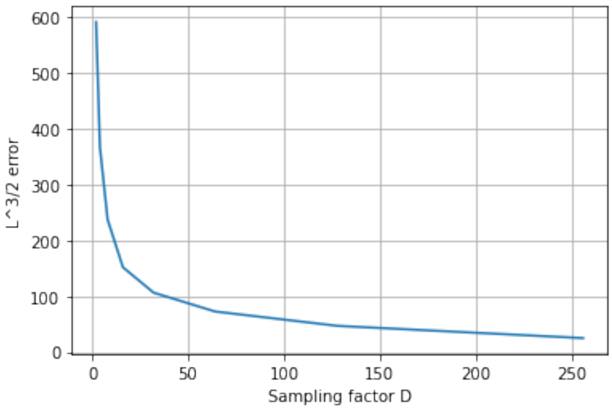
3.5 Running L^4 and $L^{\frac{3}{2}}$

We used the same L^p solver we wrote, only using $p = 4, \frac{3}{2}$.

The L^4 error we got, with respect to sampling factor D:



The $L^{\frac{3}{2}}$ error we got, with respect to sampling factor D:



As we can see from the results, the L^4 error's graph shows a much steeper decrease in error with small changes of sampling factor – and a plateau for high D values.

We can infer from this, and the resemblence to the L^2 and L^1 plots – that other values of p can also be used as a reconstruction technique.