CISC 621, Assignment 2, Question 6

Eric Abbott, Elizabeth Fellows, Bowen Hui

October 25, 1999

CLR Problem 10-2, part d.

The post-office location problem is defined as follows. We are given n points $p_1, p_2, ..., p_n$ with associated weights $w_1, w_2, ..., w_n$. We wish to find a point p (not necessarily one of the input points) that minimizes the sum $\sum_{i=1}^{n} w_i d(p, p_i)$, where $d(a, b)$ is the distance between the points a and b.

Argue that the weighted median is a best solution for the 1-dimensional post-office location problem, in which points are simply real numbers and the distance between points a and b is $d(a, b) = |a - b|$.

Let T_p be the solution with p as the post-office.

$$T_p = \sum_{i=1}^{n} w_i d(p, p_i) \tag{1}$$

Solution.

Our claim is that the weighted median p_k is the best solution to the 1-dimensional post-office location problem. To prove this, we use a proof by contradiction assuming that a better solution exists. We try to find such a solution by scanning across the real line. After failing to find a better solution, we will conclude that no such point existst and thereby resulting in contradiction. So the weighted median is a best solution.

The weighted median of $p_1, p_2, ..., p_n$ is p_k where

$$\sum_{p_i < p_k} w_i \leq \frac{1}{2}, \sum_{p_i > p_k} w_i \leq \frac{1}{2}. \tag{2}$$

Let T_{p_k} be the solution for the weighted median p_k as the post-office.

$$T_{p_k} = \sum_{i=1}^{n} w_i d(p_k, p_i) = \sum_{i=1}^{n} w_i |p_k - p_i| = \sum_{p_i < p_k} w_i (p_k - p_i) + \sum_{p_i > p_k} w_i (p_i - p_k) \tag{3}$$

Let T_{p_x} be the solution for the point p_x as the post-office where $p_x = p_k + \epsilon$ and $T_{p_x} < T_{p_k}$. Without loss of generality, we will assume that $\epsilon \geq 0$.

$$T_{p_x} = \sum_{i=1}^{n} w_i d(p_x, p_i) = \sum_{i=1}^{n} w_i |p_x - p_i| \tag{4}$$

Substituting $p_k + \epsilon$ for p_x

$$T_{p_x} = \sum_{i=1}^{n} w_i|(p_k + \epsilon) - p_i| \tag{5}$$

$$T_{p_x} = \sum_{p_i<(p_k+\epsilon)} w_i|(p_k + \epsilon) - p_i| + \sum_{p_i>(p_k+\epsilon)} w_i|(p_k + \epsilon) - p_i| \tag{6}$$

We can break the first part of (6) as follows

$$\sum_{p_i<(p_k+\epsilon)} w_i|(p_k + \epsilon) - p_i| = \sum_{p_i<p_k} w_i|(p_k + \epsilon) - p_i| + \sum_{p_k\leq p_i<(p_k+\epsilon)} w_i|(p_k + \epsilon) - p_i| \tag{7}$$

$$= \sum_{p_i<p_k} w_i(p_k + \epsilon - p_i) + \sum_{p_k\leq p_i<(p_k+\epsilon)} w_i|(p_k + \epsilon) - p_i| \tag{8}$$

$$= \sum_{p_i<p_k} w_i(p_k - p_i) + \sum_{p_i<p_k} w_i\epsilon + \sum_{p_k\leq p_i<(p_k+\epsilon)} w_i|(p_k + \epsilon) - p_i| \tag{9}$$

$$= \sum_{p_i<p_k} w_i(p_k - p_i) + \sum_{p_i<p_k} w_i\epsilon + \sum_{p_k\leq p_i<(p_k+\epsilon)} w_i((p_k + \epsilon) - p_i) \tag{10}$$

$$= \sum_{p_i<p_k} w_i(p_k - p_i) + \sum_{p_i<p_k} w_i\epsilon + (\sum_{p_k<p_i<(p_k+\epsilon)} w_i((p_k + \epsilon) - p_i) + \sum_{p_k=p_i} w_i((p_k + \epsilon) - p_i)) \tag{11}$$

$$= \sum_{p_i<p_k} w_i(p_k - p_i) + \sum_{p_i<p_k} w_i\epsilon + \sum_{p_k<p_i<(p_k+\epsilon)} w_i((p_k + \epsilon) - p_i) + (w_i(p_k + \epsilon - p_k) \tag{12}$$

$$= \sum_{p_i<p_k} w_i(p_k - p_i) + \sum_{p_i<p_k} w_i\epsilon + \sum_{p_k<p_i<(p_k+\epsilon)} w_i((p_k + \epsilon) - p_i) + w_i\epsilon \tag{13}$$

$$= \sum_{p_i<p_k} w_i(p_k - p_i) + (\sum_{p_i<p_k} w_i\epsilon + w_i\epsilon) + \sum_{p_k<p_i<(p_k+\epsilon)} w_i((p_k + \epsilon) - p_i) \tag{14}$$

$$= \sum_{p_i<p_k} w_i(p_k - p_i) + \sum_{p_i\leq p_k} w_i\epsilon + \sum_{p_k<p_i<(p_k+\epsilon)} w_i((p_k + \epsilon) - p_i) \tag{15}$$

And the second part of (6) as follows

$$\sum_{p_i>(p_k+\epsilon)} w_i|(p_k + \epsilon) - p_i| = \sum_{p_i>(p_k+\epsilon)} w_i(p_i - (p_k + \epsilon)) \tag{16}$$

$$= \sum_{p_i > (p_k+\epsilon)} w_i(p_i - p_k) - \sum_{p_i > (p_k+\epsilon)} w_i\epsilon \tag{17}$$

$$= (\sum_{p_i > p_k} w_i(p_i - p_k) - \sum_{p_k < p_i < (p_k+\epsilon)} w_i(p_i - p_k)) - \sum_{p_i > (p_k+\epsilon)} w_i\epsilon \tag{18}$$

$$= \sum_{p_i > p_k} w_i(p_i - p_k) + (\sum_{p_k < p_i < (p_k+\epsilon)} w_i(p_k - p_i)) - \sum_{p_i > (p_k+\epsilon)} w_i\epsilon \tag{19}$$

$$= \sum_{p_i > p_k} w_i(p_i - p_k) + \sum_{p_k < p_i < (p_k+\epsilon)} w_i(p_k - p_i) - (\sum_{p_i > p_k} w_i\epsilon - \sum_{p_k < p_i < (p_k+\epsilon)} w_i\epsilon) \tag{20}$$

$$= \sum_{p_i > p_k} w_i(p_i - p_k) - \sum_{p_i > p_k} w_i\epsilon + (\sum_{p_k < p_i < (p_k+\epsilon)} w_i(p_k - p_i) + \sum_{p_k < p_i < (p_k+\epsilon)} w_i\epsilon) \tag{21}$$

$$= \sum_{p_i > p_k} w_i(p_i - p_k) - \sum_{p_i > p_k} w_i\epsilon + \sum_{p_k < p_i < (p_k+\epsilon)} w_i((p_k + \epsilon) - p_i) \tag{22}$$

re-combining (15) and (22)

$$T_{p_x} = \sum_{p_i < p_k} w_i(p_k - p_i) + \sum_{p_i \le p_k} w_i\epsilon + \sum_{p_k < p_i < (p_k+\epsilon)} w_i((p_k + \epsilon) - p_i) \tag{23}$$

$$+ \sum_{p_i > p_k} w_i(p_i - p_k) - \sum_{p_i > p_k} w_i\epsilon + \sum_{p_k < p_i < (p_k+\epsilon)} w_i((p_k + \epsilon) - p_i) \tag{24}$$

$$= (\sum_{p_i < p_k} w_i(p_k - p_i) + \sum_{p_i > p_k} w_i(p_i - p_k)) + (\sum_{p_i \le p_k} w_i\epsilon - \sum_{p_i > p_k} w_i\epsilon) + 2 \sum_{p_k < p_i < (p_k+\epsilon)} w_i((p_k + \epsilon) - p_i) \tag{25}$$

$$T_{p_x} = T_{p_k} + \epsilon(\sum_{p_i \le p_k} w_i - \sum_{p_i > p_k} w_i) + 2 \sum_{p_k < p_i < (p_k+\epsilon)} w_i((p_k + \epsilon) - p_i) \tag{26}$$

In the equation (26), we find the solution T_{p_x} for p_x equals T_{p_k}, the solution for p_k plus two additional terms,

$$\epsilon(\sum_{p_i \le p_k} w_i - \sum_{p_i > p_k} w_i) \tag{27}$$

and

$$2 \sum_{p_k < p_i < (p_k+\epsilon)} w_i((p_k + \epsilon) - p_i) \tag{28}$$

The term (28) is non-negative because $p_i < (p_k + \epsilon)$.

By the definition of the weighted median, we know that $\sum_{p_i > p_k} w_i \leq \frac{1}{2}$ and that the sum of all the weights equals 1. Therefore $\sum_{p_i \leq p_k} w_i \geq \frac{1}{2}$. Hence $\sum_{p_i > p_k} w_i \leq \frac{1}{2} \leq \sum_{p_i \leq p_k} w_i$.

Therefore the term (27) is non-negative.

Since neither (27) nor (28) can be negative, according to (26), T_{p_x} cannot be less than T_{p_k}.

Since T_{p_x} cannot be less than T_{p_k}, we conclude that there is no point that better minimizes (1), which is a contradiction. Therefore T_{p_k} must be a minimization of (1) which makes the weighted median p_k a best solution.

CLR Problem 10-2, part e. Find the best solution for the 2-dimensional post-office location problem in which the points are (x, y) coordinate pairs and the distance between points $a = (x_1, y_1)$ and $b = (x_1, y_2)$ is the **Manhattan distance** $d(a, b) = |x_1 - x_2| + |y_1 - y_2|$.

Solution.

Our claim is that the weighted median of the x coordinates and the weighted median of the y coordinates will give us the coordinate pair (x, y), a best solution for the 2-dimensional post-office location problem.

The total weighted distance to be minimized is:

$$\sum_{i=1}^{n} w_i d(p, p_i) \tag{29}$$

where $d(a, b) = |x_1 - x_2| + |y_1 - y_2|$. By substitution,

$$\sum_{i=1}^{n} w_i d(p, p_i) = \sum_{i=1}^{n} w_i |x_1 - x_i| + \sum_{i=1}^{n} w_i |y_1 - y_i| \tag{30}$$

For convenience, let X be $\sum_{i=1}^{n} w_i |x_1 - x_i|$, let Y be $\sum_{i=1}^{n} w_i |y_1 - y_i|$, and let Z be $\sum_{i=1}^{n} w_i d(p, p_i)$. So we have $Z = X + Y$, and we want to minimize Z.

In part (d), the equation

$$\sum_{i=1}^{n} w_i d(p, p_i) \tag{31}$$

where $d(a, b) = |a - b|$, is minimized by setting p to be a weighted median. Notice the equation that we minimized in part (d) is identical to X which is also identical to Y. Therefore, by the correctness of (d), X is minimized by finding the weighted median with respect to all the x values. Similarly, by the correctness of (d), Y is minimized by finding the weighted median with respect to all the y values. Since the parts of Z are minimized, the sum itself is also minimized. The solution we get is (x_k, y_k), which may not necessarily be an input point. A final remark is that since part (d) only yields **a** best solution and not **the** best solution, other equally good, but not better, solutions exist for this problem as well.