Locating a circle on a sphere

Jack Brimberg
Royal Military College of Canada
and Groupe d'Études et de Recherche en Analyse des Décisions
jack.brimberg@rmc.ca

Henrik Juel
Technical University of Denmark
hj@imm.dtu.dk
Informatics and Mathematical Modelling, Richard Petersens Plads,
DTU - Bygning 321, DK-2800 Kongens Lyngby, Denmark
Phone +45 4525 3391, Fax +45 4528 1344

Anita Schöbel
Georg-August-Universität Göttingen
schoebel@math.uni-goettingen.de

Subject classifications:
Facilities/equipment planning, location, continuous:
circle on sphere.

Area of review: Optimization.

Abstract

We consider the problem of locating a spherical circle with respect to existing facilities on a sphere, such that the sum of distances between the circle and the facilities is minimized, or such that the maximum distance is minimized. The problem properties are analyzed, and we give solution procedures. When the circle to be located is restricted to be a great circle, some simplifications are possible.

1 Introduction

The location of a linear facility in space has many potential applications. For example, the facility may represent a new highway in two-dimensional space or a pipeline. It could be an electrical power line, a string of radio or mobile phone transmission towers, or radar stations, and on a smaller scale, a main electrical conduit on a circuit board.

The problem of locating a linear facility in the plane has been well studied beginning with the work of Wesolowsky [18]. Here the objective is to find a line that minimizes the weighted sum of shortest Euclidean distances from the line to a set of existing facilities representing the users or customers. A fundamental property of this problem that leads to an efficient solution procedure is that the "median" line must intersect at least two of the existing points. Further refinements and extensions to the basic model are investigated by Morris and Norback [11, 12] and Norback and Morris [13]; meanwhile Schöbel [16] examines general distance measures and other forms of generalizations to the problem. Finding a line that minimizes the maximum distance to a set of users has been studied in Schömer et al. [17], and in the context of determining the width of a set in Houle and Toussaint [5]. For a recent overview of line location in the plane, see Schöbel [16], and [9] for a survey on dimensional facility location in general.

It is well recognized that the planar model becomes inaccurate when the users are spread over larger areas of the earth's surface, and that spherical distances should be used to account for the earth's curvature. Early work on locating a point facility on a sphere was done by Drezner and Wesolowsky [2] and Katz and Cooper [7], among others, and summarized by Wesolowsky [19]. Ongoing work includes the location of points on the sphere in the presence of forbidden regions, see [3]. As the general single facility minisum problem on a sphere is nonconvex, unlike the planar model, Hansen et al. [4] propose a branch-and-bound algorithm to solve it. The idea is to divide the surface of the sphere into smaller and smaller sections, using alternative bounds provided by the authors to fathom unattractive zones, and to proceed in this fashion until the solution is found within an acceptable accuracy. The problem of locating a point facility on a sphere with the minimax objective is examined more recently by Das et al. [1] and Patel and Chidambaram [14].

The purpose of this paper is to study the problem of locating a circle on a sphere, which is a natural (yet new) extension of the line location problem on the plane. For example, it would be more suitable to model large scale linear facilities on the earth's surface, of the types noted above, as spherical circles or segments thereof, and to use geodesic distances between the facility and its users, thus accounting for the earth's curvature. Geodesic distances are also used in other totally different contexts, such as the analysis of medical or biological data (e.g., see study of rat skulls in Huckemann and Ziezold [6]). The models introduced in

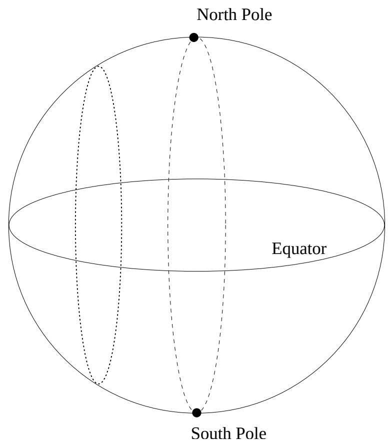

Figure 1: A sphere with north pole, south pole, equator, a great circle (dashed) and a circle (dotted).

this paper may prove useful in these different contexts now or in the future. The models presented here may also be extended in future research to the location of “orbits” in space around the sphere. In the next section, some basic concepts of spherical distances are reviewed and the notation we will use is specified. Section 3 investigates the problem of locating great circles using the minimax criterion, while section 4 deals with locating an arbitrary spherical circle. The main result is that every optimal great circle is at maximum distance from at least three existing facilities, while an arbitrary circle is even fixtured by four existing facilities. The remaining sections are devoted to finding minisum great circles and general circles. Here we will show that an optimal great circle passes through two of the existing facilities, and state some first results for the case of general circles.

2 Notation

We use the following notation, based on [8]. To facilitate the discussion, the reader is also referred to Figure 1. The sphere is denoted by S, and without loss of generality we may assume that the radius of the sphere is 1.

A point $x = (x_1, x_2)$ on the sphere is given by its latitude x_1 (angle from the equator) and its longitude x_2 (angle from the Greenwich meridian). We assume $-\frac{\pi}{2} \leq x_1 \leq \frac{\pi}{2}$, with a negative latitude denoting a point south of the equator, and $-\pi \leq x_2 \leq \pi$, with a negative longitude denoting a point west of Greenwich. Henceforth, by a point we mean a point on the sphere.

A great circle is the intersection between the sphere and a plane through the center of the sphere. The distance between two points is measured along the great circle containing the points; it is the shorter of the lengths of the two great circle arcs connecting the points, measured in radians. The largest possible distance between two points is π, realized when one point is the antipode of the other point. The distance $d(x, a)$ between the two points $x = (x_1, x_2)$ and $a = (a_1, a_2)$ may be computed from the relation

$$\cos d(x, a) = \cos x_1 \cos a_1 \cos(x_2 - a_2) + \sin x_1 \sin a_1.$$

A spherical circle is the intersection between the sphere and a plane. Henceforth, by a circle we mean a spherical circle. A circle is the locus of points with a fixed distance r from a given point $c = (c_1, c_2)$. c is called the center and r is called the radius of the circle. Denoting the circle by $C(c, r)$, we have

$$C(c, r) = \{x \in S : d(x, c) = r\}.$$

If a circle $C(c, r)$ has radius $r > \frac{\pi}{2}$, it may be viewed as a circle with center in the antipode of c and radius $\pi - r$. Thus it suffices to consider circles with radii in the interval $0 \leq r \leq \frac{\pi}{2}$, and henceforth we shall do so. A circle with radius 0 is a point, and a circle with radius $\frac{\pi}{2}$ is a great circle.

The distance between a point a and a circle $C = C(c, r)$, defined as

$$D(C, a) = \min_{x \in C} d(x, a)$$

can be calculated as follows: Consider the great circle containing a and c. If this great circle intersects the circle C in the points x and z, the point to circle distance $D(C, a)$ is given by

$$D(C, a) = \min\{d(x, a), d(z, a)\} =: d(y, a).$$

The closer of the two points x and z is called the *footpoint* y of a with respect to C. Furthermore, let P be the shorter part of the great circle connecting a and y. Note that the length of P equals $D(C, a)$.

For the special case $r = 0$, we have $D(C, a) = d(c, a)$, and for the special case $c = a$, we have $D(C, a) = r$.

It is particularly easy to compute the point to circle distance when the center of the circle is a Pole of the sphere. Suppose for instance that the center of a circle C is the North Pole; then the distance from the point $a = (a_1, a_2)$ to the circle is $r + a_1 - \frac{\pi}{2}$ if a is north of the circle, and $\frac{\pi}{2} - r - a_1$ otherwise, or $D(C, a) = |r + a_1 - \frac{\pi}{2}|$.

In general, we have

$$D(C, a) = |r - d(c, a)|.$$

For the case when C is a great circle, there is a simple relation between the point to great circle distance, $D(C, a)$, and the smallest Euclidean distance from the point to the plane H containing the great circle, $E(H, a)$, namely $\sin D(C, a) = E(H, a)$.

Let n be the number of existing facilities, located at $a_j = (a_{j1}, a_{j2}) \in S$ with positive weight w_j, for $j = 1, \ldots, n$. Denote the set of existing facility locations by A.

Then our optimization problem is finding a circle $C = C(c, r)$ with center $c = (c_1, c_2)$ and radius $r \in [0, \frac{\pi}{2}]$ so as to minimize

$$f(C) = f(c, r) = \sum_{j=1}^{n} w_j D(C(c, r), a_j)$$

or so as to minimize

$$g(C) = g(c, r) = \max_{j=1,\ldots,n} w_j D(C(c, r), a_j).$$

The first objective refers to the minisum or median problem, whereas the second objective refers to the minimax or center problem. We will mainly consider the unweighted case, i.e., the case in which all weights w_j are equal, and hence, may be set to unity without affecting the optimal solution.

Any given circle $C = C(c, r)$ separates the sphere in two parts, and it is convenient to define the index sets $J_+ = \{j : d(a_j, c) < r\}$, $J_- = \{j : d(a_j, c) > r\}$, and $J_= = \{j : d(a_j, c) = r\}$.

3 Finding minimax great circles

We consider the unit weight (or unweighted) *Great-Circle-Minimax problem* of locating a great circle on the sphere that minimizes the maximum distance to the existing facilities. This problem will be called (GCM). To avoid the trivial case, let us assume that $n \geq 3$.

Lemma 1 *Let C^* be an optimal solution of (GCM) with objective value $g(C^*)$. Then there are at least three existing facility locations $a \in A$ satisfying*

$$D(C^*, a) = g(C^*).$$

Proof: Let C^* be an optimal great circle and assume first that there exist exactly two points $a_i, a_j \in A$ with $g(C^*) = D(C^*, a_i) = D(C^*, a_j)$. Without loss of generality let $i = 1$, $j = 2$, and assume that $g(C^*) > 0$; otherwise all points $a \in A$ satisfy $D(C^*, a) = g(C^*)$.

Determine the corresponding footpoints and great circle segments $y_1 \in P_1$ and $y_2 \in P_2$. Let $\epsilon > 0$ and define two points $y_1' \in P_1$ and $y_2' \in P_2$ such that

$$
\begin{aligned}
d(a_1, y_1) - d(a_1, y_1') &= \epsilon, \\
d(a_2, y_1) - d(a_2, y_1') &= \epsilon.
\end{aligned}
$$

Since a great circle is uniquely defined by two points we define C' as the great circle passing through y_1' and y_2'. Note that the function $\mathcal{C}$ mapping two points y_1, y_2 to the great circle defined by these points is well-defined and continuous whenever $y_1 \neq y_2$ and y_1 and y_2 are not antipodes to each other. For this reason we distinguish the following three cases.

- First assume that $y_1 \neq y_2$ and the points are not antipodes to each other. Hence, $y_1' \neq y_2'$.

  We obtain for $k = 1, 2$:

  $$D(C', a_k) = \min_{x \in C'} d(x, a_k) \leq d(y_i', a_k) < d(y_i, a_k) = g(C^*)$$

  Denote $g' = \max\{D(C', a_1), D(C', a_2)\}$ and note that $g' < g(C^*)$. Since $g(C^*) > D(C^*, a)$, for all other $a \in A \setminus \{a_1, a_2\}$, the continuity of $\mathcal{C}$ yields $g' \geq D(C', a)$, if ϵ is chosen small enough.

  Together,

  $$g(C') = \max_{a \in A} D(C', a) = g' < g(C^*)$$

  contradicting the optimality of C^*.

- In the case that $y_1 = y_2$, the existing facility locations a_1 and a_2 must be on opposite sides of C^* (otherwise they would coincide). Then rotate C^* by a small amount as in the previous case, but this time about the axis through the common footpoint, $y_1 = y_2$. Again we obtain a reduction in distance, $d(a_1, y_1') = d(a_2, y_2') < g(C^*)$, leading to a similar contradiction as before.

- If y_1 and y_2 are antipodes of each other, there are two possibilities: either a_1 and a_2 are on the same side of C^*, in which case rotate C^* a small amount about the axis through y_1 and y_2; or a_1 and a_2 are on the opposite sides of C^*, in which case use the line on C^* perpendicular to (y_1, y_2) as the axis of rotation. Again we obtain a similar contradiction as before.

To exclude that there exists only one unique point a_j on C^* satisfying $g(C^*) = D(C^*, a_j)$, we proceed as follows. Let a_j be such a unique point, y be the corresponding footpoint, and P be the great circle segment between a_j and y. Similar to the first part of the proof, we fix an arbitrary point $x \in C^* \setminus \{y\}$, find $y' \in P$ such that

$$d(a_j, y) - d(a_j, y') = \epsilon$$

for some $\epsilon > 0$, and choose a new circle C' as the great circle passing through x and y'. Since $g(C') < g(C^*)$, we again have a contradiction.

QED

Now we turn our attention to computing an optimal great circle.

First, we remark that it can happen that all three existing facilities with the maximum distance to the circle may lie on the same side of the circle, as the following example demonstrates.

Consider three existing facilities all on the northern hemisphere, but all three of them close to the equator, e.g.,

$$A = \{(\epsilon, 0), (\epsilon, \frac{2}{3}\pi), (\epsilon, -\frac{2}{3}\pi)\}.$$

Using Lemma 1 and checking all great circles at equal distance to the three existing facilities yields the equator $C((\frac{\pi}{2}, 0), \frac{\pi}{2})$ with distance of ϵ to all three points as the optimal great circle.

From Lemma 1 we know that all optimal circles of (GCM) have the same positive distance to at least three points $a_j, a_i, a_k \in A$. Note that in the case that not all points are contained in one common great circle, no pair of these points a_j, a_i, and a_k can be antipodes to each other, since they all have the same positive distance to an optimal great circle. Furthermore, at least two of these points lie on the same side of C; without loss of generality let us assume that $i, k \in J_+$. Since $D(C, a) = |r - d(a, c)|$ and $i, k \in J_+$, we obtain $d(c, a_i) = \frac{\pi}{2} - D(C, a_i) = \frac{\pi}{2} - D(C, a_k) = d(c, a_k)$, i.e., the distance from both points a_i and a_k to the center c of the circle is the same. In other words, c lies in the set $B_{ik} = \{x \in S : d(x, a_i) = d(x, a_k)\}$, which is the bisector of a_i and a_k. Note that bisectors on the sphere are great circles. Hence, to find the center point c^* of an optimal circle C^*, only points on the bisectors $B_{ik}, i \neq k, i, k \in \{1, \ldots, n\}$ need to be investigated. Finding the best great circle with center c on some bisector (great circle) B_{ik} hence reduces to a one-dimensional optimization problem.

To tackle this problem we can furthermore use that the distance of an optimal great circle $D(C^*, a_j)$ to a third point a_j is the same as to the points a_i and a_k defining the bisector B_{ik}.

We hence only have to investigate the points a_j satisfying that $D(C, a_j) = D(C, a_i)$ which can be reformulated as $|\frac{\pi}{2} - d(c, a_j)| = |\frac{\pi}{2} - d(c, a_i)|$.

We need to consider two cases.

Case 1: Assume that a_j is on the same side of the optimal circle C^* as a_i and a_k. In this case, a_j satisfies

$$d(c, a_j) = \frac{\pi}{2} - D(C^*, a_j) = \frac{\pi}{2} - D(C^*, a_i) = d(c, a_i),$$

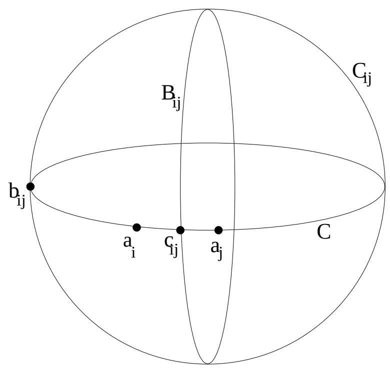

Figure 2: Construction of $B_{ij} = C(b_{ij}, \frac{\pi}{2})$ and $C_{ij} = C(c_{ij}, \frac{\pi}{2})$.

hence the set of candidates to be investigated can be determined by intersecting B_{ik} with all bisectors B_{ij}, $j = 1, \dots, n, j \neq i, j \neq k$.

Case 2: Assume that a_j lies on the opposite side of C^* as a_i and a_k do. Consequently, a_j satisfies $\frac{\pi}{2} - d(c, a_j) = -\frac{\pi}{2} + d(c, a_i)$, or, equivalently,

$$\pi = d(c, a_i) + d(c, a_j).$$

As candidates for the optimal center c we hence have to consider all points $c \in C_{ij} = \{x \in S : d(x, a_i) + d(x, a_j) = \pi\}$

For Case 1 it is well known that B_{ij} is a great circle. In the following we show that also C_{ij} is a great circle and how the centers of B_{ij} and C_{ij} can be constructed. To this end, let C denote the (unique) great circle passing through a_i and a_j. Let c_{ij} be the midpoint on the great circle segment of C joining a_i and a_j and choose $b_{ij} \in C$ such that $d(b_{ij}, c_{ij}) = \frac{\pi}{2}$, see Figure 2. Note that by construction we have $d(c_{ij}, a_i) = d(c_{ij}, a_j)$ and $d(b_{ij}, a_i) + d(b_{ij}, a_j) = \pi$. Then the following holds.

Lemma 2

1. $B_{ij} = C(b_{ij}, \frac{\pi}{2})$

2. $C_{ij} = C(c_{ij}, \frac{\pi}{2})$

Proof:

1. Consider any point $x \in S$ such that $d(x, a_i) = d(x, a_j)$. Since $d(c_{ij}, a_i) = d(c_{ij}, a_j)$ it follows by congruence that the great circle segment joining x and c_{ij} belongs to B_{ij}, or $d(b_{ij}, x) = \frac{\pi}{2}$. The reverse obviously holds: If $x \in B_{ij}$, then $d(x, a_i) = d(x, a_j)$.

2. Suppose $x \in S$ satisfies $d(x, a_i) + d(x, a_j) = \pi$, and without loss of generality, $d(x, a_i) \leq d(x, a_j)$. Using center x, draw three circles C_1, C_2, C_3 of radius $d(x, a_i), \pi/2, d(x, a_j)$, respectively. Clearly, C_2 bisects the angle $(= d(x, a_j) - d(x, a_i))$ subtended by C_1 and C_3 and the center of the sphere, and by symmetry, C_2 intersects the great circle segment joining a_i and a_j at its mid-point c_{ij}. We conclude that $x \in C_{ij}$. The argument also applies in reverse such that from $x \in C_{ij}$ we conclude that $d(x, a_i) + d(x, a_j) = \pi$.

QED

This means, in both cases we have to find the intersection of two great circles. This gives two points $x_1, x_2 \in S$. But note that x_1 and x_2 are antipodes to each other and hence define the same great circle. This means that only one of these points needs to be further investigated.

Algorithm 1 for (GCM)

Step 1: Let the candidate set $K = \emptyset$
Step 2: For all triples $(i, j, k) \in \{1, \ldots, n\}$, determine an intersection point h^1_{ijk} of the bisectors B_{ik} and B_{ij}, and an intersection point h^2_{ijk} of the bisector B_{ik} and the great circle C_{ij}; let $K = K \cup \{h^1_{ijk}, h^2_{ijk}\}$
Step 3: Evaluate all candidates $c \in K$ by calculating
$g(C(c, \frac{\pi}{2})) = \max_{j=1,\ldots,n} |d(a_j, c) - \frac{\pi}{2}|$ and take the one with the smallest objective value.

Since there are $O(n^3)$ triplets $\{a_i, a_j, a_k\}$ to be evaluated (or $O(n^3)$ candidate solutions in general), and $O(n)$ operations are required to verify each candidate solution, the resulting time complexity is $O(n^4)$. However, some simplifications of this algorithm are possible. For example, if the spherical triangle formed by the triplet $\{a_i, a_j, a_k\}$ contains a fourth existing facility, then $\{a_i, a_j, a_k\}$ cannot determine an optimal circle of the type given by case 1. We can therefore delete the candidate h^1_{ijk}. If the (spherical) convex hull of a subset of the fixed points contains all the points of the triplet within its interior, then case 2 does not apply, and we can delete the three intersection points $(h^2_{ijk}, h^2_{ikj}, h^2_{jik})$ derived for this triplet. Thus, some pre-processing steps may reduce the number of candidates considerably.

We add a short remark on the weighted great circle minimax problem. In this case we allow general positive weights for the existing facilities and minimize the maximum weighted distance to the circle. Lemma 1 then states that there must be three existing facilities i, j, k satisfying

$$w_i D(C^*, a_i) = w_j D(C^*, a_j) = w_k D(C^*, a_k) = g(C^*),$$

which may be shown by an analogous proof. Unfortunately, the sets B_{ij} and C_{ij} are no longer great circles in the weighted case, such that Algorithm 1 cannot be applied as easily. Finding h^1_{ijk} or h^2_{ijk} is equivalent to solving a system of two nonlinear equations in two unknowns ($x = (x_1, x_2)$), which appears in the weighted problem to require a numerical procedure.

Another problem, intimately related to (GCM), is that of locating a plane H through the center of the sphere, such that the maximum of the Euclidean distances between the existing facilities and the plane is minimized. This problem is called the restricted Euclidean minimax problem (REM), because the plane cannot be located anywhere in $\mathbb{R}^3$, but is restricted to contain the center of the sphere; henceforth the center of the sphere will be called the origin, O. The Euclidean distance between a point a and a plane H is found as follows: Consider a line through a, perpendicular to H, and let the line intersect H in b; the Euclidean point to plane distance is the Euclidean length of the line segment between a and b. Let us denote this distance by $E(H, a)$. (REM) may then be written as the problem of minimizing

$$G(H) = \max_{j=1,\ldots,n} E(H, a_j)$$

Consider an arbitrary great circle C and the plane containing it, H. For all existing facilities, we have $E(H, a_j) = \sin(D(C, a_j))$. Since $\sin(v)$ is increasing on the relevant interval, $0 \le v \le \frac{\pi}{2}$, the existing facility, j', that is furthest from H (measured by Euclidean distance) is also furthest from C (measured by angle). This observation means that $E(H, a_{j'}) = \max_{j=1,\ldots,n} E(H, a_j)$ and $D(C, a_{j'}) = \max_{j=1,\ldots,n} D(C, a_j)$, and allows us to characterize the relationship between the two problems.

Lemma 3 *The problems (GCM) and (REM) are equivalent: If a plane H^* solves (REM), then the great circle contained in H^* solves (GCM), and if a great circle C^* solves (GCM), then the plane containing C^* solves (REM).*

Proof: Trivial.

QED

The equivalence also holds for the weighted case. This gives an alternative proof of Lemma 1 by using Theorem 3 of [15] which states that all optimal hyperplanes of (REM), i.e., all hyperplanes in $\mathbb{R}^3$ through one specified point that minimize the maximum distance to a given set of points $a_1, \ldots, a_n$ are equidistant to at least three affinely independent points of this set.

The algorithmic implication is clear: To solve the great circle minimax problem, we can equivalently solve the restricted Euclidean minimax problem. If all candidate hyperplanes are investigated, the complexity of this approach remains the same as the complexity of Algorithm 1.

4 Finding minimax circles

Now we consider the unweighted *Circle-Minimax problem* (CM), where we relax the restriction to great circles, in order to allow any circle on the sphere. Again our goal is to minimize the maximum distance to the existing facilities. Although many more circles are allowed as feasible solutions the problem is more straightforward to solve. First of all, we prove the following result which is stronger than the result of Lemma 1. Here we need to assume that $n > 3$, and the existing facilities are not all contained in the same circle.

Theorem 1 *Let C^* be an optimal solution of (CM) with objective value $g(C^*)$. Then there are at least four existing facility locations $a \in A$ satisfying*

$$D(C^*, a) = g(C^*).$$

Furthermore, at least two of these facilities must be located inside C^* (i.e., belong to J_+), and at least two of them must be located outside C^* (i.e., belong to J_-).

Proof: Let C^* be an optimal circle which is at maximum distance from exactly $m \in \{1, 2, 3\}$ existing facilities. Without loss of generality assume that these existing facility locations are $a_1, \ldots, a_m$; thus,

$$\begin{aligned} g(C^*) &= D(C^*, a_1) > 0 \\ &\cdots \\ g(C^*) &= D(C^*, a_m) > 0 \end{aligned}$$

The goal is to define a circle C' with a smaller objective value. This is done as follows: For $j = 1, \ldots, m$, consider the footpoint y_j of a_j with respect to C^*, and the great circle segment P_j between y_j and a_j. Furthermore, let us assume that no two footpoints coincide. (Otherwise, use the chord through the common footpoint y_j and the center of the disc formed by C^* as an axis of rotation similarly as in Lemma 1). Choose $\epsilon > 0$, and for $j = 1, \ldots, m$, define a new point y'_j by moving y_j along P_j a distance ϵ closer to a_j, i.e., $y'_j \in P_j$ and $d(y_j, a_j) - d(y'_j, a_j) = \epsilon$. Furthermore, choose $3 - m$ arbitrary points in $C^* \setminus \{y_1, \ldots, y_m\}$. This defines $m + 3 - m = 3$ points which uniquely define a new circle C', and if the footpoints are different, the function mapping these 3 points to a circle C' is well-defined and continuous. Hence, we can choose $\epsilon > 0$ in such a way that

$$|D(C^*, a_j) - D(C', a_j)| \leq \delta \text{ for all } j = 1, \ldots, n.$$

To calculate the objective value of C' we first consider $j = 1, \ldots, m$, and obtain

$$\begin{aligned} D(C', a_j) &= \min_{x \in C'} d(x, a_j) \\ &\leq d(y'_j, a_j) \\ &= d(y_j, a_j) - \epsilon \\ &= g(C^*) - \epsilon. \end{aligned}$$

Defining $g' := \max_{j=1,\ldots,m} D(C', a_j)$, this yields

$$0 < \epsilon \leq g(C^*) - g' \leq \delta.$$

On the other hand, for all $j = m+1, \ldots, n$, we know that

$$D(C^*, a_j) < g(C^*);$$

hence, choosing $\delta \leq \frac{1}{2}\left(g(C^*) - \max_{j=m+1,\ldots,n} D(C^*, a_j)\right)$ implies that

$$D(C', a_j) \leq D(C^*, a_j) + \delta \leq g(C^*) - \delta \leq g'.$$

Thus, $D(C', a_j) \leq g'$ for all $j = 1, \ldots, n$, and hence, $g(C') = g' < g(C^*)$. We conclude that $m \geq 4$.

Now suppose that only one extreme point is located inside an optimal circle $C^* = C(c^*, r^*)$, while the remaining three or more extreme points are located outside. Without loss of generality, let a_1 denote the internal extreme point, and let y_1 be its footpoint on C^*. Consider a new circle C' with radius $r' = r^* + \epsilon$, and center c' displaced a distance ϵ from c^* along the great circle through y_1 and c^*, and in the direction away from y_1. Note that C' is now closer to all the external extreme points, except possibly one with common footpoint y_1, and $g(C') = g(C^*)$ for small enough positive ϵ. But this contradicts the requirement that $m \geq 4$, and we conclude that C^* cannot be an optimal circle. Similarly, if a_1 is the only extreme point outside C^*, construct C' with radius $r' = r^* - \epsilon$, and center c' displaced in the same manner as before, except distance ϵ towards y_1, to arrive at the same conclusion. Thus, we finally conclude that at least two extreme points are located on each side of C^*.

QED

It should be noted that Theorem 1 gives an analogous result as obtained by Drezner et al. [10] for the (unweighted) minimax circle in the plane. The following rather simple procedure can now be used to determine an optimal circle for the unweighted (CM). The algorithm is based on the fact that the center of the optimal circle must, as a result of Theorem 1, be an intersection point of two bisector circles.

In the unlikely event that two bisectors B_{ij} and B_{kl} are identical great circles, any point c on the common bisector becomes a candidate. In this case, we would move c along the common bisector in order to reduce the objective value until a new extreme point is added. Thus, the case where bisectors B_{ij} and B_{kl} coincide may be ignored. To simplify the procedure below, we assume that B_{ij} and B_{kl} are always different circles. Hence, they intersect at exactly two points that are antipodes of each other.

Algorithm 2 for (CM)

Step 1: Let $K = \emptyset$.
Step 2: For all pairs of pairs $\{i, j\} \subseteq \{1, \ldots, n\}$ and $\{k, l\} \subseteq \{1, \ldots, n\} \setminus \{i, j\}$, determine an intersection point h_{ijkl} of the bisectors B_{ij} and B_{kl}, and let $K = K \cup \{h_{ijkl}\}$.
Step 3: Determine the radius r_{ijkl} for each candidate in K as follows:
$r_{ijkl} = (d(h_{ijkl}, a_i) + d(h_{ijkl}, a_k))/2$.
Step 4: Evaluate all candidates, $c \in K$, by calculating
$g(C(c, r)) = \max_{j=1,\ldots,n} |d(a_j, c) - r|$,
where r is the corresponding radius obtained in step 3,
and take the one with the smallest objective value.

Note that the time complexity of this procedure is $O(n^5)$.

Again, a similar proof can be given for the weighted circle problem, in which positive weights for the existing facilities are allowed. In this case, Theorem 1 can be extended to four existing facilities at the same weighted distance to the optimal circle. Unfortunately that does not help much for finding the optimal circle in the weighted case, since in the case that $w_i \neq w_j$ the set $\{x \in S : \exists r \text{ such that } w_j(d(x, a_j) - r) = w_i(d(x, a_i) - r)\}$ contains all points x for which $\frac{w_j d(x,a_j) - w_i d(x,a_i)}{w_j - w_i}$ is positive.

For any two pairs $\{i, j\}$, $\{k, l\}$ ($\{i, j\} \cap \{k, l\} = \emptyset$), the resulting subproblem requires solving three nonlinear equations in three unknowns ($(x, r) = (x_1, x_2, r)$), namely,
$w_j d(x, a_j) - w_i d(x, a_i) = (w_j - w_i)r$,
$w_l d(x, a_l) - w_k d(x, a_k) = (w_l - w_k)r$,
$r = (w_i d(x, a_i) + w_k d(x, a_k))/(w_i + w_k)$,
which again appears to require a numerical procedure.

5 Finding minisum great circles

In this section we consider the problem of finding a great circle minimizing the sum of (weighted) distances to the given facilities. This problem will be called (GCS).

In a first result, we relate the minisum great circle problem to that of locating a plane H through the center of the sphere, such that the sum of the Euclidean distances to the points $a_1, \ldots, a_n$ is minimized. We denote this problem as a restricted Euclidean minisum problem (RES). It can be stated as the problem of minimizing

$$F(H) = \sum_{j=1,\ldots,n} w_j E(H, a_j).$$

Recall that the Euclidean distance $E(H, a_j) = \sin(D(C, a_j))$, if H is the hyperplane containing the great circle C. Unfortunately, we cannot show that (RES) and (GCS) are equivalent as in the minimax case, but we can at least use the hyperplane location problem for getting an upper and lower bound.

Lemma 4 *Let H^* be an optimal hyperplane for (RES), and let C^* be an optimal great circle for (GCS). Furthermore, let $C(H^*) = H^* \cap S$ be the great circle contained in H^* and $H(C^*)$ be the hyperplane passing through C^*. Then*

$$F(H^*) \leq F(H(C^*)) \leq f(C^*) \leq f(C(H^*)).$$

Proof:

$$\begin{array}{rcl} F(H^*) & \leq & F(H(C^*)) \\ & = & \sum\limits_{j=1,\dots,n} w_j E(H(C^*), a_j) \\ & = & \sum\limits_{j=1,\dots,n} w_j \sin D(C^*, a_j) \\ & \leq & \sum\limits_{j=1,\dots,n} w_j D(C^*, a_j) \\ & = & f(C^*) \\ & \leq & f(C(H^*)) \end{array}$$

QED

Note that for (RES) it is known that all optimal hyperplanes pass through at least two of the existing facilities (see Theorem 3 of [15]); i.e., $C(H^*)$ always contains two points a_i, a_j.

Locating a great circle may be viewed as the spherical equivalent of locating a straight line on the plane; in fact, the latter problem may be viewed as a special case where the existing facility locations remain a finite distance apart while the radius of the sphere increases without bound. The following result thus generalizes a well-known property of the optimal line in the plane (e.g., see [16]).

Lemma 5 *An optimal solution C^* of (GCS) may be found that intersects at least two of the existing facility locations.*

Proof: Consider first the trivial case where all existing facilities are located on some great circle, C. Obviously, $C^* = C$ is the optimal solution, with $f(C^*) = 0$.

Now assume that all existing facilities are not located on the same great circle. The problem is to minimize
$f(C) = f(c, \frac{\pi}{2}) = \sum_{j \in J_+} w_j(\frac{\pi}{2} - d(c, a_j)) + \sum_{j \in J_-} w_j(\frac{\pi}{2} - d(c', a_j))$,

where the index sets J_+ and J_- contain the existing facilities on each side of the great circle, and c' is the antipode of c. Suppose we have an optimal solution $C^* = (c^*, \frac{\pi}{2})$ that does not contain any existing facilities. It is known that the distance from a given point a to another point on the sphere is a convex function within a circle of radius $\frac{\pi}{2}$ and center a (e.g., see [4]). Hence $d(c, a_j)$ is convex in a local neighborhood of c^* for $j \in J_+$, and $d(c', a_j)$ is convex in a local neighborhood of $c^{*\prime}$ for $j \in J_-$. Furthermore, since $c' = c + (\pi, 0)$, the convexity extends to c for $j \in J_-$. We conclude that $f(C)$ is locally concave at c^* in a strict sense, and hence, C^* must intersect one of the existing facility locations, say a_r. Now use the line through a_r and the center of the sphere as the axis of rotation, to conclude in similar fashion (for adjusted J_+, J_-) that $f(C)$ is locally concave, and C^* may be rotated further without increasing the objective function until it intersects a second fixed point.

QED

This result permits a finite solution method for (GCS): Compute the objective function value for the great circle through each pair of existing facilities (ignoring any pairs that are antipodes of each other); the optimal solution is the great circle with lowest value. The time complexity of this algorithm is $O(n^3)$.

6 Finding minisum circles

The problem we discuss here, denoted by (CS), is to find a circle minimizing the sum of (weighted) distances to the existing facilities.

Recall that for this purpose we identify a circle $C = C(c, r) \subset S$ by its center point $c \in S$ and its radius r.

The objective function of (CS) may be written as

$f(c, r) = \sum_{j=1}^{n} w_j D(C, a_j) = \sum_{j=1}^{n} w_j |r - d(c, a_j)|$.

The first observation is the following.

Lemma 6 *There exists an optimal solution C^* to problem (CS) passing through at least one of the existing facilities.*

Proof: Let $C = C(c, r)$ be any optimal solution of (CS). Now fix the center c and consider the problem of finding the optimal radius r^*, i.e.,

$\min_r \sum_{j=1}^{n} w_j |r - d(c, a_j)|$.

This problem is equivalent to the problem of locating a point on a line for which it is well known that an optimal solution r^* exists satisfying $r^* = d(c, a_{j^*})$ for some $j^* \in \{1, \ldots, n\}$ (e.g., see [8]). Consequently, $C^* = C(c, r^*)$ contains a_{j^*}, and its objective value satisfies

$f(c, r^*) = \sum_{j=1}^{n} w_j |r^* - d(c, a_j)| \leq \sum_{j=1}^{n} w_j |r - d(c, a_j)| = f(c, r)$. QED

At the other extreme, one might conjecture the existence of an optimal circle containing three existing facilities, whenever the number of existing facilities is four or more. This conjecture is false in general, as shown by the following example: consider $n = 6$ existing facilities with locations $a_1 = (-4t, 0)$, $a_2 = (4t, 0)$, $a_3 = (-5t, 0)$, $a_4 = (5t, 0)$, $a_5 = (0, 3t)$, $a_6 = (0, -3t)$, where $t = \pi/20$, and unit weights. Using Maple, we computed the objective function values for those circles containing at least three existing facilities; they were all larger than the value for the circle centered at $(0, 0)$ with radius $4t$, intersecting exclusively a_1 and a_2. It is still an open question, if there always exists an optimal circle containing at least two existing facilities.

Finding the global minimum requires a systematic search. Using Lemma 6, it follows that for a given center c, and after sorting the distances $d(c, a_j)$ in nondecreasing order, an optimal radius $r^*(c)$ may be found in linear ($O(n)$) time. The resulting circle $C(c, r^*(c))$ intersects at least one of the a_j, and divides the set of existing facilities into three subsets, $J_=$, J_+, and J_- (that are also functions of c). The objective function becomes:

$$f(c, r^*(c)) = \textstyle\sum_{j\in J_+} w_j(r^* - d(c, a_j)) + \sum_{j\in J_-} w_j(d(c, a_j) - r^*).$$

Solving (CS) is thus reduced to finding an optimal center c^*. But $f(c, r^*(c))$ is seen to be a sum of convex and concave terms, that may in general contain several local minima. Furthermore, since J_+ and J_- also change as c and $r^*(c)$ change, the objective function will not be differentiable everywhere. These difficulties may be handled by a branch-and-bound procedure, as outlined in the simple approach below. (Note that a branch-and-bound method is also proposed for the well-known median point location problem on the sphere [4].)

Since any circle C with center c may also use the antipode c' as the center of the circle, it suffices to consider only the northern hemisphere for the purposes of finding c^*. We therefore begin by dividing the northern hemisphere into subsets S_i by drawing great circles along the longitude at equi-spaced increments of π/N_1 starting at $x_2 = 0$, and circles along the latitude at equi-spaced increments of $\pi/(2N_2)$ starting at $x_1 = 0$, where the parameters N_1 and N_2 depend on the accuracy required. Each subset (or cell) S_i may be characterized by a mid-point $c_i = (x_{i1} + \pi/(4N_2), x_{i2} + \pi/(2N_1))$, and objective function value $f(c_i, r^*(c_i))$, which may be used as an upper bound $\overline{f}_i$ for the optimal solution in S_i.

To find a lower bound $\underline{f}_i$ associated with S_i, let $\underline{d}_j(S_i) = \min_{c\in S_i} d(c, a_j)$, $\overline{d}_j(S_i) = \max_{c\in S_i} d(c, a_j)$, $\underline{r}(S_i) = \min_{c\in S_i} r^*(c)$, and $\overline{r}(S_i) = \max_{c\in S_i} r^*(c)$. Since the elements of J_+ and J_- are not known at this stage, a simple lower bound may be formulated as follows:

$$\underline{f}_i = \textstyle\sum_{j=1}^{n} w_j(\max\{0, \underline{r}(S_i) - \overline{d}_j(S_i)\} + \max\{0, \underline{d}_j(S_i) - \overline{r}(S_i)\}).$$

The maximum distance from a_j to S_i (denoted by $\overline{d}_j(S_i)$) occurs at the corner point of S_i that is furthest from a_j, or at a point on the furthest edge of S_i where a circle centered at a_j just becomes tangent to the edge. Similarly, the minimum distance $\underline{d}_j(S_i)$ may occur at the closest corner point of S_i, or a point along the closest edge of S_i where a circle centered at a_j just becomes tangent to the edge. Both distances are easily determined. This no longer holds for the minimum and maximum radii ($\underline{r}(S_i)$ and $\overline{r}(S_i)$) since the point a_{j*} that intersects $C(c, r^*)$ may change as a function of $c \in S_i$. However, the following result may be used to resolve this difficulty.

Assume that $J_+ \cup J_=$ and $J_- \cup J_=$ each contain at least three elements, $\forall c \in S$ (otherwise minor modifications are required).

Lemma 7 *Let $\overline{c}_i = \arg\max_{c \in S_i} r^*(c)$ and $\underline{c}_i = \arg\min_{c \in S_i} r^*(c)$. Then both $\overline{c}_i$ and $\underline{c}_i$ may be found satisfying one of the following conditions:*
(i) it coincides with the center of a circle intersecting three fixed points a_h, a_j, a_k (denote this center as $c(h, j, k)$);
(ii) it coincides with the intersection of a bisector B_{hj} with an edge of S_i;
(iii) it is a corner point of S_i; or
(iv) it is a point on an edge of S_i, such that $r^*(\overline{c}_i) = d(\overline{c}_i, a_{j*})$ ($r^*(\underline{c}_i) = d(\underline{c}_i, a_{j*})$), for some a_{j*}, and this point maximizes (minimizes) the distance from a_{j*} to the edge.

Proof: Referring to Lemma 6, it follows that the circle $C = C(c, r^*(c))$ may be rotated until it contains three of the existing facilities, without violating the existing median property; i.e., $\sum_{j \in J_+ \cup J_=} w_j \geq \sum_{j \in J_-} w_j$, and $\sum_{j \in J_- \cup J_=} w_j \geq \sum_{j \in J_+} w_j$. Thus, an unconstrained local maximum (or minimum) of $r^*(c)$ can only occur at an intersection point $c(h, j, k)$ that satisfies the median property. This implies that any interior point of S_i is a candidate only if it is the center of a circle intersecting three existing facilities.

Now suppose $\overline{c}_i$ ($\underline{c}_i$) occurs at a point on the boundary of S_i that does not satisfy condition (ii), (iii), or (iv). Then, clearly, we can move $\overline{c}_i$ ($\underline{c}_i$) in a direction along the edge without decreasing $r^*(\overline{c}_i)$ (increasing $r^*(\underline{c}_i)$) until one of these conditions is satisfied, whichever occurs first.

QED

A preprocessing step may be used to determine the relevant $c(h, j, k)$, the relevant bisector segments, and the closest and furthest points on the edges of S_i to each a_j. The lower bounds $\underline{f}_i$ are then set up using the maximum value $\overline{r}(S_i)$ and the minimum value $\underline{r}(S_i)$ obtained from the candidates in each S_i identified by Lemma 7. Also note that the optimal solution may actually be one of the circles intersecting three of the existing facilities. If $\underline{f}_i \geq \min_\ell \overline{f}_\ell$, we delete cell S_i; otherwise S_i is divided in four equally sized parts and the lower bounds for each subcell re-calculated. The branch-and-bound process may be terminated when the difference between the incumbent solution (i.e., best solution so far) and the minimum lower bound is within an acceptable tolerance.

The Weiszfeld procedure for Euclidean distances (e.g., see [8]) may be adapted in a similar fashion as in the well-studied Weber problem on the sphere, where a point rather than a circle is to be located (see the survey in [4]), to provide an approximate solution method for (CS). The method uses the first-order conditions for a stationary point to derive a simple single point iterative scheme. However, just as in the Weber problem on the sphere, since the objective function is generally non-convex, the stationary point may be a local minimum or maximum, or a saddle point; furthermore, convergence is not guaranteed.

In order to assure that the objective function is differentiable everywhere in the standard Weiszfeld procedure, a hyperbolic approximation or smoothing of the distance function is introduced (see [8]). In our case we propose a different approach of perturbing the radius a small amount by alternately setting $r = r^*(c) \pm \Delta$ in order to impose that $J_=$ is the null set at all times. The partial derivatives are then given by:

$\partial f/\partial c_1 = \sum_{j \in J_-} w_j(\sin c_1 \cos a_{j1} \cos(c_2 - a_{j2}) - \cos c_1 \sin a_{j1})/B_j$
$\quad - \sum_{j \in J_+} w_j(\sin c_1 \cos a_{j1} \cos(c_2 - a_{j2}) - \cos c_1 \sin a_{j1})/B_j,$
$\partial f/\partial c_2 = \sum_{j \in J_-} w_j \cos c_1 \cos a_{j1} \sin(c_2 - a_{j2})/B_j$
$\quad - \sum_{j \in J_+} w_j \cos c_1 \cos a_{j1} \sin(c_2 - a_{j2})/B_j$, where
$B_j = \sin(\arccos(\cos c_1 \cos a_{j1} \cos(c_2 - a_{j2}) + \sin c_1 \sin a_{j1}))$.

Setting the two derivatives equal to zero and simplifying considerably yields

$$\tan c_2 = \frac{\sum_{j \in J_+} w_j \cos a_{j1}(\sin a_{j2})/B_j - \sum_{j \in J_-} w_j \cos a_{j1}(\sin a_{j2})/B_j}{\sum_{j \in J_+} w_j \cos a_{j1}(\cos a_{j2})/B_j - \sum_{j \in J_-} w_j \cos a_{j1}(\cos a_{j2})/B_j},$$

$$\frac{\tan c_1}{\sin c_2} = \frac{\sum_{j \in J_+} w_j(\sin a_{j1})/B_j - \sum_{j \in J_-} w_j(\sin a_{j1})/B_j}{\sum_{j \in J_+} w_j \cos a_{j1}(\sin a_{j2})/B_j - \sum_{j \in J_-} w_j \cos a_{j1}(\sin a_{j2})/B_j}$$

Now a procedure for finding a local minimum may be outlined. We start by choosing an arbitrary point c on the sphere and use this as the center of a circle. Given this center, the optimal radius, r^*, is easily found by solving the equivalent median problem of locating a point facility on a line. Given c and the perturbed radius r, the index sets, J_+ and J_- ($J_= = \emptyset$), may be specified. Now the expression for $\tan c_2$ is used for finding a new value for c_2, and the expression for $\tan c_1/\sin c_2$ is then used to find a new value for c_1. The procedure is continued iteratively until significant changes in the three decision variables no longer occur, indicating that a stationary point is being reached.

7 Conclusions

In this paper we present a new model, the location of a circle on a sphere, which is a natural extension of the linear facility location problem in the plane. The minimax (or center) problem and the minisum (or median) problem are formulated for two cases: location of great circles and general circles. Problem properties are developed that lead to polynomial algorithms for the minimax problem (both cases) and the minisum great circle problem. A branch-and-bound procedure and a single point iterative search method are outlined for the general minisum problem.

Future research directions include further investigation of the proposed solution methods, in particular the branch-and-bound. A related problem of potential interest concerns the location of an "orbit" above a sphere where a third dimension of altitude is introduced.

8 Acknowledgment

We thank two anonymous referees for comments leading to significant improvement of the paper.

References

[1] Das, P., N.R. Chakraborti, and P.K. Chaudhuri, Spherical minimax location problem, *Computational Optimization and Applications*, vol. 18, pp. 311-326, 2001.

[2] Drezner, Z. and G.O. Wesolowsky, Facility location on a sphere, *Journal of the Operational Research Society*, vol. 29, pp. 997-1004, 1978.

[3] Hamacher, H.W. and M. Jayasundara, Spherical location problems, ISOLDE 9, Fredericton, Canada, 2002.

[4] Hansen, P., B. Jaumard, and S. Krau, An algorithm for Weber's problem on the sphere, *Location Science*, vol. 3, pp. 217-237, 1995.

[5] Houle, M.E. and G.T. Toussaint, Computing the width of a set, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 10, pp. 760-765, 1988.

[6] Huckemann, S. and H. Ziezold, Principal component analysis based on geodesics for distributions on spheres with an application to triangular shape spaces, *Advances of Applied Probability*, to appear.

[7] Katz, I.N. and L. Cooper, Optimal location on a sphere, *Computers and Mathematics with Applications*, vol. 6, pp. 175-196, 1980.

[8] Love, R.F., J.G. Morris, and G.O. Wesolowsky, *Facilities Location - Models & Methods*, North-Holland, New York, 1988.

[9] Díaz-Báñez, J.M., J.A. Mesa, and A. Schöbel, Continuous location of dimensional structures, *European Journal of Operational Research*, to appear

[10] Drezner, Z., G. Steiner, and G.O. Wesolowsky, On the circle closest to a set of points, *Computers and Operations Research*, vol. 29, pp. 637-650, 2002.

[11] Morris, J.G. and J.P. Norback, A simple approach to linear facility location, *Transportation Science*, vol. 14, pp. 1-8, 1980.

[12] Morris, J.G. and J.P. Norback, Linear facility location - solving extensions of the basic problem, *European Journal of Operational Research*, vol. 12, pp. 90-94, 1983.

[13] Norback, J.P. and J.G. Morris, Fitting hyperplanes by minimizing orthogonal deviations, *Mathematical Programming*, vol. 19, pp. 102-105, 1980.

[14] Patel, M.H. and A. Chidambaram, A new method for minimax location on a sphere, *International Journal of Industrial Engineering*, vol. 9, pp. 96-102, 2002.

[15] Schöbel, A., Anchored hyperplane location problems, *Discrete & Computational Geometry*, vol. 2, no. 29, pp. 229-238, 2003.

[16] Schöbel, A., *Locating Lines and Hyperplanes*, Kluwer, Dordrecht, 1999.

[17] Schömer, E., J. Sellen, M. Teichmann, and Ch. Yap, Efficient algorithms for the smallest enclosing cylinders problem, in: *Proceedings of the 8th Canadian Conference on Computational Geometry*, 1996.

[18] Wesolowsky, G.O., Location of the median line for weighted points, *Environment and Planning A*, vol. 7, pp. 163-170, 1975.

[19] Wesolowsky, G.O., Location problems on a sphere, *Regional Science and Urban Economics*, vol. 12, pp. 495-508, 1982.