F. PLASTRIA

Optimal gridpositioning or single facility location on the torus

<http://www.numdam.org/item?id=RO_1991__25_1_19_0>

OPTIMAL GRIDPOSITIONING OR SINGLE FACILITY LOCATION ON THE TORUS (*)

by F. PLASTRIA ([1])

Abstract. — *A finite set of points in the plane must be approximated by gridpoints of an orthogonal grid of fixed orientation and mesh. The problem is to find the position of the grid which minimises the total approximation error. We show how this problem may be viewed as a single facility location problem on the two-dimensional torus, which can be solved by an adapted big square small square method for general errormeasures. Two particular errormeasures — the sum of rectilinear errorlengths and the sum of squared euclidean errorlengths — give rise to very efficient solution methods. In these cases the problem can be decomposed into two independent location problems on a circular network, solvable in linear time after sorting.*

Keywords : Approximation; continuous location problem; circular network.

Résumé. — *Un ensemble fini de points du plan doit être approximé par des points d'un réseau orthogonal d'orientation et maille fixés. Le problème consiste en la recherche de la position du réseau qui minimise l'erreur globale de l'approximation. Nous montrons comment ce problème se traduit en un problème de localisation d'une facilité sur le tore, résoluble pour des mesures d'erreur générales par un algorithme « big square small square » adapté. Pour deux mesures d'erreur particulières — la somme des distances rectilinéaires et la somme des carrés des distances euclidiennes — ce problème peut être résolu très efficacement. Dans ces deux cas le problème peut se décomposer en deux problèmes indépendants de localisation sur un réseau circulaire, résolubles en temps linéaire après un tri.*

Mots clés : Approximation; problème de localisation continue; réseau circulaire.

1. INTRODUCTION

Consider a finite set A of points on the plane. An orthogonal grid G with sides parallel to the axes and fixed mesh must be used to approximate the given set. Each point, $a \in A$ will be approximated by the nearest grid point $a(G)$ of G, yielding the approximation $A(G)$ of A by the grid G. Since the origin of the grid G is not fixed, the different positions of the grid yield

various approximations of A. It is required to position the grid so as to obtain the best approximation of A.

The following three examples show that this problem arises in several practical applications.

When a digitalised image is to be displayed on a low resolution screen or is to be printed by a dot-matrix printer one wants to obtain the most resembling representation, although the pixel density will be much lower than that of the original image. This calls for a good choice of the origin of the coarser grid.

Compression of the numerical representation of spatial data may be accomplished by reducing the coordinates to integer values representable by a small but fixed number of bits, with minimal loss of information concerning relative position of the data. When the absolute position must be stored, it will suffice to add the constant translation and scaling factors to the dataset, enabling the easy reconstruction of the (approximated) original data.

Many automated machines can move in only two directions to position themselves, the movement being discrete with fixed step. The starting point (usually a rest position) must be chosen so as to be able to reach given service points as accurately as possible.

Several measures for the goodness of fit of the approximation $A(G)$ to the dataset A exist. We propose here a large family of such errormeasures: any continuous nondecreasing function of the l_p-distances between a and $a(G)$ (a ranging over A) is allowed. In section 3 it will be indicated how an adapted version of the Big Square Small Square (BSSS) method of Hansen *et al.* (1985) may be used to solve the gridpositioning problem for such general errormeasures. This method is however not very efficient. The two first applications suggested above call for very efficient methods since the datasets concerned are usually quite large and/or response is often needed quickly. For two particular errormeasures, the sum of rectilinear distances and the sum of squared euclidean distances, it is shown in sections 4 and 5 that the gridpositioning problem can be efficiently solved in $O(|A|\log|A|)$ time. These two efficiently solvable cases are very important ones. The first corresponds to the l_1-estimation technique used in statistics to obtain more robust estimation in which the influence of outliers and/or observation errors is less than in the classical "sum of squares" estimation methods, which correspond to the second case we consider. This latter errormeasure however may often be preferred since it penalizes more the greater errors. This will often lead to a subjectively better global approximation.

2. GRIDPOSITIONING AS FACILITY LOCATION ON THE TORUS

A grid G in the plane $\mathbb{R}^2$ with origin $O^G = (g_1, g_2) \in \mathbb{R}^2$ and mesh $(m_1, m_2) \in \mathbb{R}^2$ is the set of gridpoints

$$\{ \overset{N}{G(i_1, i_2)} = (g_1 + i_1 m_1, g_2 + i_2 m_2) \,|\, (i_1, i_2) \in \mathbb{Z}^2 \}.$$

Distances in $\mathbb{R}^2$ will be measured by some l_p-norm with fixed p

$$l_p(x, y) = \begin{cases} \left(\sum_{j=1}^{2} |x_j - y_j|^p \right)^{1/p} & \text{for} \quad 1 \le p < \infty \\ \text{Max}\{ |x_1 - y_1|, |x_2 - y_2| \} & \text{for} \quad p = \infty, \end{cases}$$

where $x = (x_1, x_2)$ and $y = (y_1, y_2)$.

Clearly the gridpoint $a(G)$ of grid G closest to a given point $a \in \mathbb{R}^2$ is then given by $a(G) = G(i_1, i_2)$ with $i_j = \text{round}((a_j - g_j)/m_j)$ $(j = 1, 2)$, where round (r) denotes the integer closest to $r \in \mathbb{R}$ [in the doubtful case where $r = i + 0.5$ (i integer) we take round $(r) = i + 1$]. The error made by approximating a point a by gridpoint $a(G)$ is given by their distance $l_p(a, a(G))$.

If $A \subset \mathbb{R}^2$ is the set of fixed points to be approximated by gridpoints of a grid G of fixed mesh, we will measure the global error made by way of a function of the individual errors $F : \mathbb{R}_+^4 \to \mathbb{R}$, supposed to be nondecreasing and continuous. Good candidates for F would be the sum, sum of squares of Max operators.

The gridpositioning problem may then be formulated as

(GP) $\qquad \text{Min}\{ F((l_p(a, a(G))_{a \in A}) \,|\, O^G \in \mathbb{R}^2 \}$

When a grid G' has origin at a gridpoint of grid G, both consist of the same gridpoints, since we consider the mesh fixed. Therefore both are equivalent for the gridpositioning problem and should be identified. Thus we may restrict our attention to grids with origin $O^G = (g_1, g_2)$ with $0 \le g_j \le m_j$ $(j = 1, 2)$ and the space of these grids is a torus $T = S^1 \times S^1$, where S^1 denotes the topological circle. By rescaling both axes of $\mathbb{R}^2$ so that the mesh is 1×1, we may view this gridspace as the topological quotient group $T = \mathbb{R}^2/\mathbb{Z}^2$ (the torus) with addition modulo 1 as operator. The quotient map $\mathbb{R}^2 \to \mathbb{R}^2/\mathbb{Z}^2$ is the "mod 1" operator, mapping each point of $\mathbb{R}^2$ to the point obtained by replacing each coordinate by its nonnegative fractional part (also denoted by mod 1); e. g. $(1.7, -3.4) \bmod 1 = (1.7 \bmod 1, -3.4 \bmod 1) = (0.7, 0.6)$.

In this way each grid G is defined by its origin $(x_1, x_2) \in T$, where $x_j = (g_j/m_j) \bmod 1$ $(j = 1, 2)$, hence $0 \leq x_j < 1$.

By the same rescaling and quotient map, each point $a \in \mathbb{R}^2$ is mapped to the point $a^T \in T$, where $a_j^T = (a_j/m_j) \bmod 1$ $(j = 1, 2)$.

We denote by d the "shortest path" distance on the circle S^1 (viewed as the interval $[0, 1]$ with endpoints identified), $i.e.$ for $s, t \in S^1$ $(0 \leq s, t < 1)$

$$d(s, t) = \begin{cases} |t - s| & \text{if} \quad |t - s| \leq 1/2 \\ 1 - |t - s| & \text{if} \quad |t - s| > 1/2, \end{cases}$$

or

$$d(s, t) = \text{Min}\left\{ (s - t) \bmod 1, (t - s) \bmod 1 \right\}.$$

The "$l_p -$ distance with mesh (m_1, m_2)" on the torus T defined by

$$l_p^T(x, y) = \begin{cases} \left[\displaystyle\sum_{j=1}^{2} (m_j d(x_j, y_j))^p \right]^{1/p} & \text{for} \quad (1 \leq p < \infty) \\[4mm] \underset{j}{\text{Max}}\, m_j d(x_j, y_j) & \text{for} \quad p = \infty \end{cases}$$

is easily seen to be a metric on T, which is invariant for translation modulo 1, $i.e.$ $l_p^T(x, y) = l_p^T(x +_1 z, y +_1 z)$ for any $z \in T$, where $+_1$ denotes the coordinatewise addition modulo 1 and also the addition modulo 1 on S^1:

$$x +_1 z = (x_1 +_1 z_1, x_2 +_1 z_2) = ((x_1 + z_1) \bmod 1, (x_2 + z_2) \bmod 1)$$

Since for any real value r we have $|r - \text{round}(r)| = d(0, r \bmod 1)$ we find that, for any point $a \in \mathbb{R}^2$ and grid G defined by origin $x = (x_1, x_2) \in T$, $l_p(a, a(G)) = l_p^T(x, a^T)$.

It follows that the gridpositioning problem (GP) may be reformulated as the following single facility location problem on the torus T, with destination set $A^T = \{ a^T \mid a \in A \}$:

(GPT) $\text{Min}\left\{ F((l_p^T(x, a^T))_{a^T \in A^T} \mid x \in T \right\}.$

By continuity of F and the distance measure, and by the compactness of T, an optimal solution is guaranteed to exist. The sequel consists of the description of solution methods for this location problem on the torus. We will further drop the superscript T and consider A as a finite set of points on the torus.

In some particular cases the problem on the torus may be reduced to a problem on the plane; this is possible whenever the set A is sufficiently concentrated, as will be presently shown.

Let us consider a finite set B on the circle S^1 and suppose that its elements b_i $(i=1, \ldots, n)$ are sorted: $0 \leq b_1 \leq b_2 \leq \ldots \leq b_n < 1$. For circularity we denote b_n also b_0. For s, $t \in S^1$ the interval $[s, t]_1$ on S^1 and its length $l(s, t)$ are the set of points of S^1 on the positively ordered path from s to t and the length of it, *i.e.*

If $s \leq t$ then $[s, t]_1 = [s, t]$ and $l(s, t) = t - s$;

If $s > t$ then $[s, t]_1 = [s, 1[\cup [0, t]$ and $l(s, t) = t - s + 1$.
We also have $d(s, t) = \text{Min} \{ l(s, t), l(t, s) \} \leq 1/2$.

LEMMA 1: *If for some* i $l(b_i, b_{i-1}) < 1/2$ *then no point of* $[b_{i-1}, b_i]_1$ *is efficient with respect to the distances* $d(., b_k)$ $(k=1, \ldots, n)$, *i.e. for every* $s \in [b_{i-1}, b_i]_1$ *there exists a* $t \in S^1$ *such that* $d(s, b_k) > d(t, b_k)$ *for all* $k = 1, \ldots, n$.

Proof: Let $m = b_i +_1 (1/2) l(b_i, b_{i-1})$ be the "midpoint" of $[b_i, b_{i-1}]_1$ and $m' = m +_1 1/2$ be its antipode. Denote also by h the midpoint of $[m', m]_1$, then its antipode $h' = h +_1 1/2$ is the midpoint of $[m, m']_1$. Since $l(b_i, b_{i-1}) < 1/2$ we have that $B \subset [h, h']_1 \setminus \{ h, h' \}$.

For any $s \in [h', h]_1 \setminus \{ h, h' \}$ we define its "mirrorpoint" by $t = h +_1 l(s, h)$ (see *fig.* 1). We claim that $d(s, b_k) > d(t, b_k)$ for all $k = 1, \ldots, n$.
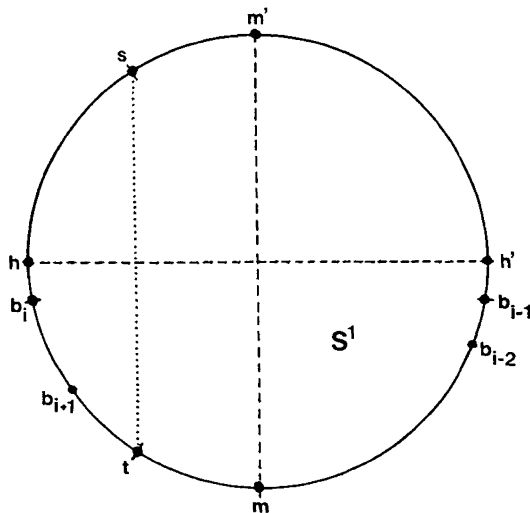

Figure 1. − Illustration of notations in lemma 1.

Clearly $t \in [h, h']_1$, and since $B \subset [h, h']_1$ we have either $b_k \in [b_i, t]_1$, and then $d(t, b_k) = l(b_k, t)$, or $b_k \in [t, b_{i-1}]_1$, and then $d(t, b_k) = l(t, b_k)$, since the other path between b_k and t includes both h and h' and thus has length greater than $1/2$. By symmetry it is sufficient to consider only the first case, where we have

$$
\begin{aligned}
l(s, b_k) &= l(s, h) + l(h, b_k) \\
&= l(h, t) + l(h, b_k) & \text{by definition of } t \\
&= 2\, l(h, b_k) + l(b_k, t) & \text{since } b_k \in [h, t]_1 \\
&> d(t, b_k) & \text{since } b_k \neq h
\end{aligned}
$$

and

$$
\begin{aligned}
l(b_k, s) &= l(b_k, t) + l(t, s) & \text{since } t \in [b_k, s]_1 \\
&> d(t, b_k) & \text{since } s \neq t
\end{aligned}
$$

hence $d(s, b_k) > d(t, b_k)$.

It remains to show the same for the points $s \in [h, h']_1 \setminus [b_i, b_{i-1}]_1$. However the interval $[h, h']_1$ being half the circle S^1 is isometric to an interval on the real line, on which the convex hull property holds [*see* Wendell and Hurter (1973)]. ∎

COROLLARY: *If $A \subset T$ is contained within a "quarter" of T, i.e. if there exist h_1, $h_2 \in S^1$ such that $a_j \in [h_j, h'_j]_1$ ($j = 1, 2$) for each $a \in A$, then the single facility location problem on the torus is equivalent to a planar one.*

Proof: By the preceeding lemma applied to each coordinate and since F is nondecreasing we may restrict search of an optimal solution to the "square' $[h_1, h'_1]_1 \times [h_2, h'_2]_1$ which is isometric to the planar square $[0, 1/2] \times [0, 1/2]$. ∎

Hence in this special case one may use any algorithm for planar single facility location. However in most cases A is not sufficiently concentrated and one needs to take the special topology of the torus into account, which is the subject of the following sections.

3. THE GENERAL CASE

For general l_p-distances and arbitrary nondecreasing and continuous functions F we need a global optimisation method for solving problem (GPT). An excellent candidate is the Big Square Small Square method of Hansen *et al.* (1985). This branch and bound method relies on subdivision of squares

into four congruent subsquares for branching, and the calculation of a lower bound for the objective values within each square. As shown below both subdivision and lower bounds carry over in a natural way to the torus, leading to an algorithm for the gridpositioning problem with a large class of errormeasures.

Initially the torus $T = S^1 \times S^1$ will be covered by four "squares" by choosing any $(h_1, h_2) \in T$ and considering the four set-theoretic products $[m_1, m_1']_1 \times [m_2, m_2']_1$, where m_1 (resp. m_2) is either h_1 or h_1' (resp. h_2 or h_2'); e. g. choosing $h_1 = h_2 = 1/2$ leads to cutting the half open unit square into four squares.

Any generated "square" $[s_1, t_1]_1 \times [s_2, t_2]_1$ will be cut into four sub "squares" in the standard way: let $u_j = s_j +_1 (1/2) l(s_j, t_j)$ be the midpoint of $[s_j, t_j]_1$ ($j = 1, 2$), then the subsquares are found as product of either $[s_j, u_j]_1$ or $[u_j, t_j]_1$ ($j = 1, 2$).

In order to calculate a lower bound on the objective values within the "square" $S = [s_1, t_1]_1 \times [s_2, t_2]_1$ it is sufficient to calculate for each $a \in A$ the distance to S and apply F to these. The distance to S of some $a \in A$ is easily calculated as follows. Denote again by u_j the midpoint of $[s_j, t_j]_1$. Extending the sides of S into full circles, and cutting these at $u_j' (j = 1, 2)$, T is divided into nine regions. For points in each of these regions the closest point of S,
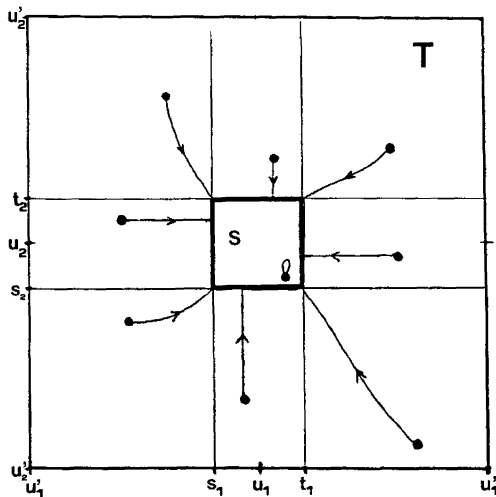

Figure 2. – Projection of the torus T on the "square" S.

and thus the distance up to S is easily found as shown on figure 2, where opposite sides should be identified.

The corresponding calculation of $l_p(a, S)$ may be carried out as follows:

For $j = 1, 2$:

If $a_j \in [u'_j, s_j]_1$ then $w_j = s_j$.

If $a_j \in [s_j, t_j]_1$ then $w_j = a_j$.

If $a_j \in [t_j, u'_j]_1$ then $w_j = t_j$.

Then $l_p(a, S) = l_p(a, w)$ where $w = (w_1, w_2)$.

4. THE RECTILINEAR MINISUM PROBLEM

When the errormeasure is chosen to be the sum of rectilinear (l_1) distances, the gridpositioning problem is much easier. The objective is now

$$F((l_1(x, a))_{a \in A}) = \sum_{a \in A} (m_1 d(x_1, a_1) + m_2 d(x_2, a_2))$$

$$= m_1 \sum_{a \in A} d(x_1, a_1) + m_2 \sum_{a \in A} d(x_2, a_2).$$

Hence (GPT) decomposes into two independent location problems on the circle S^1 of type

(LPC) $\underset{z \in S^1}{\text{Min}} \ f(z) = \sum_{b \in B} w_b d(z, b),$

where $B = \{ a_j : a \in A \}$ and $w_b = |\{ a : a_j = b \}|$ is the "covering number" of $b \in B$ ($j = 1$ or 2 respectively). The mesh m_j is ignored, being a constant factor not affecting the minimisation.

(LPC) is a minisum single facility location problem on the circular network S^1 with node set B. By Hakimi's node optimality theorem (1964) there exists an optimal solution at some node $b \in B$. This trivially leads to an algorithm of $O(|B|^2)$ by evaluating each node in turn, each evaluation requiring $O(|B|)$ time. Goldman (1971) showed that on circular networks the problem is solvable in linear time. Applying this algorithm on both coordinates of the original problem leads to a linear time solution of the rectilinear minisum gridpositioning problem, after sorting along both coordinates in order to obtain an adequate description of the circular networks on which (LPC) is to be solved. Taking this sorting into account we have an $O(|A| \log |A|)$ method, since projection of the original planar point set onto the torus requires $O(|A|)$ time.

5. THE SUM OF SQUARED EUCLIDEAN DISTANCES PROBLEM

Consider now as errormeasure the sum of squares of the euclidean (l_2) distances. The objective is then

$$F((l_2(x,a))_{a \in A})$$

$$= \sum_{a \in A} [(m_1 d_1(x_1,a_1))^2 + (m_2 d_2(x_2,a_2))^2]$$

$$= m_1^2 \sum_{a \in A} d_1(x_1,a_1)^2 + m_2^2 \sum_{a \in A} d_2(x_2,a_2)^2$$

The problem decomposes again into two independent location problems on the circle S^1 of type

(SSE) $$\operatorname*{Min}_{z \in S^2} g(z) = \sum_{b \in B} w_b d(z,b)^2,$$

where B and w_b are defined as in section 4.

We proceed by deriving a linear time algorithm for solving (SSE) when B is sorted increasingly as $0 \leq b_1 \leq b_2 < \ldots < b_n < 1$. We suppose that lemma 1 does not apply, since otherwise the well-known center of gravity result on the line applies.

In order to avoid calculations modulo 1, we work on $[0,2[$ considered as two consecutive copies of S^1, in which x and $x+1$ should be seen as identified. For $z \in [1/2, 3/2[$ (which also is a copy of S^1, translated modulo 1 by $1/2$) we then have for any $b \in S^1$:

$$d(z,b)^2 = \begin{cases} (b-z)^2 & \text{when } z \leq b+1/2 \\ (b+1-z)^2 & \text{when } z \geq b+1/2 \end{cases}$$

We define now the intervals J_k for $k=0, \ldots, n-1$ as $J_k = [b_k + 1/2, b_{k+1} + 1/2]$, where $b_{n+1} \stackrel{N}{=} b_1 + 1$ and $b_0 \stackrel{N}{=} b_n - 1$, and the functions $f_k : \mathbb{R} \to \mathbb{R}$ by

$$f_k(x) = \sum_{m=1}^{k} w_m (b_m + 1 - x)^2 + \sum_{m=k+1}^{n} w_m (b_m - x)^2.$$

Then f_k and f coincide on J_k, showing that f is a piecewise quadratic function with pieces J_k.

The absolute minimum of f_k lies at the "mean" value

$$c_k = \frac{1}{w} \left(\sum_{m=1}^{k} w_m (b_m + 1) + \sum_{m=k+1}^{n} w_m b_m \right) \tag{1}$$

where $w = \sum_{m=1}^{n} w_m$, and one easily derives

$$f_k(x) = f_k(c_k) + w(x - c_k)^2. \tag{2}$$

For the constrained minimum of f_k on J_k three cases arise:

case I_k: $c_k < b_k + 1/2$; f_k reaches its minimum on J_k at $b_k + 1/2$.

case II_k: $b_k + 1/2 \leq c_k \leq b_{k+1} + 1/2$; f_k reaches its minimum on J_k at c_k.

case III_k: $b_{k+1} + 1/2 < c_k$; f_k reaches its minimum on J_k at $b_{k+1} + 1/2$.

Since f and f_k coincide on J_k, the same rules apply to the constrained minimum of f on J_k. Therefore the minimum of f on $[1/2, 3/2[$, is reached either at some antipode $b_k + 1/2$, or at some mean value $c_k \in J_k$.

LEMMA 2: *No antipode $b_k + 1/2$ is a global minimum of f.*

Proof: For $b_k + 1/2$ to be a global minimum of f, it should be the minimum of f_{k-1} on J_{k-1} and of f_k on J_k. Thus cases III_{k-1} and I_k should arise simultaneously, *i.e.* $b_k + 1/2 < c_{k-1}$ and $c_k < b_k + 1/2$, hence $c_k < c_{k-1}$, which is impossible, since from (1) one easily derives $c_k = c_{k-1} + w_k/w$. ∎

It follows that we may restrict our search for the minimum of f to the mean values c_k, the only candidates among which being those satisfying $c_k \in J_k$. In passing this also proves that there exists at least one k for which $c_k \in J_k$!

This search is easily accomplished by an iteration on $k = 0, \ldots, n-1$, using the following recursion formulae:

$$c_k = c_{k-1} + w_k/w$$

$$f_k(c_k) = f_{k-1}(c_{k-1}) + w_k \left[1 + \frac{w_k}{w} + 2(b_k - c_k) \right] \tag{3}$$

This last formula is derived as follows:

$$\begin{aligned} f_k(x) &= f_{k-1}(x) - w_k(b_k - x)^2 + w_k(b_k + 1 - x)^2 \\ &= f_{k-1}(x) + w_k(2(b_k - x) + 1) \end{aligned} \tag{4}$$

By (2) and (3):

$$f_{k-1}(c_k) = f_{k-1}(c_{k-1}) + w(c_{k-1} - c_k)^2$$

$$= f_{k-1}(c_{k-1}) + \frac{w_k^2}{w}.$$

The claimed formula is then obtained from (4) for $x = c_k$. ■

Hence we obtain the following linear time algorithm:

Initialisation step: $k = 0$

Calculate $c = (1/w) \sum_{m=1}^{n} w_m b_m$ and $f = \sum_{m=1}^{n} w_m (b_m - c)^2$.

If $b_n - 1/2 \leq c \leq b_1 + 1/2$ then set $\text{Min} = f$ and $\text{Opt} = c$ else set $\text{Min} = +\infty$.

Iteration steps for $k = 1, \ldots, n-1$

Set $c = c + (w_k/w)$ and $f = f + w_k(1 + (w_k/w) + 2(b_k - c))$.

If $b_k + 1/2 \leq c \leq b_{k+1} + 1/2$ then if $f < \text{Min}$ set $\text{Min} = f$ and $\text{Opt} = c$.

After these n steps the optimal solution is given by Opt mod 1.

Except for the first step, which is $O(n)$, all iteration steps are of constant time, hence the global complexity is $O(n)$, once B is sorted, leading to an overall $O(n \log n)$ complexity. Since the gridpositioning problem with sum of squares of euclidean distances calls for solving (SSE) twice, we have shown it to be solvable in $O(|A| \log |A|)$ time.

REFERENCES

A. J. GOLDMAN, Optimal Center Location in Simple Networks, *Transportation Science*, 1971, *5*, pp. 212-221.

S. L. HAKIMI, Optimal Locations of Switching Centers and the Absolute Centers and Medians of a Graph, *Operations Research*, 1964, *12*, pp. 450-459.

P. HANSEN, D. PEETERS, D. RICHARD and J. F. THISSE, The Minisum and Minimax Location Problems Revisited, *Operations Research*, 1985, *33*, pp. 1251-1265.

R. E. WENDELL and A. P. HURTER, Location Theory-Dominance and Convexity, *Operations Research*, 1973, *21*, pp. 314-320.