Locating the Center of a Set of Points on a Curve
Jayadev Misra
6/15/00

1 Problem Description

Given is a finite set of points, A, on a simple closed curve. Henceforth, *point* refers to an arbitrary point on the curve, and a point in A is called an *anchor*. For any two points x, y, the *distance* between them, $d(x, y)$, is the length of the shorter segment (of the curve) joining x, y. The *metric* of a point with respect to the given anchors is the sum of the distances between the anchors and the point, i.e., for a point p, its metric with respect to A, $M(p, A)$, is $\sum_{y \in A} d(p, y)$. It is required to find a point with the smallest metric; we call such a point a *center* of the given set of anchors, and we denote its metric by $M(A)$.

First, we solve the problem when the anchors are on a simple *open* curve; in that case, there is a unique segment of the curve joining any two points. We give a simple characterization of the center in this case. Next, we use the result for open curves to locate the center in a closed curve.

2 Locating the center on an open curve

For an open curve the solution is quite easy: for an odd number of anchors, the middle anchor is the (unique) center; for an even number of anchors any of the middle two anchors, or any point in between, is a center. (Thus, in all cases one of the anchors is a center.) The proof is by induction on the number of anchors.

1. $|A| = 0$: center can be any point because $M(\emptyset) = 0$.

2. $|A| = 1$: center is the unique point in A. Then $M(A) = 0$.

3. $|A| \geq 2$: Let q, r be the two extreme points in A. Let $T = A - \{q, r\}$. We note, without proof, that the center lies in the closed interval $[q, r]$. Below, p is quantified over all points in this closed interval.

$$
\begin{array}{cl}
 & M(A) \\
= & \{\text{definition}\} \\
 & (\min p :: \sum_{y \in A} d(p, y)) \\
= & \{\text{arithmetic; for empty } T, \sum_{y \in T} d(p, y) \text{ is 0, below}\} \\
 & (\min p :: d(p, q) + d(p, r) + \sum_{y \in T} d(p, y)) \\
= & \{p \text{ is between } q, r \text{ on the curve}\} \\
 & (\min p :: d(q, r) + \sum_{y \in T} d(p, y)) \\
= & \{\text{arithmetic}\} \\
 & d(q, r) + (\min p :: \sum_{y \in T} d(p, y)) \\
= & \{\text{definition of center}\} \\
 & d(q, r) + M(T)
\end{array}
$$

From the equation $M(A) = d(q,r) + M(T)$, any center of T that is between the two extreme anchors of A is a center of A. Conversely, a point, p, that is not a center of T will have $M(p,A) \geq d(q,r) + M(p,T) > d(q,r) + M(T) = M(A)$. Therefore, only and all centers of T are centers of A. By the induction hypothesis, center of T is given by: if the number of anchors is odd then the middle point is the center, and for a non-zero even number of anchors any point between the two innermost anchors is a possible center. The middle point of T is the middle point of A in case T has an odd number of anchors, and two innermost anchors of T are the two innermost anchors of A when T has a non-zero even number of anchors.

Calculating the metric of the center We develop the necessary notation and a formula for $M(A)$ that we employ in the solution for the closed curve. Let the anchors in an open curve be successively labelled $0, 1, ..., t$ by going from one extreme point to another. Define a *segment* to be the portion of the curve between two adjacent anchors. Let the length of the segment between anchors i and $i+1$ be s_i, $0 \leq i < t$. The distance between anchor i and the center c, $d(i,c)$, is:

$$(+k : i \leq k < c : s_k), \text{ if } i \leq c$$
$$(+k : c \leq k < i : s_k), \text{ if } i \geq c$$

Now,

$$
\begin{array}{ll}
 & M(A) \\
= & \{\text{definition of metric}\} \\
 & (+i : 0 \leq i \leq t : d(i,c)) \\
= & \{\text{arithmetic}\} \\
 & (+i : 0 \leq i < c : d(i,c)) + (+i : c \leq i \leq t : d(i,c)) \\
= & \{\text{writing the definition of } d(i,c)\} \\
 & (+i : 0 \leq i < c : (+k : i \leq k < c : s_k)) \\
 & \;+(+i : c \leq i \leq t : (+k : c \leq k < i : s_k)) \\
= & \{\text{arithmetic}\} \\
 & (+i : 0 \leq i < c : (i+1)s_i)) + (+i : c \leq i \leq t : (t-i)s_i))
\end{array}
$$

Next, we introduce some notations that make it easier to manipulate the terms in the expression above. For a finite list of reals, B, $B = \langle b_0, b_1, b_2, ...\rangle$, let $\overline{B}$ denote the sum of its elements, i.e., $\overline{B} = b_0 + b_1 + b_2 + ...$, and $\widehat{B} = b_0 + 2 \times b_1 + 3 \times b_2 + ...$, i.e., $\widehat{B}$ is the weighted sum of the elements. Then, $M(A) = \widehat{L} + \widehat{R}$ where L is the list of segment lengths from the extreme left to the center, i.e, $L = \langle s_0, s_1, ..., s_{c-1}\rangle$, and R is the list of segment lengths from the extreme right to the center, i.e., $R = \langle s_{t-1}, s_{t-2}, ..., s_c\rangle$.

We note a few properties of sum and weighted sum. In the following, B is a list, u is a single real number, Bu is the list obtained by appending u

to the end of B, and uB is defined analogously. Let E denote the empty list.

- $\overline{E} = 0$, $\widehat{E} = 0$.
- $\overline{Bu} = \overline{B} + u$, $\overline{uB} = u + \overline{B}$
- $\widehat{Bu} = \widehat{B} + (|B| + 1) \times u$, where $|B|$ is the length of B, and $\widehat{uB} = \overline{uB} + \widehat{B}$

3 Locating the center on a closed curve

Let c be a center on a closed curve; let a be the point exactly half way around the curve from c. Assume, for the moment, that a is not an anchor. Let x be an anchor; the length of the shorter of the two paths from x to c is at most half the length of the curve. Therefore, this path does not include a as an intermediate point, because length of such a path, $d(x, a) + d(a, c)$, exceeds half the length of the curve. Let p, q be the two adjacent anchors that flank a, i.e., a belongs to the closed interval $[p, q]$. This interval is uncovered by any path from an anchor to c because any such path would include a as an intermediate point. Hence, c is the center on the open curve that is obtained from the closed curve after removing the interval $[p, q]$. This shows that one of the anchors is a possible center. A minor modification of this argument can be used to establish this result when point a is an anchor: both paths from a to c are of the same length, and we choose the anti-clockwise path to connect them; then, the segment between a and its clockwise adjacent anchor is uncovered by any path connecting anchors to c.

Our algorithm for center location on a closed curve is as follows. Remove a segment s, locate a center, c_s, of the open curve (using the characterization of the previous section) and compute the metric, M_s, $M_s = M(c_s)$, of this center. Consider all the segments in turn to find the one that results in the smallest value of M_s. Computation of M_s, for any s, takes linear time (in the size of A); therefore, the straightforward calculation of the center takes quadratic time. We show, however, that given M_s it is possible to compute M_t, where t is the segment adjacent to s, in constant time. Therefore, all M_s can be computed in linear time.

Assume, henceforth, that there are at least 3 anchors. In Figure 1, below, s, t are two adjacent segments; c and g are the centers of the open curves when segment s and t are removed, respectively. From the characterization in the previous section, c, g are adjacent. The length of the segment between c, g is u and between g and the next anchor is v. Let X be the sequence of segment lengths starting at the anchor at the left end of s and ending at g; the last element of X is v. Similarly, Y is the list of segment lengths starting after t and ending at the anchor c. Observe that if the number of anchors is odd then $|Y| = |X|$, and if it is even we let $|Y| = |X|+1$. Let P be the total length of the curve; then $\overline{sX}+\overline{tY}+u = P$.

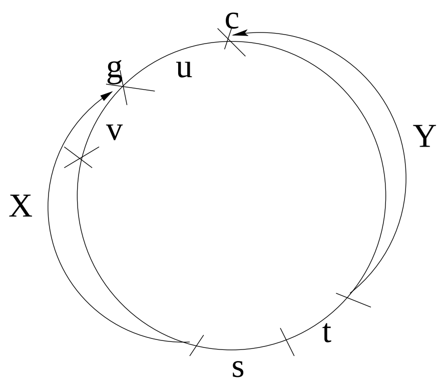

Figure 1: s, t are adjacent segments; c, g corresponding centers

The metric of c, M_s, is $\widehat{Xu} + \widehat{tY}$. Similarly, M_t is $\widehat{sX} + \widehat{Yu}$. Let $\mathit{diff}_s = M_t - M_s$. We calculate diff_s for the case where the number of anchors is odd.

$$\begin{array}{ll} & \mathit{diff}_s \\ = & \{\text{definition}\} \\ & M_t - M_s \\ = & \{\text{rewriting } M_t, M_s\} \\ & \widehat{sX} + \widehat{Yu} - [\widehat{Xu} + \widehat{tY}] \\ = & \{\text{expanding the weighted sums}\} \\ & \overline{sX} + \widehat{X} + \widehat{Y} + (|Y| + 1) \times u \\ & -[\widehat{X} + (|X| + 1) \times u + \overline{tY} + \widehat{Y}] \\ = & \{\text{simplifying}\} \\ & \overline{sX} - \overline{tY} + (|Y| + 1) \times u - (|X| + 1) \times u \\ = & \{|X| = |Y| \text{ since the number of anchors is odd}\} \\ & \overline{sX} - \overline{tY} \\ = & \{\overline{sX} + \overline{tY} + u = P; \text{ Hence, } \overline{sX} = P - \overline{tY} - u\} \\ & P - 2 \times \overline{tY} - u \end{array}$$

A similar analysis shows that in case the number of anchors is even then $\mathit{diff}_s = P - 2 \times \overline{tY}$. Given M_s we can compute the metric, M_t, corresponding to the next segment t, by adding diff_s to M_s. However, the current definition of diff_s still requires linear amount of computation, for the term $\overline{tY}$. Therefore, we compute the second difference, diff'_s, which is defined to be $\mathit{diff}_t - \mathit{diff}_s$. First, we do the analysis for the case where there are an odd number of anchors.

$$
\begin{array}{ll}
 & \mathit{diff}'_s \\
= & \{\text{definition}\} \\
 & \mathit{diff}_t - \mathit{diff}_s \\
= & \{\mathit{diff}_s = P - 2 \times \overline{tY} - u;\ \text{similarly},\ \mathit{diff}_t = P - 2 \times \overline{Yu} - v\} \\
 & P - 2 \times \overline{Yu} - v - [P - 2 \times \overline{tY} - u] \\
= & \{\text{arithmetic}\} \\
 & 2 \times [\overline{tY} - \overline{Yu}] + u - v \\
= & \{\text{expanding } \overline{tY} \text{ and } \overline{Yu}\} \\
 & 2 \times [t + \overline{Y} - \overline{Y} - u] + u - v \\
= & \{\text{simplifying}\} \\
 & 2 \times t - u - v
\end{array}
$$

A similar calculation for even number of anchors shows that $\mathit{diff}'_s = 2 \times t$. Therefore, diff'_s can be computed in constant time in all cases, for any segment s.

The over all calculation strategy is as follows. First, M_s, diff_s and diff'_s are computed for some segment s, which can be done in linear time. These quantities are then computed for the next segment t in constant time: M_t is $M_s + \mathit{diff}_s$, diff_t is $\mathit{diff}_s + \mathit{diff}'_s$, and diff'_t can be computed in constant time. Therefore, M_s, for all s, can be computed in linear time.

Remark If there are exactly 3 anchors in a closed curve then it can be shown that the longest segment should be removed. This strategy does not work for higher number of anchors.

Acknowledgement I am indebted to the Eindhoven Tuesday Afternoon Club (ETAC) for a thorough review, and helpful comments on a first draft of the manuscript.