Task 2:

Preliminary:

We first solve easier problem:

$$0 = y_1 < \cdots < y_N < y_{N+1} = 1 \qquad (N+1) \text{ points}$$

are fixed.

We define $S = \left\{ f \in [0,1]^{[0,1]} \;\middle|\; f \text{ non-decreasing}, \; f([0,1]) = \{y_1, \ldots, y_{N+1}\} \right\}$

We prove that $\underset{f \in S}{\operatorname{argmin}} \int_0^1 (f(x) - x)^2 \, d\mu(x)$ exists.

Proof: (assume that μ has probability density function and thus it is atomless)

we have:

$$\int_0^1 \left(\min_{i \in [\![1, N+1]\!]} (y_i - x) \right)^2 d\mu(x) \leq$$

$$\leq \inf_{f \in S} \int_0^1 (f(x) - x)^2 \, d\mu(x)$$

(as $f([0,1]) = \{y_1, \ldots, y_{N+1}\}$)

Let's define $g(x) = y_{\underset{i \in [\![1, N+1]\!]}{\operatorname{argmin}} |y_i - x|}$

When $x = \dfrac{y_i + y_{i+1}}{2}$ for $i \in [\![1, N]\!]$, we choose any of $\{i, i+1\}$, as the measure is atomless this will not affect the integral. Let's fix the rule and always take i.

We show that $g \in S$.

- $g([0,1]) \subseteq \{y_1, \dots, y_{N+1}\}$

  $g(y_i) = y_i \qquad i \in [\![1, N+1]\!]$

  thus $\quad g([0,1]) = \{y_1, \dots, y_{N+1}\}$

- With the convention above (making g left-continuous)

$$g = \sum_{i \in [\![1, N-1]\!]} y_{i+1} \mathbb{1}_{(t_i, t_{i+1}]} + \mathbb{1}_{(t_N, 1]} \quad \text{for } t_i = \frac{y_i + y_{i+1}}{2}$$

$$i \in [\![1, N]\!]$$

thus g is non-decreasing as so is $\{y_1, \dots, y_{N+1}\}$

This proves that $g \in S$. ■

Now, we introduce $h : [0,1] \to [0,1]$ increasing, smooth.

We prove that $\underset{f \in S}{\operatorname{argmin}} \int_0^1 \left(h(f(x)) - h(x) \right)^2 d\mu(x)$ exists.

Proof:

We have
$$\int_0^1 \left(\min_{i \in [[1, N+1]]} \left(h(y_i) - h(x) \right) \right)^2 d\mu(x) \leq$$

$$\leq \inf_{f \in S} \int_0^1 \left(h(f(x)) - h(x) \right)^2 d\mu(x)$$

(as $h(f([0,1])) = \{ h(y_1), \dots, h(y_{N+1}) \}$

Furthermore, as h is increasing:

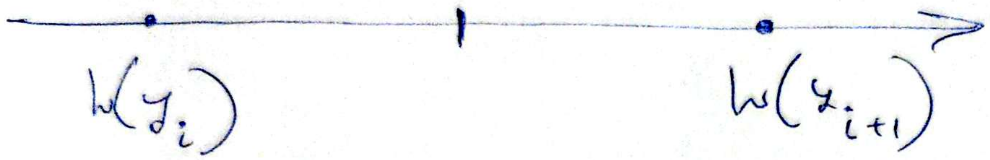
$$0 \leq h(y_1) < \dots < h(y_{N+1}) \leq 1$$

Let's define $g(x) = y_{\underset{i \in [[1, N+1]]}{\operatorname{argmin}} |h(y_i) - h(x)|}$

To prove $g \in S$, we have to first simplify the form of g

$$0 \leq h(y_1) < \dots < h(y_{N+1}) \leq 1$$

$$s_i = \frac{(h(y_i) + h(y_{i+1}))}{2}$$



$i \in [[1, N]]$

We know that h is invertible (it is increasing and continuous) so when

- $x \in [y_i, t_i = h^{-1}(s_i)]$

  $h(x) \in [h(y_i), s_i]$ and thus

  $|h(y_i) - h(x)| \leq |h(y_{i+1}) - h(x)|$.

- $x \in [t_i = h^{-1}(s_i), y_{i+1}]$

  $h(x) \in [s_i, h(y_{i+1})]$ and thus

  $|h(y_{i+1}) - h(x)| \leq |h(y_i) - h(x)|$.

This show that g can be represented as:

$$g = \sum_{i \in [[1, N-1]]} y_{i+1} \mathbb{1}(t_i, t_{i+1}] + \mathbb{1}(t_N, 1]$$

for $t_i = h^{-1}(s_i) = h^{-1}\left(\frac{h(y_i) + h(y_{i+1})}{2}\right)$

$i \in [[1, N]]$

4.

Now, h^{-1} is increasing as h is. Also

$$0 \leq h(y_1) < s_1 < \cdots < s_N < h(y_{N+1}) \leq 1$$

thus

$$0 < t_1 < \cdots < t_N < 1$$

This above, and the form of g shows $g \in S$ ■

Task 2:

6. For each function, once we fix $0<a<b<c<1$. the values $0=y_1<\cdots\cdots<y_{N=2^8}<y_{2^8+1}=1$ will be fixed following the definition of S. We have to also fix the discontinuity points of f.

For each $i\in[\![1,N]\!]$ there is a discontinuity point between y_i, y_{i+1} so we have to also optimize their placement. This is additional $N=2^8$ parameters.

In total $\underline{N+3=2^8+3}$ parameters to optimize.

However, we saw in preliminary results that optimal discontinuity points are $t_i=\frac{y_i+y_{i+1}}{2}$ $i\in[\![1,N]\!]$.

So in facts for the numerical optimizations for ex.7. we only need to optimize $\underline{0<a<b<c<1}$.

7. To find optimal parameters (a, b, c) we run gradient descent. For testing purposes we use different densities μ, and we notice that (a, b, c) parameters move towards higher densities. This is expected, as the approximation error has highest weight there and we need increased "resolution" in such areas.

For sanity check, we make sure numerical results agree with analytical when μ is uniform. $\left(a = \frac{1}{4},\ b = \frac{1}{2},\ c = \frac{3}{4}\right)$

Proof: we introduce $\Delta_1 = a$, $\Delta_2 = b - a$, $\Delta_3 = c - b$, $\Delta_4 = 1 - c$

$$\Delta_i \geq 0 \qquad \sum_i \Delta_i = 1.$$

Now on each segments, integrals between points of discontinuity can be divided as $I_j = \int_{t_j}^{t_{j+1}} \left(\frac{t_{j+1} - t_j}{2} - x \right)^2 dx$ for

$$j \in [\![0, M]\!]$$

$$M = N/4.$$

We see $I_j \propto (t_{j+1} - t_j)^3$ and therefore

$$\mathcal{L}(f) \propto \Delta_1^3 + \Delta_2^3 + \Delta_3^3 + \Delta_4^3. = \sum_i \Delta_i^3$$

Now, $\phi(x) = x^3$ is convex on $[0, 1]$ so by Jensen's inequality:

$$\phi\left(\frac{1}{4} \sum_i \Delta_i\right) \leq \frac{1}{4}\left(\sum_i \phi(\Delta_i)\right) = \frac{1}{4} \sum_i \Delta_i^3$$

Thus $\sum_i \Delta_i^3 \geq 1/16$ and this is attained for $\Delta_i = \frac{1}{4}$.

By definition of Δ_i, $a = \frac{1}{4}$, $b = \frac{1}{2}$, $c = \frac{3}{4}$

7

8. For this problem, in the preliminaries we also proved that optimal placement of discontinuity points are $t_i = g^{-1}\left(\frac{g(y_i) + g(y_{i+1})}{2}\right)$ for $i \in [[1, N]]$.

Once $\{t_i \mid i \in [[1, N]]\}$ are fixed we use gradient descent with pytorch to optimize $0 < a < b < c < 1$ triplet.