Introduction to Data Processing and Representation- HW1

Gal Kesten: 316353176 Chen Pery: 313283657

June 2024

1 Theory

1.1 Solving the L^p problem using the L^2 solution

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

Solution:

Since w is a constant, we can take it out of the integral.
The argmin of the optimization problem is not affected by multiplying by a constant (though the minimum value itself will change).
Therefore, we get the same optimization problems we already saw in class for L^1 and L^2 norms.

For $p = 1$: The optimal $\hat{f}_p(x) = \sum_{i=1}^{N} median_i \cdot 1\Delta_i(x)$ where $median_i$ is the median of f over each interval I_i.
For $p = 2$: The optimal $\hat{f}_p(x) = \sum_{i=1}^{N} average_i \cdot 1\Delta_i(x)$ where $average_i = \frac{1}{|\Delta_i|} \int_{\Delta_i} f(x)\,dx$

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

Solution:

We aim to minimize the weighted L^2 error:

$$\mathcal{E}^2(f, \hat{f}) = \int_0^1 |f(x) - \hat{f}(x)|^2 w(x)\, dx$$

By the additivity property of the integral we can write

$$\mathcal{E}^2(f, \hat{f}) = \sum_{i=1}^{N} \int_{\Delta_i} |f(x) - \hat{f}(x)|^2 w(x)\, dx$$

In each interval I_i $\hat{f}(x) = \hat{f}_i$ where $\hat{f}_i \in \mathcal{R}$

Therefore, we get:

$$\mathcal{E}^2(f, \hat{f}) = \sum_{i=1}^{N} \int_{\Delta_i} |f(x) - \hat{f}_i|^2 w(x)\, dx \overset{(1)}{=}$$

$$\int_0^1 f(x)^2 w(x)\, dx - 2\sum_{i=1}^{N} \hat{f}_i \int_{\Delta_i} f(x) w(x)\, dx + \sum_{i=1}^{N} \hat{f}_i^2 \int_{\Delta_i} w(x)\, dx$$

Therefore we can write our optimization problem as follows:

$$\min_{\hat{f}_i \ldots \hat{f_N}} \int_0^1 f(x)^2 w(x)\, dx - 2\sum_{i=1}^{N} \hat{f}_i \int_{\Delta_i} f(x) w(x)\, dx + \sum_{i=1}^{N} \hat{f}_i^2 \int_{\Delta_i} w(x) dx$$

This problem is a convex optimization problem(local minimum is the global minimum) and the objective function is differential; Therefore, it is sufficient to check when $\nabla \mathcal{E} = 0$.

$$\frac{\partial}{\partial \hat{f}_i} \left(\int_0^1 f(x)^2 w(x)\, dx - 2\sum_{i=1}^{N} \hat{f}_i \int_{\Delta_i} f(x) w(x)\, dx + \sum_{i=1}^{N} \hat{f}_i^2 \int_{\Delta_i} w(x)\, dx \right) \overset{(2)}{=}$$

$$-2\int_{\Delta_i} f(x) w(x)\, dx + 2\hat{f}_i \int_{\Delta_i} w(x)\, dx$$

Solving for $\frac{\partial \mathcal{E}}{\partial \hat{f}_i} = 0$ we get:

$$2\hat{f}_i \int_{\Delta_i} w(x)\,dx = 2\int_{\Delta_i} f(x)w(x)\,dx \implies \hat{f}_i = \frac{\int_{\Delta_i} f(x)w(x)\,dx}{\int_{\Delta_i} w(x)\,dx}$$

Therefore, the optimal $\hat{f}_p$ when $p = 2$ is:

$$\boxed{\hat{f}_p(x) = \sum_{i=1}^{N} \left(\frac{\int_{\Delta_i} f(x)w(x)\,dx}{\int_{\Delta_i} w(x)\,dx} \right) 1_{\Delta_i}(x)}$$

Explanations
1. Linearity + additivity of the integral 2. Linearity of partial derivative

c. For general w, what is the optimal $\hat{f}_p$ when $p = 1$?

Solution:

We aim to minimize the weighted L^1 error:

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

Since f is a piecewise constant (PC) function, we can write:

$$\mathcal{E}^1(f, \hat{f}) = \sum_{i=1}^{N} \int_{\Delta_i} |f(x) - \hat{f}_i| w(x)\,dx,\ \hat{f}_i \in \mathcal{R}$$

As we saw in lectures, to minimize $\mathcal{E}^1$, we should search for $\nabla \mathcal{E}^1 = 0$

$$\frac{\partial \mathcal{E}^1}{\partial \hat{f}_i} \overset{(1)}{=} \sum_{i=1}^{N} \int_{\Delta_i} \frac{\partial}{\partial \hat{f}_i} |f(x) - \hat{f}_i| w(x)\,dx \overset{(2)}{=} \int_{\Delta_i} -\text{sign}(f(x) - \hat{f}_i) w(x)\,dx$$

Setting the derivative to zero, we get:

$$\int_{\Delta_i} -\text{sign}(f(x) - \hat{f}_i) w(x)\,dx = 0$$

Which implies:

$$\int_{\Delta_i} \text{sign}(f(x) - \hat{f}_i) w(x)\, dx = 0$$

The integral will be zero when the weight of the signs of the expressions $\text{sign}(f(x) - \hat{f}_i)$ cancels out. This is achieved when $\hat{f}_i$ is the weighted median of $f(x)$ over Δ_i with weight $w(x)$.

The weighted median of $f(x)$ can be interpreted as: the weight for which $f(x)$ is below $\hat{f}_i$ is equal to the weight for which $f(x)$ is above $\hat{f}_i$:

$$\int_{I:f(x)<\hat{f}_i} w(x)\, dx = \int_{I:f(x)>\hat{f}_i} w(x)\, dx$$

(There can also be a compensation term for the case where $\hat{f}_i = f(x)$.)

Therefore, the optimal $\hat{f}_p$ when $p = 1$ is:

$$\boxed{\hat{f}_p(x) = \sum_{i=1}^{N} \text{weighted median in } \Delta_i * 1_{\Delta_i}(x)}$$

Explanations

1. Leibniz rule + Linearity of the partial derivative
2. $\frac{\partial}{\partial \hat{f}_i} |f(x) - \hat{f}_i| = \text{sign}(f(x) - \hat{f}_i)$

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.

Solution:
The objective function is:

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

By the additivity of the integral, we can write the equation as follows:

$$\mathcal{E}^p(f, \hat{f}) = \sum_{i=1}^{N} \int_{\Delta_i} |f(x) - \hat{f}(x)|^p w(x)\, dx$$

$\hat{f}(x)$ is a piecewise constant (PC), so we have $\hat{f}(x) = \sum_{i=1}^{N} \hat{f}_i 1_{\Delta_i}(x)$ where $\hat{f}_i \in \mathbb{R}$.

We will denote the restriction of $\hat{f}(x)$ to the interval Δ_i as $\hat{f}_i(x)$. Since this is a PC function we can simply write $\hat{f}_i(x) = \hat{f}_i$.
We will also denote the restriction of $f(x)$ to the interval Δ_i as $f_i(x)$

Therefore, we can write:

$$\mathcal{E}^p(f, \hat{f}) = \sum_{i=1}^{N} \int_{\Delta_i} |f(x) - \hat{f}(x)|^p w(x)\, dx = \sum_{i=1}^{N} \int_{\Delta_i} |f_i(x) - \hat{f}_i(x)|^p w(x)\, dx =$$

$$\sum_{i=1}^{N} \int_{\Delta_i} |f_i(x) - \hat{f}_i|^p w(x)\, dx$$

For each interval I_i, we define:

$$\mathcal{E}_i^p(f_i, \hat{f}_i) = \sum_{i=1}^{N} \int_{\Delta_i} |f_i(x) - \hat{f}_i|^p w(x)\, dx$$

where $f_i(x)$ and $\hat{f}_i(x) = \hat{f}_i$ are the restrictions of $f(x)$ and $\hat{f}(x)$ to the interval Δ_i.

Thus, we get:

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

as requested.

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

Solution:
Define

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

$w_{f_i,\hat{f}_i}(x)$ is defined since we know that $f_i(x) \neq \hat{f}_i(x)$. It's a positive function since $|f_i(x) - \hat{f}_i(x)|$ is positive and therefore $|f_i(x) - \hat{f}_i(x)|^p$ for $p \geq 1$.

In addition, we get:

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

Therefore we found the requested function $w_{f_i,\hat{f}_i}(x)$.

e. 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 $\hat{f}_i$.

Solution:

Given:

$$\mathcal{E}_i^p(f_i, \hat{f}_i) = \int_{\Delta_i} |f_i(x) - \hat{f}_i(x)|^p w(x)\, dx$$

Using the relation:

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

We can rewrite the objective function as follows:

$$\mathcal{E}_i^p(f_i, \hat{f}_i) = \int_{\Delta_i} w_{f_i,\hat{f}_i}(x)(f_i(x) - \hat{f}_i(x))^2 w(x)\, dx = \int_{\Delta_i} (f_i(x) - \hat{f}_i(x))^2 w'_{f_i,\hat{f}_i}(x)$$

where $w'_{f_i,\hat{f}_i}(x) = w_{f_i,\hat{f}_i}(x)w(x)$
Since $\hat{f}_i(x) = \hat{f}_i$ for $\hat{f}_i \in \mathbb{R}$
The problem can be rewritten as

$$\min_{\hat{f}_i \in \mathbb{R}} \mathcal{E}_i^p(f_i, \hat{f}_i) = \int_{\Delta_i} (f_i(x) - \hat{f}_i)^2 w'_{f_i,\hat{f}_i}(x)$$

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

Solution:
If the weight function is independent of $\hat{f}_i$ we get a quadratic objective function w.r.t to $\hat{f}_i$:

$$\mathcal{E}_i^p(f_i, \hat{f}_i) = \int_{\Delta_i} (f_i(x) - \hat{f}_i)^2 w(x)\, dx = f_i^2 \int_{\Delta_i} w(x)\, dx - 2\hat{f}_i \int_{\Delta_i} f_i(x)w(x)\, dx + \int_{\Delta_i} f_i(x)^2 w(x)\, dx$$

Since $w(x) > 0$, we also have $\int_{\Delta_i} w(x)\, dx > 0$ (monotonicity of the integral). Minimizing this objective w.r.t $\hat{f}_i$ is a convex optimization problem with a closed analytical solution (weighted average as we saw in the previous exercises).
However, if $w'_{f_i,\hat{f}_i}$ depends on $\hat{f}_i$, our optimization problem with respect to $\hat{f}_i$ is:

$$\min_{\hat{f}_i} \mathcal{E}_i^p(f_i, \hat{f}_i) = \min_{\hat{f}_i} \int_{\Delta_i} (f_i(x) - \hat{f}_i)^2 w'_{f_i,\hat{f}_i}(x)\, dx$$

This objective function is not necessarily convex w.r.t $\hat{f}_i$, so local minima may not be global minima as before. We also won't be able to take out the variable $\hat{f}_i$ from the integral as before, and we will have to use the Leibniz rule when calculating the partial derivative w.r.t $\hat{f}_i$. This means that we will be left with calculating:

$$\int_{\Delta_i} \frac{\partial}{\partial \hat{f}_i} \left[(f_i(x) - \hat{f}_i)^2 w'_{f_i,\hat{f}_i}(x) \right] \, dx = 0$$

This problem is much more complex to solve and there might not be a closed analytical solution to this problem.

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

Solution:
We prefer this function because of the following reasons:
1) Since

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

for $f_i(x) = \hat{f}_i(x)$, the weight function is not defined. The new formulation ensures that the weight function is defined over the entire domain, preventing division by zero or undefined values.
2)In the previous formulation, the weight function can become extremely large when $f_i(x)$ is close to $\hat{f}_i$ and $p \leq 2$. This can lead to numerical instability and overly large weights that skew the optimization process. Conversely, when $p > 2$, the weight function can become extremely large if there is a significant difference between $f_i(x)$ and $\hat{f}_i$. This also introduces numerical instability and can make the optimization process unreliable. The new weight function $\tilde{w}_{f_i,\hat{f}_i}(x)$ is bounded by $\frac{1}{\epsilon}$, ensuring that the weights do not become excessively large or small. This boundedness leads to more numerically stable optimization, avoiding the extremes that could disrupt the optimization algorithm.

e.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 $\hat{f}_i^{\text{next}}$ using this fixed w'_i. Repeat the process using $\hat{f}_i^{\text{next}}$ as $\hat{f}_i$. Write down in pseudo-code an algorithm implementing +this idea.

Solution:

Algorithm: SolveLpUsingL2SingleInterval

Input: function $f_i(x)$, weights $w(x), \epsilon > 0, p$

1. Initialize: set/guess $\hat{f}_i^{next}$

2. While stopping condition is not met:

   a. Compute the weight function :
   $w'_i = w'_{f_i,\hat{f}_i^{next}}(x) = \min\left\{\frac{1}{\epsilon}, \frac{|f_i(x)-\hat{f}_i^{(next)}|^p}{|f_i(x)-\hat{f}_i^{(next)}|^2} w(x)\right\}$

   b. Compute the new approximation $\hat{f}_i^{(next)}$:
   $\hat{f}_i^{(next)} = \frac{\int_{\Delta_i} f_i(x) w'_i(x)\, dx}{\int_{\Delta_i} w'_i(x)\, dx}$

3. Return $\hat{f}_i^{(next)}$

possible stopping conditions are reaching a number of iterations (FOR loop), the error changes by less than a threshold

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

Solution:

```
Algorithm: SolveLpUsingL2NIntervals

Input: function f(x) : [0,1] → ℝ, weights w(x), ε > 0, N, p

1. Initialize result ← {}

2. For i = 1 to N: (can be computed simultaneously)
   a. Restrict f to the interval Δ_i: f_i(x) = f(x) · 1_{Δ_i}(x)
   b. Compute f̂_i using SolveLpUsingL2SingleInterval:
   f̂_i ← SolveLpUsingL2SingleInterval(f_i(x), w(x), ε, p)
   c. Append f̂_i to result: result ← result ∪{f̂_i}

3. Return result
```

We chose to use the algorithm from the previous section to solve the entire L^p problem by solving N

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

Solution: IRLS

2. Signal Discretization using a Piecewise-Linear Approximation

a. Show that for a positive integer k:

$$\int_{t\in\Delta_i} (t-t_i)^k \, dt = \begin{cases} 0, & \text{if } k \text{ is odd} \\ \frac{|\Delta_i|^{k+1}}{2^k(k+1)}, & \text{if } k \text{ is even} \end{cases}$$

where $|\Delta_i|$ is the size of the interval.

Solution:

Since the interval is defined as $\Delta_i = \left[\frac{i-1}{N}, \frac{i}{N}\right)$, the center t_i of the interval Δ_i is:

$$t_i = \frac{(i-1)+i}{2N} = \frac{2i-1}{2N}$$

We need to calculate

$$\int_{t\in\Delta_i} (t-t_i)^k \, dt$$

Using U-substitution for solving the integral:

$$u = t - t_i, du = dt, u_{\min} = \frac{i-1}{N} - t_i = -\frac{1}{2N}, u_{\max} = \frac{i}{N} - t_i = \frac{1}{2N}$$

we get the following integral:

$$\int_{t\in\Delta_i} (t-t_i)^k \, dt = \int_{-\frac{1}{2N}}^{\frac{1}{2N}} u^k \, du = \left.\frac{u^{k+1}}{k+1}\right|_{-\frac{1}{2N}}^{\frac{1}{2N}} = \frac{1}{k+1}((\frac{1}{2N})^{k+1} - (-\frac{1}{2N})^{k+1})$$

$$= \frac{1}{k+1}(\frac{1}{2^{k+1}}|\Delta_i|^{k+1}(1-(-1)^{k+1}) = \begin{cases} 0 & \text{if } k \text{ is odd} \\ \frac{2|\Delta_i|^{k+1}}{2^{k+1}(k+1)} = \frac{|\Delta_i|^{k+1}}{2^k(k+1)}, & \text{if } k \text{ is even} \end{cases}$$

b. What are the optimal coefficients a_i and c_i that minimize the MSE of representing the entire signal using N intervals?

Solution:

We aim to solve the following optimization problem:

$$\min_{\hat{\phi}} \int_0^1 (\phi(t) - \hat{\phi}(t))^2 \, dt = \min_{\hat{\phi}} \Psi_{MSE}(\phi \to \hat{\phi}))$$

It holds that

$$\Psi_{MSE}(\phi \to \hat{\phi}) = \int_0^1 (\phi(t) - \hat{\phi}(t))^2 \, dt \overset{(1)}{=} \sum_{I=1}^{N} \int_{\Delta_i} (\phi(t) - \hat{\phi}(t))^2 \, dt =$$

$$\sum_{I=1}^{N} (\int_{\Delta_i} \phi(t)^2 \, dt - 2 \int_{\Delta_i} \hat{\phi}(t)\phi(t) \, dt + \int_{\Delta_i} \hat{\phi}(t)^2 \, dt) \overset{(2)}{=}$$

$$\int_0^1 \phi(t)^2 \, dt - 2 \sum_{i=1}^{n} \int_{\Delta_i} \hat{\phi}(t)\phi(t) \, dt + \sum_{i=1}^{N} \int_{\Delta_i} \hat{\phi}(t)^2 \, dt =$$

Since $\hat{\phi}(t) = a_i(t - t_i) + c_i$ for $t_i \in \Delta_i$, we can write:

$$\Psi_{MSE}(\phi \to \hat{\phi}) = \int_0^1 \phi(t)^2 \, dt - 2 \sum_{i=1}^{n} \int_{\Delta_i} (a_i(t - t_i) + c_i)\phi(t) \, dt + \sum_{i=1}^{N} \int_{\Delta_i} (a_i(t - t_i) + c_i)^2 \, dt) =$$

$$\int_0^1 \phi(t)^2 \, dt - 2 \sum_{i=1}^{n} \int_{\Delta_i} a_i(t - t_i)\phi(t) + c_i\phi(t) \, dt + \sum_{i=1}^{N} \int_{\Delta_i} a_i^2(t - t_i)^2 + 2a_i c_i(t - t_i) + c_i^2 \, dt \overset{(3)}{=}$$

$$\int_0^1 \phi(t)^2 \, dt - 2 \sum_{i=1}^{N} a_i \int_{\Delta_i} (t - t_i)\phi(t) \, dt - 2 \sum_{i=1}^{N} c_i \int_{\Delta_i} \phi(t) \, dt + \sum_{i=1}^{N} a_i^2 \int_{\Delta_i} (t - t_i)^2 \, dt$$

$$+ \sum_{i=1}^{N} a_i c_i \int_{\Delta_i} (t - t_i) \, dt + c_i^2 \sum_{i=1}^{N} \int_{\Delta_i} dt \overset{(4)}{=}$$

$$\int_0^1 \phi(t)^2 \, dt - 2 \sum_{i=1}^{N} a_i \int_{\Delta_i} (t - t_i)\phi(t) \, dt - 2 \sum_{i=1}^{N} c_i \int_{\Delta_i} \phi(t) \, dt + \sum_{i=1}^{N} a_i^2 \frac{|\Delta_i|^3}{2^2 3} + \sum_{i=1}^{N} c_i^2 |\Delta_i| =$$

$$\int_0^1 \phi(t)^2\,dt - 2\sum_{i=1}^N a_i \int_{\Delta_i} (t-t_i)\phi(t)\,dt - 2\sum_{i=1}^N c_i \int_{\Delta_i} \phi(t)\,dt + \sum_{i=1}^N a_i^2 \frac{|\Delta_i|^3}{12} + \sum_{i=1}^N c_i^2 |\Delta_i|$$

We get convex $\Psi_{MSE}(\phi \to \hat{\phi})$ w.r.t a_i, c_i since the coefficients of a_i^2, c_i^2 are positive (only these terms will be different than zero in the hessian matrix and they will be on diagonal so we get PD hessian). Therefore, to minimize $\Psi_{MSE}(\phi \to \hat{\phi})$, we need to take the partial derivatives w.r.t a_i, c_i and set them to 0.

$$\frac{\partial \Psi_{MSE}(\phi \to \hat{\phi})}{\partial a_i} \overset{(5)}{=} 2a_i \frac{|\Delta_i|^2 3}{12} - 2\int_{\Delta_i} (t-t_i)\phi(t)\,dt$$

$$a_i \frac{|\Delta_i|^3}{6} - 2\int_{\Delta_i} (t-t_i)\phi(t)\,dt = 0 \Rightarrow \boxed{a_i^* = \frac{12\int_{\Delta_i}(t-t_i)\phi(t)\,dt}{\Delta_i^3} = 12N^3 \int_{\Delta_i} (t-t_i)\phi(t)\,dt}$$

$$\frac{\partial \Psi_{MSE}(\phi \to \hat{\phi})}{\partial c_i} \overset{(6)}{=} 2c_i|\Delta_i| - 2\int_{\Delta_i} \phi(t)\,dt$$

$$2c_i|\Delta_i| - 2\int_{\Delta_i} \phi(t)\,dt = 0 \Rightarrow \boxed{c_i = \frac{\int_{\Delta_i}\phi(t)\,dt}{|\Delta_i|} = \int_{\Delta_i} \phi(t) \cdot N}$$

Explanations

(1) Additivity of integration
(2) Linearity and additivity of the Integral
(3)Linearity of the Integral
(4) We proofed $\int_{\Delta_i}(t-t_i)\,dt = 0$ and $\int_{\Delta_i}(t-t_i)^2\,dt = \frac{|\Delta_i|^{2+1}}{2^2(2+1)}$
(5) Linearity of the partial derivative
(6) Linearity of the partial derivative

c. Formulate the minimal MSE of representing the entire signal using N intervals?

Solution:

We saw in the previous exercise that :

$$\Psi_{MSE}(\phi \to \hat{\phi}) = \int_0^1 \phi(t)\,dt - 2\sum_{i=1}^{N} a_i \int_{\Delta_i} (t-t_i)\phi(t)\,dt - 2\sum_{i=1}^{N} c_i \int_{\Delta_i} \phi(t)\,dt + \sum_{i=1}^{N} a_i^2 \frac{|\Delta_i|^3}{12} + \sum_{i=1}^{N} c_i^2 |\Delta_i|$$

Plug in optimal $a_i^* = 12N^3 \int_{\Delta_i}(t-t_i)\phi(t)\,dt$, $c_i = N\int_{\Delta_i}\phi(t)\,dt$

$\Psi_{MSE}(\phi \to \hat{\phi}^*) =$

$$\int_0^1 \phi(t)\,dt - 2\sum_{i=1}^{N} 12N^3 \int_{\Delta_i}(t-t_i)\phi(t)\,dt \int_{\Delta_i}(t-t_i)\phi(t)\,dt$$

$$-2\sum_{i=1}^{N} N\int_{\Delta_i}\phi(t)\,dt \int_{\Delta_i}\phi(t)\,dt + \sum_{i=1}^{N}(a_i^*)^2 \frac{|\Delta_i|^3}{12} + \sum_{i=1}^{N}(c_i^*)^2|\Delta_i| =$$

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

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

d.Compare the minimal MSE for using piecewise-linear approximation and the minimal MSE for using piecewise-constant approximation

Solution:
We saw in the tutorial that the optimal (MSE criterion) piecewise-constant approximation for a signal $\phi(t)$ (under uniform sampling) is $\phi^{PC}(t) = \sum_{i=1}^{N} 1_{\Delta_i}(t)\hat{\phi}_i^{\,*}$, where $\hat{\phi}_i^{\,*} = \frac{1}{|\Delta_i|}\int_{\Delta_i}\phi(t)\,dt = N\int_{\Delta_i}\phi(t)\,dt$.

The minimal MSE we get using the optimal piecewise-constant function is

$$\Psi_{MSE}(\phi \to \phi^{*PC}) = \int_0^1 \phi(t)^2\,dt - \frac{1}{N}\sum_{i=1}^{N}(\hat{\phi}_i^{\,*})^2$$

Comparing this to the optimal MSE we get using the optimal piecewise-linear function:

$$\Psi_{MSE}(\phi \to \phi^{*PL}) = \int_0^1 \phi(t)\,dt - \frac{1}{12N^3}\sum_{i=1}^{N}(a_i^*)^2 - \frac{1}{N}\sum_{i=1}^{N}(c_i^*)^2$$

It holds that $c_i^* = N\int_{\Delta_i}\phi(t)\,dt = \hat{\phi}_i^{\,*}$ and also $(a_i^*)^2$ is a positive number. Therefore, we get

$$\Psi_{MSE}(\phi \to \phi^{*PL}) = \Psi_{MSE}(\phi \to \phi^{*PC}) - \frac{1}{12N^3}\sum_{i=1}^{N}(a_i^*)^2 \leq \Psi_{MSE}(\phi \to \phi^{*PC})$$

So linear approximation has lower optimal MSE.