Introduction to Data Processing and Representation
(236201)
Spring 2025

Homework 1

- **Published date: 09/04/2024**
- **Deadline date: 10/05/2025 23:59**

Guidelines:

- Submission is in pairs only.
- Submit your entire solution (including the theoretical part, the Python part and the Python code) electronically via the course website. The file should be a zip file containing the your PDF submission and Python code.
- The submission should be in English and in a clear printed form edited via computer software. We highly recommend using Latex through Overleaf.
- Rigorous mathematical proofs and reasoning are required for theoretical questions. Vague answers and unjustified claims will not be accepted.

I Theory

1. Solving the L^p problem using the L^2 solution

In this exercise we revisit the L^p uniform sampling problem for p a real scalar $p \geq 1$. Let N bet the number of samples. We partition $[0, 1]$ into N uniform intervals I_i. For a real function f defined on $[0, 1]$, we sample it by considering piece-wise constant functions $\hat{f}$ constant on each I_i. The weighted L^p sampling problem consists in solving the following optimization problem:

$$\min_{\hat{f}} \mathcal{E}^p(f, \hat{f}) = \min_{\hat{f}} \int_0^1 |f(x) - \hat{f}(x)|^p w(x) dx,$$

subject to the sampling constraint: $\hat{f}$ needs to be constant on each interval I_i, and where $w(x) > 0$ is a strictly positive weight function independent of f and $\hat{f}$. In this exercise we assume we are provided with a method capable of computing integrals.

a. Assume here that w is a constant function. Give, without proof, what is the optimal $\hat{f}_p$ when $p = 1$ and when $p = 2$.

b. For general w, what is the optimal $\hat{f}_p$ when $p = 2$?

c. For general w, what is the optimal $\hat{f}_p$ when $p = 1$? You may use the same level of preciseness as in the lectures rather than in the tutorial.

d. Prove that the optimization problem can be rewritten as a sum of N independent optimization problems depending solely on what happens in each interval. That is find $\mathcal{E}_i^p(f_i, \hat{f}_i)$ such that $\mathcal{E}^p(f, \hat{f}) = \sum_{i=1}^{N} \mathcal{E}_i^p(f_i, \hat{f}_i)$ where f_i and $\hat{f}_i$ are the functions f and $\hat{f}$ restrained to the interval I_i.

e. As in the case where $p = 1$, explicitly computing the values of $\hat{f}_p$ is non-trivial when $p \neq 2$. We thus wish to use the simplicity of the L^2 optimization problem to solve the general L^p optimization. In this question, fix $i \in \{1, \cdots, N\}$ and work in I_i, thus focus on $\mathcal{E}_i^p$.

   i. Assume that $f_i(x) \neq \hat{f}_i(x)$ for all $x \in I_i$. Find a positive function $w_{f_i,\hat{f}_i}$ depending on f_i and $\hat{f}_i$ such that $|f_i(x) - \hat{f}_i(x)|^p = w_{f_i,\hat{f}_i}(x)(f_i(x) - \hat{f}_i(x))^2$.

   ii. Under the same assumption, rewrite the optimization of $\mathcal{E}_i^p$ as a weighted L^2-like optimization problem except that in this new formulation the positive weight function $w'_{f_i,\hat{f}_i}$ may depend on f_i and $\hat{f}_i$.

iii. Under the same assumption, solving this L^2-like optimization problem is hard because the $w'_{f_i,\hat{f}_i}$ is not necessarily independent of $\hat{f}_i$. It would be much simpler if the weight function was independent of it. Why?

iv. When we remove the previous assumption, why do we prefer to use the function $\tilde{w}_{f_i,\hat{f}_i}(x) = \min\{\frac{1}{\varepsilon}, w_{f_i,\hat{f}_i}(x)\}$ instead of $w_{f_i,\hat{f}_i}(x)$, where $\varepsilon > 0$ a small fixed number?

From now on, we assume we have done this replacement. In particular, we replaced it in the previous new formulation of the optimization problem.

v. A classic algorithmic trick to solve this challenging problem is the following. Given a current $\hat{f}_i$ approximating f_i, assume that this provides a weight function $w'_i = w'_{f_i,\hat{f}_i}$. Consider that at the next step w'_i is a function independent of our next choice of $\hat{f}_i$. Find a new approximation $\hat{f}_i^{next}$ using this fixed w'_i. Repeat the process using $\hat{f}_i^{next}$ as $\hat{f}_i$. Write down in pseudo-code an algorithm implementing this idea.

f. Write a pseudo code for approximately solving the weighted L^p optimization problem using only L^2 optimizations.

g. What is the name of this algorithm? No points will be awarded to this question and we will not penalise the ignorant.

2. Haar matrix and Walsh-Hadamard matrix

Given $t \in [0, 1]$, consider the signal as

$$\phi(t) = a + b\cos(2\pi t) + c\cos^2(\pi t) \tag{1}$$

where a, b, and $c \in \mathbb{R}$ are constants. The procedures considered in this question for the approximation of $\phi(t)$ should be optimal with respect to the minimization of the approximation MSE, calculated over the continuous domain $[0, 1]$.

a. The 4×4 Haar matrix is given by

$$\boldsymbol{H}_4 = \frac{1}{2}\begin{bmatrix} 1 & 1 & \sqrt{2} & 0 \\ 1 & 1 & -\sqrt{2} & 0 \\ 1 & -1 & 0 & \sqrt{2} \\ 1 & -1 & 0 & -\sqrt{2} \end{bmatrix} \tag{2}$$

and its columns are used to form a set of 4 orthonormal functions, $\{\psi_i^H(t)\}_{i=1}^4$, defined for $t \in [0, 1]$, by using the change of basis from the standard basis with this matrix.

(i) Prove that $\boldsymbol{H}_4$ is unitary.

(ii) Show the set of orthonormal Haar functions $\{\psi_i^H(t)\}_{i=1}^4$. The functions should be presented using graphs with explicit notation of relevant values on the two axes.

(iii) What is the best approximation of ϕ using this Haar basis? What is the associated MSE?

(iv) Assume $a \geq b \geq 0$ and $c \geq 0$. What is the best 1-term approximation of ϕ? What is the best 2-term approximation of ϕ? What is the best 3-term approximation of ϕ? What is the best 4-term approximation of ϕ?

(v) Assume $a = \frac{1}{\pi}$, $b = 1$, and $c = \frac{3}{2}$. What is the best 1-term approximation of ϕ? What is the best 2-term approximation of ϕ? What is the best 3-term approximation of ϕ? What is the best 4-term approximation of ϕ?

b. The 4×4 Walsh-Hadamard matrix is given by

$$\boldsymbol{W}_4 = \frac{1}{2}\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & -1 & 1 & -1 \end{bmatrix} \tag{3}$$

and its columns are used to form a set of 4 orthonormal functions, $\{\chi_i^W(t)\}_{i=1}^4$, defined for $t \in [0, 1]$.

(i) Prove that $\boldsymbol{W}_4$ is unitary.

(ii) Show the set of orthonormal Walsh-Hadamard functions $\{\psi_i^W(t)\}_{i=1}^4$. The functions should be presented using graphs with explicit notation of relevant values on the two axes.

(iii) What is the best approximation of ϕ using this Walsh-Hadamard basis? What is the associated MSE?

(iv) Assume $a \geq b \geq 0$ and $c \geq 0$. What is the best 1-term approximation of ϕ? What is the best 2-term approximation of ϕ? What is the best 3-term approximation of ϕ? What is the best 4-term approximation of ϕ?

(v) Assume $a = \frac{1}{\pi}$, $b = 1$, and $c = \frac{3}{2}$. What is the best 1-term approximation of ϕ? What is the best 2-term approximation of ϕ? What is the best 3-term approximation of ϕ? What is the best 4-term approximation of ϕ?

II Implementation

Instruction:

- In this part use a gray-scale image (of 256 gray levels) and of a size of at least 512×512 pixels.
- Along this part show the image in the same size and range.
- Figures should be appropriately titled and the font size should be set appropriately clear.
- **Your submission should include a report describing your results and conclusions.** The code should be well documented.
- **Figures and graphs should be presented in the report in conjunction to explanations showing your understanding of the exercise and its results.**
- **Important: You should implement by yourselves the requested algorithms and procedures (such as quantizers, samplers and error-calculation functions), and not rely on the corresponding functions available in Python or on the web.**

1. Quantization

1. We would like to estimate the probability density function (pdf) of the gray levels in the image using the image histogram. If the histogram seems too uniform, please pick another image with a non-uniform distribution.

2. Apply uniform quantization on the image using b bits per pixel.

   a. Show the MSE as a function of the bit-budget b for $b = 1, ..., 8$.

   b. Plot the decision and representation levels for representative b values.

3. Implement the Max-Lloyd algorithm. This should be a function taking as input a histogram pdf, a vector of initial decision levels, and a small value $\varepsilon > 0$ for convergence tolerance. The function should return the converged decision levels and the converged representation levels. In order to handle numerical approximations, we use ε as a stopping condition: when the MSE improves by less than ε we stop the algorithm.

4. Apply the Max-Lloyd quantizer starting with uniform quantization.

a. Show the MSE as a function of the bit-budget b for $b = 1, ..., 8$.

b. Plot the decision and representation levels for representative b values.

c. Compare the results to those of the uniform quantizer. Explain the differences.

2. Subsampling and Reconstruction

1. Consider an image as a discrete 2D sampled signal denoted as $\boldsymbol{I}(x, y)$, where x and y are the position indices on the images. Crop or resize your image such that its number of rows and columns are a power of 2 greater than 8. For integer sub-sampling factors $D = 2^1, 2^2, \cdots, 2^8$, grid the image domain uniformly in x and y, giving $N_x \times N_y$ uniform rectangular grid sample regions. Each region will be subsampled using a unique optimal number in some sense.

a. In the MSE sense, present the sub-sampled image for all different D. Denote these sub-sampled images by $\{\tilde{\boldsymbol{J}}_i\}_{i=1}^{8}$. Show the MSE as a function of the integer sub-sampling factor D.

b. In the MAD sense, present the sub-sampled image for all different D. Denote these sub-sampled images by $\{\hat{\boldsymbol{J}}_i\}_{i=1}^{8}$. Show the MAD as a function of the integer sub-sampling factor D.

2. Reconstruct in the standard way $\{\tilde{\boldsymbol{J}}_i\}_{i=1}^{8}$ and $\{\hat{\boldsymbol{J}}_i\}_{i=1}^{8}$ back to the same size of the original image, denoted by $\{\tilde{\boldsymbol{K}}_i\}_{i=1}^{8}$ and $\{\hat{\boldsymbol{K}}_i\}_{i=1}^{8}$. Present $\{\tilde{\boldsymbol{K}}_i\}_{i=1}^{8}$ and $\{\hat{\boldsymbol{K}}_i\}_{i=1}^{8}$.

3. Discuss how the integer sub-sampling factor D affects the result.

3. Hadamard, Hadamard-Walsh, and Haar matrices

a. Implement Hadamard matrices $\boldsymbol{H}_{2^n}$. This should be a function taking as input the level n. This function should return a $2^n \times 2^n$ matrix.

b. Take the two orthonormal families $\boldsymbol{H}_{2^n}$ and $\{\sqrt{2^n}\mathbf{1}_{\Delta_i}(t)\}_{i=1}^{2^n}$ into a new set of functions $\{h_i(t)\}_{i=1}^{2^n}$ by

$$\begin{pmatrix} h_1(t) \\ h_2(t) \\ \vdots \\ h_{2^n}(t) \end{pmatrix} = \boldsymbol{H}_{2^n}^{\top} \begin{pmatrix} \sqrt{2^n}\mathbf{1}_{\Delta_1}(t) \\ \sqrt{2^n}\mathbf{1}_{\Delta_2}(t) \\ \vdots \\ \sqrt{2^n}\mathbf{1}_{\Delta_{2^n}}(t) \end{pmatrix} \tag{4}$$

Plot the functions $\{h_i(t)\}_{i=1}^{2^n}$ for $n = 2, ..., 6$.

c. Implement Walsh-Hadamard matrices $\widetilde{\boldsymbol{H}}_{2^n}$. This should be a function taking as input Hadamard matrices $\boldsymbol{H}_{2^n}$. This function should return a $2^n \times 2^n$ matrix.

d. Take the two orthonormal families $\widetilde{\boldsymbol{H}}_{2^n}$ and $\{\sqrt{2^n}\mathbf{1}_{\Delta_i}(t)\}_{i=1}^{2^n}$ into a new set of functions $\{hw_i(t)\}_{i=1}^{2^n}$ by

$$\begin{pmatrix} hw_1(t) \\ hw_2(t) \\ \vdots \\ hw_{2^n}(t) \end{pmatrix} = \widetilde{\boldsymbol{H}}_{2^n}^\top \begin{pmatrix} \sqrt{2^n}\mathbf{1}_{\Delta_1}(t) \\ \sqrt{2^n}\mathbf{1}_{\Delta_2}(t) \\ \vdots \\ \sqrt{2^n}\mathbf{1}_{\Delta_{2^n}}(t) \end{pmatrix} \tag{5}$$

Plot the functions $\{hw_i(t)\}_{i=1}^{2^n}$ for $n = 2, ..., 6$.

e. Implement Haar matrices $\hat{\boldsymbol{H}}_{2^n}$. This should be a function taking as input the level n. This function should return a $2^n \times 2^n$ matrix. Haar matrices are defined recursively as follows. $H_{2N} = \begin{pmatrix} H_N \otimes (1,1) \\ I_N \otimes (1,-1) \end{pmatrix}$, with $H_2 = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$. Recall the definition of the Kronecker product between A and B is

$$A \otimes B = \begin{pmatrix} a_{1,1}B & a_{1,2}B & \dots & a_{1,n}B \\ a_{2,1}B & a_{2,2}B & \dots & a_{2,n}B \\ \vdots & \ddots & \ddots & \vdots \\ a_{n,1}B & a_{n,2}B & \dots & a_{n,n}B \end{pmatrix}.$$

f. Take the two orthonormal families $\hat{\boldsymbol{H}}_{2^n}$ and $\{\sqrt{2^n}\mathbf{1}_{\Delta_i}(t)\}_{i=1}^{2^n}$ into a new set of functions $\{ha_i(t)\}_{i=1}^{2^n}$ by

$$\begin{pmatrix} ha_1(t) \\ ha_2(t) \\ \vdots \\ ha_{2^n}(t) \end{pmatrix} = \hat{\boldsymbol{H}}_{2^n}^\top \begin{pmatrix} \sqrt{2^n}\mathbf{1}_{\Delta_1}(t) \\ \sqrt{2^n}\mathbf{1}_{\Delta_2}(t) \\ \vdots \\ \sqrt{2^n}\mathbf{1}_{\Delta_{2^n}}(t) \end{pmatrix} \tag{6}$$

Plot the functions $\{ha_i(t)\}_{i=1}^{2^n}$ for $n = 2, ..., 6$.

g. Given $t \in [-4, 5]$, consider a function

$$\phi(t) = t\exp(t). \tag{7}$$

Consider $n = 2$, what are the best k-term approximation of $\phi(t)$ for $k = 1, ..., 2^n$ in each basis? Present the results on a graph. What are the corresponding MSE errors?