ELSEVIER

European Journal of Operational Research 119 (1999) 158–180

EUROPEAN JOURNAL OF OPERATIONAL RESEARCH

www.elsevier.com/locate/orms

Theory and Methodology

Undesirable facility location with minimal covering objectives [1]

Frank Plastria [a,*], Emilio Carrizosa [b,2]

[a] *Department of Management Informatics, Vrije Universiteit Brussel, Pleinlaan 2, B 1050 Brussels, Belgium*

[b] *Departamento de Estadística e Investigación Operativa, Universidad de Sevilla, Tarfia s/n, 41012 Sevilla, Spain*

Received 1 April 1997; accepted 1 September 1998

Abstract

An undesirable facility is to be located within some feasible region of any shape in the plane or on a planar network. Population is supposed to be concentrated at a finite number n of points. Two criteria are taken into account: a radius of influence to be maximised, indicating within which distance from the facility population disturbance is taken into consideration, and the total covered population, i.e. lying within the influence radius from the facility, which should be minimised. Low complexity polynomial algorithms are derived to determine all nondominated solutions, of which there are only $O(n^3)$ for a fixed feasible region or $O(n^2)$ when locating on a planar network. Once obtained, this information allows almost instant answers and a trade-off sensitivity analysis to questions such as minimising the population within a given radius (minimal covering problem) or finding the largest circle not covering more than a given total population.

Keywords: Undesirable facility location; Bicriterion covering problem; Euclidean distance; Minimal covering; Largest circle; Voronoi diagram

1. Introduction

Many necessary facilities have an unwanted impact on their environment, be it potentially for hazardous facilities or continuously in case of (ob)noxious facilities, leading to the so-called 'not in my backyard' (NIMBY) effect. The location decision is then a delicate question leading to many trade-off considerations. Although fundamentally of multicriteria nature, with several rather qualitative and even subjective criteria (see e.g. the surveys by Erkut and Neuman (1989) and Kleindorfer and Kunreuther (1994)), it turns out that in most cases the main objective will be to choose a site as far as possible from any kind of population (or other) centre which might be affected by the facility's presence. Such an aim, taken literally, leads to *maximin*-type location problems, i.e. seeking the site which maximises the distance to the closest (and hence most affected) population (cf. e.g. Drezner and Wesolowsky (1980) and the recent survey, Plastria (1996)).

* Corresponding author. E-mail: frank.plastria@vub.ac.be

[1] Partially supported by Grant PB96-1416-C02-02 of the D.G.E.S., Spain.

[2] E-mail: ecarriz@cica.es

0377-2217/99/$ – see front matter
PII: S 0 3 7 7 - 2 2 1 7 (9 8) 0 0 3 3 5 - X

Typically, however, legal rules on proximity of noxious and/or hazardous facilities, use a simple threshold rule, stating that a population centre is considered as affected only when the intensity of the facility's undesirable effect felt at this centre is beyond this threshold value. And in our densely populated countries there will always be such affected people, wherever the facility will be located. Therefore the simple maximin objective might better be replaced by the minimisation of the total population affected by the facility, which leads to so-called *minimal covering problems*.

At the one hand, however, legal rules, and threshold values in particular, do change under various socio-economic pressures. At the other hand many choices of technology and/or investment in pollution abatement installations may be under consideration, strongly influencing the intensity of the undesirable effects of the facility. Therefore it is of great interest to be able to measure the trade-offs between the threshold intensity on the one hand and the affected population on the other hand. Thus the aim of this paper is threefold:

- first to introduce the ensuing *bicriteria model* seeking *lowest affection of population at the highest level of protection*,
- second to prove that in quite realistic context the set of *efficient (Pareto-optimal) solutions is finite* and
- third to develop fast polynomial algorithms to *construct the complete trade-off curve* between both objectives together with corresponding efficient solutions.

Many undesirable effects, such as radioactivity, radio interference, noise, heat, odour etc. are felt continuously over a relatively small geographical space and there is a clear decreasing relationship between the intensity felt at a point and its Euclidean distance to the facility's site, e.g. an inverse squared distance type of law, (cf. e.g. Melachrinoudis and Cullinane, 1986). Therefore we may consider the location problem as stated in the *Euclidean plane*, and the threshold intensity is interpreted as a threshold distance r. The affected points are then those lying within a circular disk of radius r centered at the facility's (unknown) site x. Both the site x and the radius r are taken as variables, leading to location of a circular disk of unknown radius.

Note that whereas use of Euclidean distance seems to exclude applications involving airborne pollution due to the presence of winds, it is possible to adapt our methodology also to this case, as will be explained in the concluding section.

In order to be applicable in the real world any location model, particularly for undesirable facilities, must take into account many types of locational constraints. First of all regional planning rules allow such facilities in certain areas only. Second the topography and local geological and/or meteorological conditions may rule out other areas. Finally law may forbid closeness to certain points. As a result the remaining area from which the facility's site should be chosen has quite irregular and strange shapes, as a rule. Avoiding such feasible regions (often termed 'pathological') for the sake of mathematical elegance is a guarantee that the model will *not* be applied in practice. Therefore we prefer to directly tackle feasible regions of very general shapes, thereby accepting a possible loss in terms of shortness and simplicity. Fortunately it turns out that this does not lead to significant new difficulties, a point that is mostly not discussed in other work.

A very special type of feasible region, but of particular interest in practice, is a *network* which is part of the geography, thus in our model embedded in the plane. Most facilities should be directly accessible using the existing road-network, and are not sufficiently important to allow for the extension of the existing network in order to make the facility reachable. Therefore a site along (part of) the existing network is sought, while the undesirable effects are of course still felt continuously over space. It turns out that this special type of problem, which seems never to have been tackled before, is in fact easier to solve than the fully continuous version.

The paper is organised as follows: Section 2 states the model in formal terms as a bicriteria problem of locating (Euclidean) disks in the plane. Section 3 describes in more detail how the information obtained from a full solution of this bicriteria problem may be used in different ways. In Section 4 necessary conditions are derived for a disk to be efficient. These conditions lead in Section 5 to the announced finiteness result from which the basic ideas follow for an algorithm to construct the full trade-off curve with corresponding efficient disks. Section 7 analyses further the geometric nature of the problem leading to lower complexity algorithms which are slightly more complex. In Section 8 we indicate several extensions and make suggestions for further research.

2. Problem statement and scope

Let A be a finite subset of $\mathbb{R}^2$ of points of the plane, e.g. population centres. Let also be given a function $w: A \to \mathbb{R}_+$, associating with each $a \in A$ some positive weight $w(a)$ (e.g. population at a, value of installations at a, etc.).

We consider open circular disks of varying radius r and center x lying in some closed feasible region $S \subset \mathbb{R}^2$:

$$iB(x,r) = \{y \in \mathbb{R}^2 \mid \|y - x\| < r\}, \tag{1}$$

where $\|.\|$ denotes the usual Euclidean norm. In the context of the location of an undesirable facility such a disk corresponds to the region of the plane affected by the facility, under the following assumptions:

- the facility is located at site $x \in S$;
- the intensity of the undesirable effect of the facility decreases with the Euclidean distance from it;
- a point is supposed to be affected if the intensity felt at this point is above some threshold value (e.g. legal limit) as represented by the radius r.

Thus, we may say that the 'solutions', or *disks*, we are interested in are given by pairs (x, r), where x is a site in S and r is some nonnegative radius.

For a disk (x, r), the set of *affected destinations* is given by

$$A(x,r) = iB(x,r) \cap A = \{a \in A \mid \|a - x\| < r\} \tag{2}$$

and the global affected weight (or population), called the *coverage*, is then

$$\operatorname{cov}(x,r) = \sum_{a \in A(x,r)} w(a). \tag{3}$$

In this context the following two contradicting objectives are quite natural:

1. to reduce as much as possible the affected population, in other words to minimise the coverage;
2. to increase as much as possible the radius of the disk of affected points, possibly in order to raise the level of legal protection by lowering the threshold, or, to use less costly technology for the facility, implying an increase of the intensity felt at a given distance.

We therefore seek to construct the *set of efficient* (Pareto-optimal or nondominated, cf. Chankong and Haimes (1983)) *disks* for the resulting bi-objective problem (BP):

$$\begin{aligned} \min \quad & \operatorname{cov}(x,r) \\ \max \quad & r, \\ & x \in S, \quad r \in \mathbb{R}_+. \end{aligned}$$

We will say that a solution of (BP) (x', r') *(strictly) dominates* the solution (x, r) iff $r' \geqslant r$ and $\operatorname{cov}(x', r') \leqslant \operatorname{cov}(x, r)$ (with at least one strict inequality). (x, r) is an *efficient solution* of (BP) iff there does not exist another solution (x', r') which strictly dominates (x, r).

Describing the efficient set is possible thanks to the result shown in Section 5 that for fixed compact regions S of fairly general shape there are at most $O(n^3)$ efficient disks. Furthermore, Section 7 develops ways to effectively construct and evaluate these for particular shapes S in $O(n^3 \log n)$ time by relatively simple geometric constructions, reducible to $O(n^3)$ in the case of equally weighted points or when S is (part of) a network. In this way a full trade-off curve is obtained between the coverage and the radius of disks.

An important question, in particular for the applications mentioned in the next section is whether any dominated disk is also dominated by some efficient disk. Clearly this is not true in general, since any disk covering A fully is dominated by any larger enclosing disk (same coverage, larger radius). Fortunately this is the only exception, as shown by the next lemma.

Lemma 1. *If S is compact, every dominated disk which does not completely cover A is dominated by at least one efficient disk.*

Proof. Let (x_0, r_0) represent a disk which does not completely cover A, i.e.,

$$\mathrm{cov}(x_0, r_0) < \sum_{a \in A} w(a). \tag{4}$$

Define ω_0 as

$$\omega_0 := \min\{\mathrm{cov}(x, r) \mid x \in S, r \geqslant r_0\}, \tag{5}$$

which is achieved due to the fact that the function cov takes only a finite number of values. Hence it follows from Eq. (4) that

$$\omega_0 \leqslant \mathrm{cov}(x_0, r_0) < \sum_{a \in A} w(a).$$

Let $(\hat{x}, \hat{r})$ be an optimal solution for

$$\begin{aligned} \max \quad & r \\ \text{s.t.} \quad & \mathrm{cov}(x, r) \leqslant \omega_0, \\ & x \in S, \\ & r \geqslant r_0. \end{aligned} \tag{6}$$

Observe that such optimal solution exists by Eq. (4), the compactness of S and the lower semicontinuity of cov. Then, $(\hat{x}, \hat{r})$ is an efficient disk; indeed, else there would exist a feasible (x^*, r^*) satisfying

$$\begin{aligned} & r^* \geqslant \hat{r} \ (\leqslant r_0), \\ & \mathrm{cov}(x^*, r^*) \leqslant \mathrm{cov}(\hat{x}, \hat{r}) \ (\leqslant \omega_0), \end{aligned} \tag{7}$$

with at least one inequality strict. In particular, (x^*, r^*) would be feasible for Problem (6), implying

$$\hat{r} = r^* \geqslant r_0.$$

Moreover, (x^*, r^*) would also be feasible for problem (5), thus

$$\omega_0 \leqslant \mathrm{cov}(x^*, r^*) \leqslant \mathrm{cov}(\hat{x}, \hat{r}) \leqslant \omega_0,$$

thus $\mathrm{cov}(x^*, r^*) = \mathrm{cov}(\hat{x}, \hat{r})$, contradicting the fact that, in Eq. (7), one of the inequalities should be strict.

Hence, $(\hat{x}, \hat{r})$ is efficient, as asserted. □

It is clear that with an aim of protecting at least some of the points of A from the undesirable effects of the facility to be located, disks fully covering A are of no interest. Therefore the restriction in Lemma 1 is not limiting in practice, whence our attention may further be restricted to efficient disks only.

We now show that, in fact, only a finite set of efficient disks need to be considered. The cardinality of such a set and how to construct it are the aim of Sections 4–7.

Lemma 2. *There exists a finite set C of efficient disks such that any disk (x, r) not completely covering A is dominated by some disk in C.*

Proof. The scheme of the proof is similar to Lemma 1: for any nonempty $B \subset A, B \neq A$, define (x_B, r_B) to be an optimal solution for

$$\begin{aligned} \max \quad & r \\ \text{s.t.} \quad & \mathrm{cov}(x, r) \leqslant \sum_{b \in B} w(b), \\ & x \in S, \\ & r \in \mathbb{R}_+. \end{aligned}$$

By Lemma 1, there exists an efficient disk (x_B^*, r_B^*) dominating (x_B, r_B).

It then follows that the list $\{(x_B^*, r_B^*) : \emptyset \neq B \neq A\}$ can be taken as C. Indeed, any disk (x, r) will be dominated by (x_B^*, r_B^*) with $B = A(x, r)$. □

3. Applications

As explained in the introduction the location problem at hand is normally a bi-objective one. In practice such problems are usually reduced to a single-objective problem, either by constructing a utility function aggregating the two objectives or putting one objective as a constraint by limiting it by some threshold value and optimizing the other objective, see Chankong and Haimes (1983). The other possibility is to tackle the biobjective problem directly, perhaps with the aid of interactive methods, see Chankong and Haimes (1983). It turns out that knowledge of the full trade-off curve as advocated in this paper is invaluable in all these approaches, as briefly discussed below.

3.1. Utility functions

One approach is the minimisation of some (dis)utility function combining the radius in a decreasing and the coverage in an increasing way. This supposes that one is able to translate both radius and coverage into some comparable values, in other words that one is able to evaluate the practical trade-offs between these two conflicting criteria. Although this is not always perfectly possible it might be of interest to attempt.

The main advantage of this approach is that a single objective problem arises for which the notion of optimality is clear-cut. Evidently any optimal solution for this single objective problem will be Pareto-optimal for the bicriteria problem (BP) and thus an efficient disk. Since these are only finite in number, a simple scanning of these candidate solutions will yield an optimal one.

It is quite remarkable that we thus obtain a finite method of fairly low complexity to solve an at first glance totally intractable continuous optimisation problem, the objective being neither convex nor concave nor even continuous!

3.2. Minimal covering problems

Suppose the radius R of (undesirable) influence of the facility is given. Putting this objective as constraint, one obtains the *minimal covering problem*, which asks for the center of a disk of radius at least R covering the least possible weight. This problem was studied by Drezner and Wesolowsky (1994). Note that the circular region they consider is supposed to include the full disk, and thus in our interpretation of the feasible region we must consider only the points at least at distance R from the boundary, which, in this particular case, is still a circle.

In our approach, one first constructs the finite list C defined in Lemma 2, and, for the R given, one finds the disk $(x^*, R^*) \in C$ with R^* being the lowest radius in the list C greater than or equal to R. Such disk is clearly optimal for the minimum covering problem and its radius may be even greater than the threshold value R. The complexity of this task will be derived in Corollary 14.

This strategy enables one to analyse directly the sensitivity of changes in the radius, a feature which is extremely important to decision makers in practice.

Moreover, the answer thus obtained is even more instructive than Drezner and Wesolowsky's: it not only gives some minimal covering disk of radius R, but among all such disks (and there are usually a continuous set of such) immediately produces one which maximises the minimal distance to the uncovered points, thus offering the best protection possible in maximin sense.

3.3. Largest circle problems

Instead of fixing the radius and minimising the covered weight, it is a politically relevant question to fix the maximal weight W which is allowable to be covered and maximise the radius of the disk, i.e. the level of protection for the uncovered points. We will call such a question a *largest circle problem* or a *maxquantile location problem*. The first term generalises the better known *largest empty circle problem*, in which $W = 0$, and it is required to find the (center of the) disk of largest radius not containing any destination point (see Toussaint, 1983 and Preparata and Shamos, 1985). The second term explains that the objective to be maximised may be considered as a quantile of the distribution of distances to the given points, weighted by the w_a (considered to sum up to 1). This latter term is perhaps better suited when other distance measures are used, as suggested in Section 8, and may be seen as the undesirable counterpart of the minquantile objective introduced by Carrizosa and Plastria (1995) in general and Carrizosa and Plastria (1999) in the context of attracting facility location.

Classically the largest empty circle problem is solved most efficiently by way of the Voronoi-diagram (see Preparata and Shamos, 1985, p. 249–252). This does not extend directly to nonzero coverage since for a fixed W we do not know which order Voronoi diagram is to be constructed (compare with Section 8.1). However, it will be shown in Section 7.4 that considering the Voronoi diagrams of all orders yields the full set of efficient disks.

Evidently any W-largest circle (with W not too high in order to allow for nonfull coverage of A) is either an efficient disk or (by Lemma 1) there exists an efficient disk of the same radius with a smaller coverage. Therefore our approach enables an almost instant answer to any largest circle query: it suffices to find the largest efficient disk with coverage at most W. The complexity of this task will be derived in Corollary 15.

In a sense we thus obtain the 'ideal' W-largest circle, since it is one with smallest possible coverage, among all the circles with optimal radius. Since this is obtained for all possible weights W we have a full sensitivity analysis with respect to the covered weight.

3.4. MCDM

The bi-objective problem may also be further analysed directly by any technique of multicriteria analysis. Any such method should at least eliminate dominated solutions and thus select a nondominated or efficient one. Since, by Lemma 2 we end up here with a finite list of such, any method, even those devised for finite situations might be used, (see e.g. Vincke et al., 1992), with the additional advantage of allowing inclusion of more criteria.

It should however be stressed that when additional criteria are to be taken into account, which ought to always be the case (compare the comments of Erkut and Neuman (1989)), care must be taken to include all alternate efficient disks (if any) in the final analysis, since these may be far from equivalent in view of the new criteria. Therefore the algorithms presented here will have to be slightly adapted in order to make sure that no solutions equivalent to an efficient one are abandoned.

4. Necessary conditions for efficiency

We will start in this section by deriving conditions that any disk should satisfy in order to be eligible for efficiency. These conditions will very strongly limit the candidate disks, as indicated for fairly general regions S in next section. In this section S is merely assumed compact.

We will call point $a \in A$ *active* at solution (x, r) iff a lies exactly at distance r from x. The set of all active points at (x, r) is denoted by

$$\mathrm{act}(x,r) = \{a \in A \mid \|a - x\| = r\}. \tag{8}$$

Please observe that active points are considered *not* to be covered, contrary to the attractive case (see Carrizosa and Plastria, 1999).

We recall that, by the definition of efficiency, in order to prove that some (x, r) is not efficient it will suffice either to exhibit another solution which strictly dominates it, or to exhibit another equivalent solution (same radius, same coverage) which is known not to be efficient.

Lemma 3. *If (x, r) is efficient then active points exist, i.e.* $\mathrm{act}(x, r) \neq \emptyset$.

Proof. Suppose to the contrary that $\mathrm{act}(x, r) = \emptyset$, then for any $a \notin A(x, r)$ we have $\|a - x\| > r$. Since there are only finitely many of these, we will also have $\|a - x\| > r + \epsilon$ for all $a \notin A(x, r)$ for some sufficiently small $\epsilon > 0$. This means, however, that $A(x, r + \epsilon) = A(x, r)$, and thus $\mathrm{cov}(x, r + \epsilon) = \mathrm{cov}(x, r)$, while $r + \epsilon > r$. Hence (x, r) is dominated by $(x, r + \epsilon)$, and therefore cannot be efficient. □

For some nonempty subset B of A define the distance $d_B(y)$ of y from B by

$$d_B(y) = \min_{b \in B} \|b - y\|.$$

Lemma 4. *If (x, r) is efficient then x is a local maximum in S of the distance from* $\mathrm{act}(x, r)$.

Proof. By continuity of distance for each $a \in A$ such that $\|a - x\| < r$, there exists a neighbourhood V_a of x, such that $\|a - y\| < r$ for all $y \in V_a$. Similarly, for all $b \in A$ such that $\|b - x\| > r$, there is a neighbourhood V_b of x, such that $\|b - y\| > r$ for all $y \in V_b$. Let V be the (finite) intersection of all these V_a and V_b, then V is also a neighbourhood of x.

Suppose now that x is not a local maximum in S of the distance from $B = \mathrm{act}(x, r)$, which is nonempty by Lemma 3. Then there exists some $z \in V \cap S$ such that $d_B(z) > d_B(x)$. By definition of $B = \mathrm{act}(x, r)$ we then have for all $a \in \mathrm{act}(x, r)$:

$$\|a - z\| > d_B(x) = r.$$

By construction of V it follows that $A(z, r) = A(x, r)$ with $z \in S$, hence that (x, r) is equivalent to

(z, r), while $\mathrm{act}(z, r) = \emptyset$. By Lemma 3 it follows that (z, r), and hence (x, r), is not efficient. □

We may therefore restrict ourselves to studying such local maxima in S. This will be done in the lemmas below, but first we need some preliminary notions and results.

For any fixed point a the distance function d_a defined by $d_a(x) = \|a - x\|$ is a convex function, differentiable at any point $x \neq a$ with gradient $\nabla d_a(x) = (x - a)/\|x - a\|$. At such a point x the directions of (at least initial) ascent are given by the closed halfspace

$$H_a^+ = \{p \mid \langle\ p\ ;\ x - a\ \rangle \geqslant 0\},$$

where $\langle\ .\ ;\ .\ \rangle$ denotes the scalar product. Notice that the directions orthogonal to $x - a$ are also ascent directions of d_a at x since this function has circular level sets, which are strictly convex. At $x = a$, which is the unique minimum of d_a, every direction is one of ascent. (Note that for simplicity we consider the zero vector also a direction of ascent!)

Let B be some finite set of points, then, clearly, a vector p is a *direction of ascent* at x for d_B iff it is a direction of ascent at x for every active $b \in B$, i.e. any b such that $d_b(x) = d_B(x)$. In the cases we are interested in, when $B = \mathrm{act}(x, r)$, all $b \in B$ are active at x, and $x \notin B$. In such a case, it follows that p is a direction of ascent of d_B at x iff it is an ascent direction for d_b for each $b \in B$, i.e. $p \in \bigcap_{b \in B} H_b^+(x)$. (Note that after replacing B by $B \setminus \{x\}$, the same property would also apply for $x \in B$.)

A direction $p \neq 0$ is *feasible* at $x \in S$ iff the open halfline with direction p emanating from x has points in common with S within any neighbourhood of x. Evidently, a point x is a local maximum in S of some continuous function f iff no nonzero direction of ascent of f at x is a feasible direction for S.

We call a point $x \neq a$ *S-remote* (or anti-visible) from a iff $x \in S$ and there exists no other point $y \in S$ such that $x \in]y, a[$ (see Hansen et al., 1981). A point x is *locally S-remote* from a when it is $S \cap V$-remote from a for some neighbourhood V of x. Note that any locally S-remote point is always a boundary point of S.

The following definitions extend some well known notions in convex analysis (cf. Hiriart-Urruty and Lemarechal, 1993) to nonconvex regions. Note that we have to use local notions due to the loss of convexity.

A point $x \in S$ is a *local extreme point* of S iff there exists some neighborhood V of x such that x does not lie on any open segment joining two points of $S \cap V$.

Finally the *local normal cone* to S at $x \in S$ is defined as

$$N_S(x) = \{p \mid \exists V \text{ neighbourhood of } x, \text{ such that } \forall y \in S \cap V\colon \langle\ p\ ;\ y - x\ \rangle \leqslant 0\}.$$

Lemma 5. *Let* (x, r) *be efficient and* $\mathrm{act}(x, r) = \{a\}$, *then*:

(i) x *is a boundary point of* S, *locally S-remote from* a,

(ii) x *is a local extreme point of* S,

(iii) $x - a$ *belongs to the local normal cone to* S *at* x.

Proof. The function d_a given by $d_a(x) = \|a - x\|$ is continuous and convex. According to Lemma 4, x should be a local maximum of $d_a = d_{\{a\}}$ in S, i.e. x should be the global maximum in S of d_a within some neighbourhood V of x, hence within $S_x = V \cap S$.

The first assertion then follows from the fact that d_a linearly increases along any halfline emanating from a. Similarly the second assertion is a well known fact about the global maximum of a convex function with strictly convex level sets here applied to d_a on S_x. Finally if no nonzero feasible direction of S at x lies in $H_a^+(x)$, this means exactly that $x - a$ belongs to the local normal cone to S at x, which corresponds to the last assertion. □

Fig. 1 illustrates these results. In the situation shown x is a candidate to be the center of an efficient disk because it satisfies all properties stated in Lemma 5; $(y, \|a - y\|)$ is not a candidate efficient disk since y is not locally S-remote from a; nor are $(z, \|a - z\|)$ and $(t, \|a - t\|)$, because at the one hand z is not a local extreme point of S and at the other hand $t - a$ does not belong to $N_S(t)$.

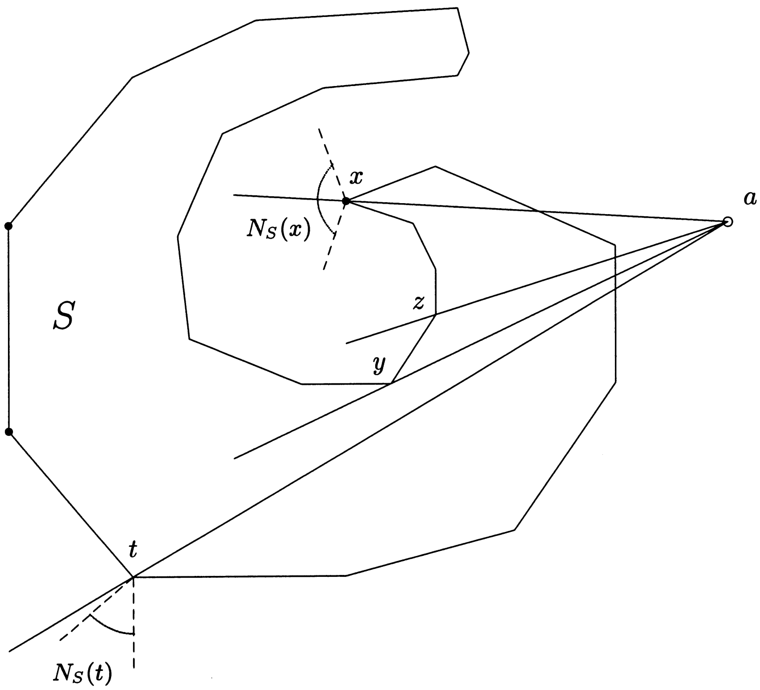

Fig. 1. Examples when $\mathrm{act}(x, r) = \{a\}$.

Observe that of the 23 points defining S's boundary only the three indicated by a black dot satisfy all conditions of Lemma 5.

For two points a and b we define the *mediatrix* $\mathrm{med}(a, b)$ as the set of all points at the same distance from a and b:

$$\mathrm{med}(a, b) = \{x \mid \|a - x\| = \|b - x\|\}.$$

Lemma 6. *Let* (x, r) *be efficient and* $\mathrm{act}(x, r) = \{a, b\}$ *with* $a \neq b$, *then*:

(i) x *lies on* $\mathrm{med}(a, b)$ *and is a boundary point of* S,

(ii) *either* $x \neq m$ *and* x *is locally S-remote from* $m = \frac{1}{2}(a + b)$, *or* $x = m$ *and this is an isolated point of* $S \cap \mathrm{med}(a, b)$.

(iii) *For any S-feasible direction* $p \neq 0$ *at* x *we have* $\langle\, p \,;\, x - a \,\rangle < 0$ or $\langle\, p \,;\, x - b \,\rangle < 0$.

Proof. (i) Since $\mathrm{act}(x, r) = \{a, b\}$ this means that $r = \|a - x\| = \|b - x\|$, hence $x \in \mathrm{med}(a, b)$. This is a straight line passing through $m = \frac{1}{2}(a + b)$, along which the distance to a and to b strictly increases with the distance to m. If x were an interior point of S, the line $\mathrm{med}(a, b)$ would indicate a feasible direction of increase of $d_{\{a,b\}}$, contrary to Lemma 4.

(ii) For exactly the same reason, when $x \neq m$, x must be locally S-remote from m, while when $x = m$ it must be an isolated point of $S \cap \mathrm{med}(a, b)$. (Please note that this latter situation can normally arise only when S contains some one-dimensional pieces, e.g. when S is a network, and when additionally m falls exactly on S – in other words, this case should be considered as quite exceptional.)

(iii) Let $p \neq 0$ be an S-feasible direction at x, then p cannot be an ascent direction of $d_{\{a,b\}}$ at x since by Lemma 4 x should be a local maximum of this function. This means $p \notin H_a^+(x) \cap H_b^+(x)$, whence the conclusion. □

Fig. 2 illustrates Lemma 6. Here $(y, \|a - y\|)$ is not a candidate efficient disk because y is not locally S-remote from m, while $(x, \|a - x\|)$ (resp. $(z, \|a - z\|)$) is not a candidate disk since the S-feasible direction p at x (resp. q at z) shown in the figure has a positive projection on both $x - a$ and

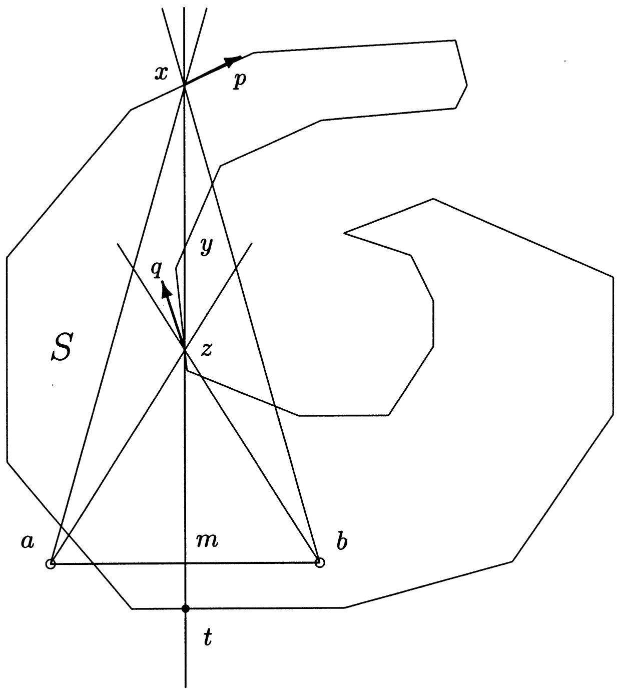

Fig. 2. Examples when $\mathrm{act}(x,r) = \{a,b\}$.

$x - b$ (resp. $z - a$ and $z - b$). So of the 4 points of intersection between S's boundary and $\mathrm{med}(a,b)$, which would be candidates according to Lemma 6(i), only the single point t generates a candidate for being efficient disks with active set $\{a,b\}$.

In general, of the possible candidates following Lemma 6(i), i.e. points of intersection between S's boundary and $\mathrm{med}(a,b)$, due to the remoteness condition in Lemma 6(ii), only approximately half of the finite number of isolated and/or local extreme points will generate candidate efficient solutions, while condition Lemma 6(iii) will possibly limit these remaining choices even further.

Lemma 7. *Let* (x,r) *be efficient and* $|\ \mathrm{act}(x,r)\ | \geqslant 3$.

(i) *If* x *lies in the convex hull of* $\mathrm{act}(x,r)$, *there exist three different noncollinear destinations* $a,b,c \in \mathrm{act}(x,r)$ *such that*:

(a) $x \in S$ *lies in the* (*nondegenerate*) *triangle* $\Delta(a,b,c)$,

(b) *x is the circumcentre of a, b, c, i.e. the centre of the unique circle passing through a, b and c*,

(c) *Triangle* $\Delta(a,b,c)$ *is acute*.

(ii) *If x does not lie in the convex hull of* $\mathrm{act}(x,r)$, *then x is a boundary point of S and there exist two different destinations a and b on which the conditions of Lemma 6 apply.*

Proof. Denote $\mathrm{act}(x,r)$ by B.

(i) If x lies in B's convex hull, then, by Carathéodory's theorem there exist three pairwise distinct points a,b,c of B (we are in $\mathbb{R}^2$!) such that x lies in their convex hull, i.e. the triangle they form. Since $B = \mathrm{act}(x,r)$, x must be at the same distance of a, b and c, which is only possible when these latter points are noncollinear, and x is the centre of the (unique) circle through them. The centre of the circumscribed circle of a triangle lies inside the triangle only when the triangle is acute.

(ii) Suppose now x does not lie in B's convex hull. Then x may be linearly separated from B, i.e. there exists some $p \neq 0$ such that $\langle\ p\ ;\ x\ \rangle \geqslant \langle\ p\ ;\ b\ \rangle$ for all $b \in B$. But this means $p \in \bigcap_{b\in B} H_b^+(x)$, and hence p is a direction of ascent of d_B. This cannot happen when

x is interior to S, since then p would be S-feasible at x, contradicting Lemma 4. Therefore x will be a boundary point (not necessarily extreme) of the convex hull of $B \cup \{x\}$. Calling a and b the two extreme points adjacent to x along this boundary, we have that p is a direction of ascent of d_B iff it is a direction of ascent of $d_{\{a,b\}}$. This shows that the conditions of Lemma 6, applied to $\{a, b\}$, must hold true. □

5. The basic algorithm

The set of candidate efficient sites, i.e. satisfying the necessary conditions of Lemmas 5–7, is usually rather reduced. To simplify matters we assume in this section that

Assumption 8. S is a fixed planar region, possibly disconnected and/or with holes, bounded by a finite number of linear pieces or arcs of quadratic curves.

More general shapes for S are also manageable, but may lead to an important increase in the calculations needed for point inclusion and determination of boundary points either on given straight lines or with certain normal directions. The simplifying assumptions made here are, however, sufficiently general to enable extremely close approximations of any shapes, and already encompass all the more usual shapes used in Geographical Information Systems and suggested in the literature, including their unions and/or intersections, such as polygonal regions (see e.g. Hansen et al., 1981; Plastria, 1992b), circles (see e.g. Drezner and Wesolowsky, 1980, 1994) or complements of circles (see e.g. Hamacher and Nickel, 1995) useful for modeling the NIMBY (Not In My Backyard) principle. Also the case when S is a network embedded in the plane is included, in which case $S = \mathrm{bd}\, S$, the *boundary* of S.

Theorem 9. *For fixed S satisfying Assumption (8), a finite candidate set of efficient disks of cardinality* $O(n^3)$ *may be constructed, such that any disk which does not fully cover A is either dominated or equivalent to some candidate disk.*

Proof. Let us consider each of the three types of candidate efficient disks as described in Lemmas 5, 6 and 7 in turn.

(i) For each fixed $a \in A$ the set $C(a)$ of boundary points of S satisfying the conditions of Lemma 5 is a finite set as a rule, except in one case, when part of bd S consists of some piece(s) of a circle centered at a.

In that particular case, let S_a denote a connected component of the intersection of $\mathrm{bd}\, S$ with $\mathrm{bd}\, B(a, r) = \{x \mid d_a(x) = r\}$, and suppose S_a is not a singleton. Of all disks (x, r) with $x \in S_a$ having the same radius r, only those minimising $\mathrm{cov}(x, r)$ are candidates to be efficient. A linear time search similar to the one proposed in Drezner and Wesolowsky (1994) will yield such an x and in $C(a)$ the full set S_a may then be replaced by this x, all other candidate disks (y, r) with $y \in S_a$ being either dominated or equivalent to the chosen (x, r).

Note: In fact S_a decomposes into connected pieces on which $A(x, r)$, and hence $\mathrm{cov}(x, r)$ is constant, the value changing each time (and only when) S_a intersects one of the circles $\mathrm{bd}\, B(b, r)$ with $b \in A \setminus \{a\}$. Therefore one may easily construct the (possibly several sets of) disks equivalent to the chosen one. Thus, in order to construct all efficient disks one may in this exceptional case additionally store this information next to the chosen x.

It follows that in this way a finite set $C(a)$ arises, with cardinality bounded by the number of pieces on the boundary of S. Correspondingly, a finite set of uniformly bounded cardinality of disks (x, r) with $x \in C(a)$ and $r = d_a(x)$ is obtained for each $a \in A$, yielding in total a set C_1 of $O(n)$ candidate efficient disks (possibly with pointers to additional equivalent ones).

(ii) For each pair of points $\{a, b\} \subset A$ $(a \neq b)$ the (possibly empty) set of points $C(a, b)$ satisfying the conditions of Lemma 6 is a finite set of uniformly bounded cardinality.

Indeed $C(a, b)$ is a subset of the intersection of the straight line $\mathrm{med}(a, b)$ with S's boundary, which by the assumptions on the shape of S, is a finite set, except for the possible presence of a finite number of straight line segments (when $\mathrm{bd}\, S$ contains linear pieces lying – by

chance – on med(a,b)), in each of which only the endpoint farthest from $\frac{1}{2}(a+b)$ has to be taken into account, thanks to Lemma 6(ii).
All disks (x,r) with $x \in C(a,b)$ and $r = d_a(x) = d_b(x)$ yields thus a second set C_2 of $O(n^2)$ candidate efficient disks.
(iii) For each triplet $\{a,b,c\} \subset A$ which form a nondegenerate acute triangle, the centre x of the circumscribed circle is unique. Only when $x \in S$ does one obtain a unique candidate efficient disk (x,r) taking $r = d_a(x) = d_b(x) = d_c(x)$. These yield a third and final set C_3 of $O(n^3)$ candidate efficient disks.

By Lemmas 5, 6 and 7, $C = C_1 \cup C_2 \cup C_3$ contains all efficient disks (or at least for each efficient disk a representative equivalent one), and by Lemma 1 any noncovering dominated disk is indeed dominated by a disk in C. □

Many of these candidates will, however, still be dominated, and have to be deleted from C in order to obtain the desired list of efficient disks. Thanks to Lemma 1, and by transitivity of the domination relation, the identification of the dominated disks in C may be done by comparison among elements of C only. Note that no disk in C fully covers A, since each one has at least one point of A on its boundary, and so does not cover it. This may be done as follows.

- First one has to calculate cov(x,r) for each $(x,r) \in C$. Clearly this may be achieved in a naive way in $O(n^4)$ time: for each of the $O(n^3)$ disks in C simply check each of the n points of A and sum the weights of the covered ones.
- The next and final step consists in building the ordered list of nondominated or efficient disks, by internal comparisons within C.

  The simplest way of achieving this, in general, is by first sorting C into nonincreasing order of radius, and then 'weeding', i.e. scanning the list in this order, deleting each disk with higher coverage than the previous one (since dominated by the latter).

  Since C has $O(n^3)$ elements, the sorting step takes $O(n^3 \log n^3) = O(n^3 \log n)$ and the weeding may be done in $O(n^3)$ time.

  Note that in the unweighted case, when all $w(a)$ are equal, to 1 say, the coverage values may range only over the integers $0, 1, \ldots, n-1$, and thus this final step may trivially be done in $O(n^3)$ time by way of a simple bucketing method.

We have thus obtained a *basic algorithm*, summarised as follows:

1. Disk generation – $O(n^3)$
   (a) Generate C_1 – $O(n)$
   (b) Generate C_2 – $O(n^2)$
   (c) Generate C_3 – $O(n^3)$
2. Disk evaluation – $O(n^4)$
   Calculate cov(x,r) for each $(x,r) \in C = C_1 \cup C_2 \cup C_3$
3. Dominated disk deletion – $O(n^3 \log n)$, or $O(n^3)$ when unweighted.
   Weighted case: sort and weed.
   Unweighted case: bucket.

This basic algorithm is illustrated by the example problem discussed in next section.

Note that the derived complexity is an exact one, and not merely a worst case situation: in all instances the method will require exactly $O(n^4)$ steps. Although this $O(n^4)$ complexity looks like a nice polynomiality result, it will often be too expensive for practice. Indeed, descriptions closely fitting the real-world situation will call for high values of n, particularly when continuously distributed affected individuals are approximated by aggregation in a finite number of discrete points. The difference between the (unavoidable) $O(n^3)$ and $O(n^4)$ obtained above then becomes very important. Therefore it is useful to seek algorithms of lower complexity.

In Section 7 we show that when all constraints are of polygonal type, the disk generation step 1 and the disk evaluation step 2 may better be done simultaneously, yielding techniques of lower complexity. When S is a network the 'naive' basic algorithm leads to another improvement of the complexity.

6. An example problem

As an example consider as feasible set S the nonconvex region already depicted in Figs. 1 and 2, whose vertices, sorted counterclockwise have the

Table 1
Vertices of S

1	(0, 16)	13	(26, 20)
2	(11, 3)	14	(16, 24)
3	(30, 3)	15	(15, 33)
4	(45, 7)	16	(19, 42)
5	(54, 20)	17	(28, 46)
6	(54, 32)	18	(40, 47)
7	(38, 39)	19	(41, 49)
8	(30, 36)	20	(40, 53)
9	(36, 34)	21	(22, 52)
10	(38, 30)	22	(11, 47)
11	(38, 26)	23	(0, 34)
12	(34, 20)	24	(0, 16)

Table 2
The set A: coordinates and weights

i	a_i	ω_i
1	(64, 34)	6/25
2	(24, 40)	1/25
3	(20, 31)	1/25
4	(20, 52)	4/25
5	(27.8, 7)	3/25
6	(45, 55)	1/25
7	(3.8, 7)	6/25
8	(9, 36)	1/25
9	(50, 38)	1/25
10	(60, 19)	1/25

coordinates given in Table 1; the set A of affected points consists of 10 points, with coordinates given in Table 2. Two choices of weights are considered: the first one is given in Table 2, while the second instance assumes all the weights to be equal and summing up to 1.

Lemmas 5–7 enable us to construct a finite dominating set $C := C_1 \cup C_2 \cup C_3$, with cardinality $120 = 14 + 80 + 26$, which is, as discussed above, independent of the choice of the weights. In Fig. 3 we see the feasible region S, the affected destinations A (empty circles) and the centers of the candidate disks (solid circles).

Choosing the weights given in Table 2 and deleting from the list C the dominated disks one obtains the list of 14 efficient disks given in Table 3 and depicted in Fig. 4 (the solid circles represent the centers of the efficient disks). The corresponding trade-off curve is given in Fig. 5. From this we may read off directly that for influence radius of e.g. 30 the minimal feasible coverage is

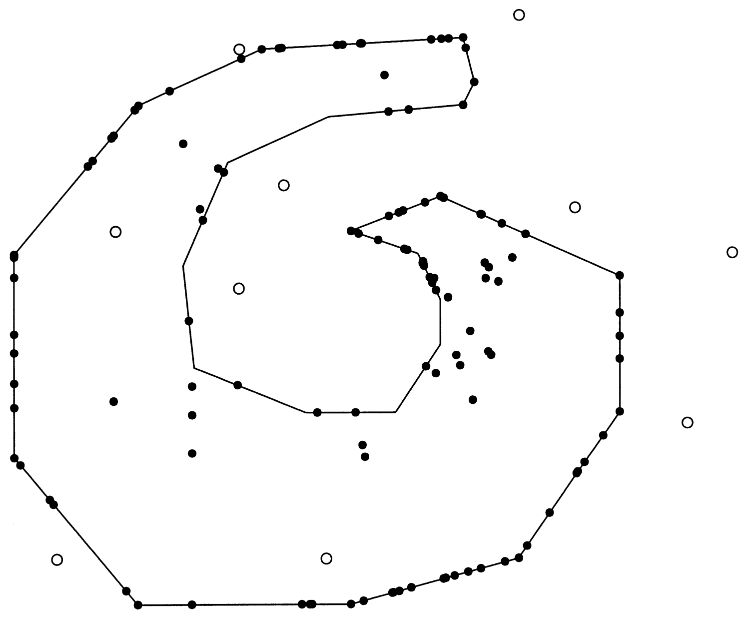

Fig. 3. Centers of candidate disks.

Table 3
Efficient disks (weighted case)

x	r	$\text{cov}(x, r)$
(40.891, 21.074)	19.221	0/25
(50.152, 14.442)	23.558	1/25
(0, 32.008)	25.296	2/25
(33.818, 4.018)	30.166	3/25
(41.627, 6.100)	32.980	4/25
(38.469, 5.258)	34.713	5/25
(40.494, 5.798)	36.713	6/25
(11, 47)	40.025	8/25
(22, 52)	45.372	9/25
(11, 3)	49.820	12/25
(0, 16)	54.626	16/25
(0.547, 15.353)	59.564	17/25
(0, 34)	61.847	18/25
(0, 16)	66.483	19/25

0.12, while in order to reduce coverage to at most 0.08 the influence radius will have to be decreased to at most 25.296.

It is interesting to observe that, when constructing the list C_1, the same center may appear associated with different radii; this happens to be the case even in the final list of efficient disks, where $(0, 16)$ is center of two efficient disks.

Although the list C of candidate disks is independent of the weights, one cannot expect this to happen with the list of efficient disks (See Fig. 6). Indeed, taking now all the weights to be equal, we obtain the list of efficient disks given in Table 4, and the trade-off curve given in Fig. 7.

7. Implementation details and lower complexity methods for polygonal constraints

In this section we address in more detail efficient implementations of the (slightly modified) basic algorithm. The all-over complexity results are expressed for a fixed S. However, where pos-

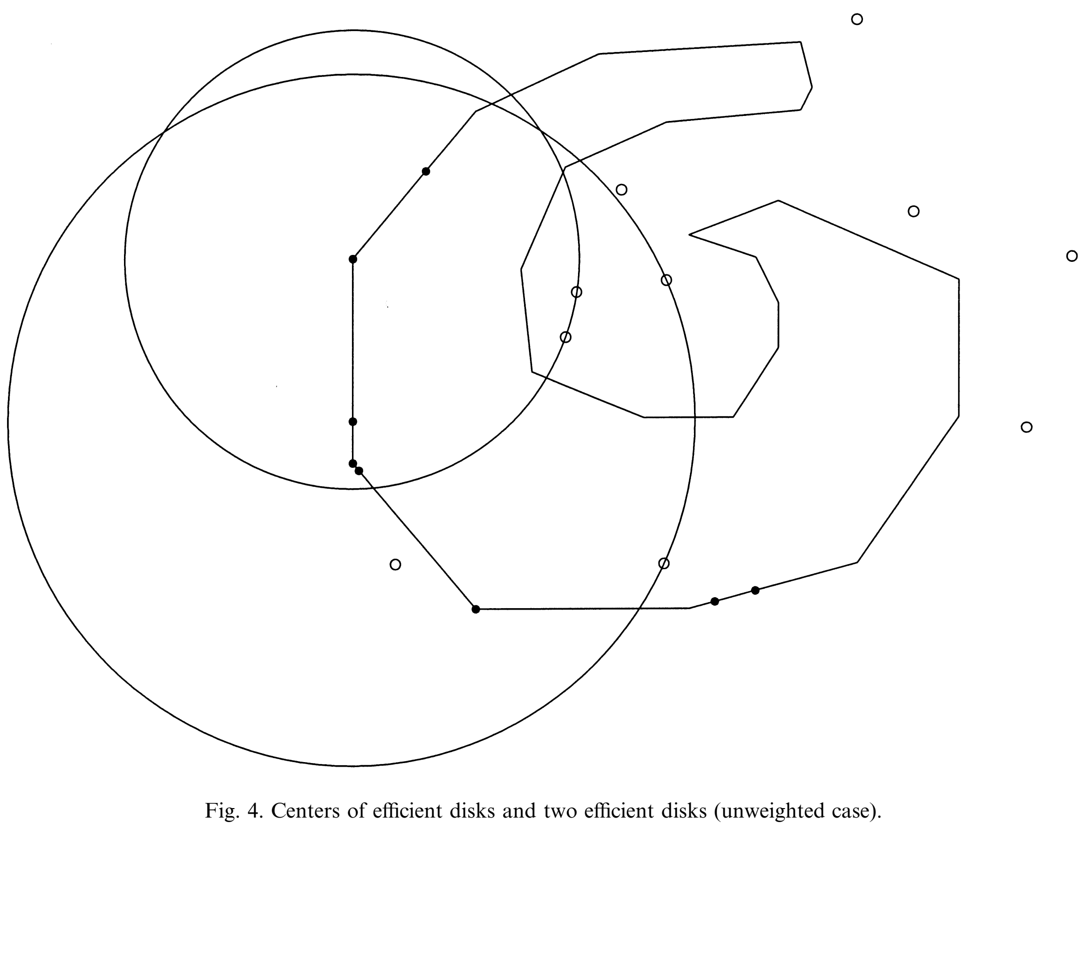

Fig. 4. Centers of efficient disks and two efficient disks (unweighted case).

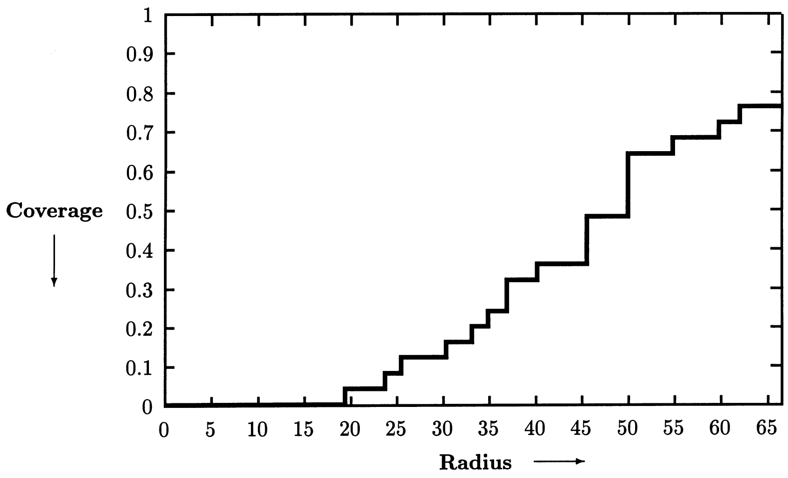

Fig. 5. Trade-off curve (weighted case).

sible in the details about the steps we indicate more precise complexity results related to the size and shape of the feasible region S, which, for sake of simplicity, is assumed here to be polygonal. Most of the described techniques carry over to the more general case discussed in previous section, but the details of these would just complicate matters even more without too much added value.

Let therefore S be a *finite union of p (not necessarily convex) polygons* S_k $(k = 1, \ldots, p)$. Each S_k is fully described by its boundary, hence by a finite circular sequence of s_k vertices. It may be observed that s_k may only be 1 if S_k is a point, 2 if S_k is a line segment, and will usually be at least 3, as soon as S_k has nonvoid interior. Let $s = \sum_{k=1}^{p} s_k$ be the total number of vertices of S.

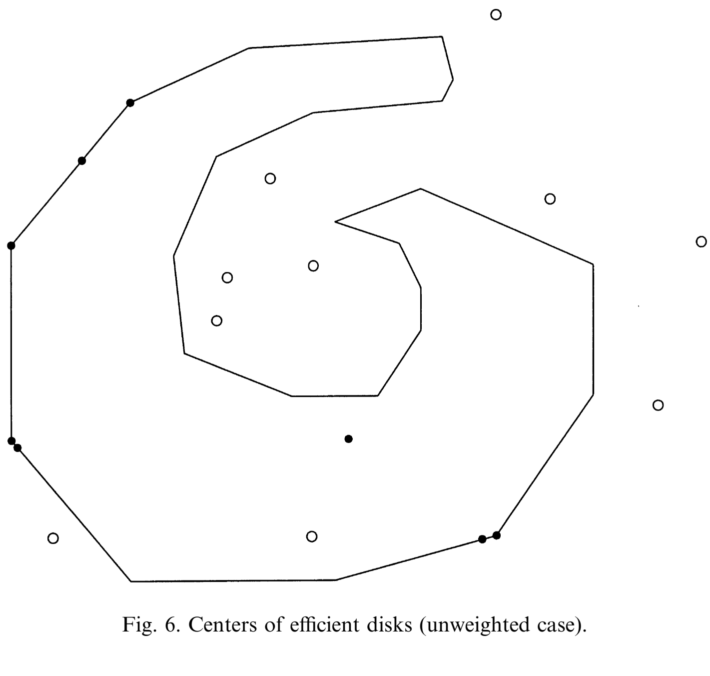

Fig. 6. Centers of efficient disks (unweighted case).

Table 4
Efficient disks (unweighted case)

x	r	$\mathrm{cov}(x, r)$
(40.891, 21.074)	19.221	0/10
(33.818, 4.018)	30.166	1/10
(41.627, 6.100)	32.980	2/10
(26.542, 3)	37.087	3/10
(11, 3)	39.217	4/10
(11, 3)	49.820	5/10
(0, 16)	54.626	6/10
(0.547, 15.353)	59.564	7/10
(0, 34)	61.847	8/10
(0, 16)	66.483	9/10

Where possible we will indicate the complexity of our algorithms in terms of both n, the number of possibly affected points, and s or another indicator of the input size of S. When only n appears in the complexity it means that the constraint set S is taken as fixed.

7.1. Generation and evaluation of C_1

For each $a \in A$ we must construct $C(a)$. By Lemma 5(ii) only 'local extreme' points have to be checked, which for a polygonal region are always 'convex' vertices, i.e. vertices with an inside angle of less than π. Let s_c denote the number of convex vertices of S. Then $s_c = s$ iff all S_k are convex, while for general shapes s_c may be much smaller than s, as happens, e.g., when S is given by a convex region intersected with the complement of a convex polygon with a very large number of vertices. The list Γ of all convex vertices of S may evidently be constructed in $O(s)$ time by moving sequentially along S's boundary and checking each vertex for convexity. Once this list obtained, the construction of each $C(a)$ may be done in $O(s_c)$ time, by checking for each vertex in Γ in $O(1)$ time whether it satisfies the additional condition in Lemma 5(iii). This yields C_1 in an overall $O(s + n.s_c)$. Calculating the coverage for each of the $O(n.s_c)$ disks in C_1 may thus be done in $O(n^2.s_c)$ overall time.

Although in general the number of vertices in $C(a)$ may be of $O(s_c)$, for many a it will be much smaller, due to the additional condition in Lemma 5(iii). In fact the separate vertex checks may be avoided by organising Γ as a circular list, sorted counterclockwise on the normal of their entering edge (the 'preceding normal'). The normal cone at some convex vertex is the cone generated by its preceding normal and the normal of the leaving edge (the 'succeeding normal'). In order to find, for a fixed $a \in A$ all convex vertices for which $x - a$ is in their normal cone, one may first determine by binary search along Γ the 'first' convex vertex with preceding normal less than $x - a$ (this takes $O(\log s_c)$ similar to Theorem 2.2 in Preparata and

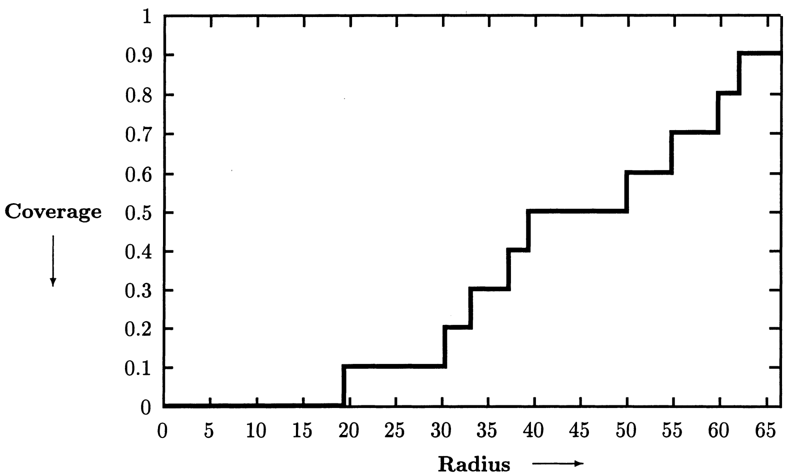

Fig. 7. Trade-off curve (unweighted case).

Shamos (1985, p. 43)), and then read the vertices sequentially from Γ as long as the succeeding normal remains counterclockwise higher than $x - a$. In this way the amount of 'unproductive' work has been reduced to $O(\log s_c)$ for each $a \in A$, at the expense of an extra $O(s_c \log s_c)$ for sorting list Γ.

7.2. Generation and evaluation of C_2

For each pair $\{a, b\} \subset A$ $(a \neq b)$ the following steps must be taken in order to generate $C(a, b)$:

1. Determine which edges of S intersect $\mathrm{med}(a, b)$.
2. For each such edge e:
   (a) Calculate x_e, the intersection point of e and $\mathrm{med}(a, b)$.
   (a) Check whether x_e is remote from $\frac{1}{2}(a + b)$, and if not remove e from further consideration.
   (c) Check the condition in Lemma 6(iii).

Step 1 may be naively done in $O(s)$ time, just checking for each vertex on which side of $\mathrm{med}(a, b)$ it lies. Since this must be done for $O(n^2)$ different lines $\mathrm{med}(a, b)$, it might be worthwhile to consider using some $O(\log s)$ technique, such as a modification of the method described in Edelsbrunner (1987 p. 373, in order to generate the intersected set of segments instead of just the count of them. The preprocessing time and extra data storage necessary may however be too important for practice.

The number of edges produced by step 1 is difficult to predict. Usually it will be quite small compared to s, but in the worst case it may be $O(s)$. Therefore the number of repetitions of step 2 will be $O(s)$.

Step 2 may be carried out in constant time for each fixed edge e. Here the outer normals q_e calculated for each edge e during the generation of C_1 will be very useful. When x_e is an inner point of e, 2(b) reduces to checking whether $\langle\, x_e - 1/2(a + b)\ ;\ q_e \,\rangle > 0$, while 2(c) corresponds to checking whether q_e strictly lies in the convex cone generated by $x_e - a$ and $x_e - b$. In the very exceptional and degenerate case where x_e is a vertex, the condition of Lemma 6(ii) says that x_e is locally farthest from a or b, hence it will already have been found in C_1, and should not be checked again here.

The overall complexity of constructing C_2 will thus be $O(n^2 s)$. A posteriori calculation of the coverage of all these disks would lead to an $O(n^3 s)$ method. Although acceptable it will be shown in the next subsection that both generation and evaluation of C_2 may better be merged together with the same steps for C_3..2

7.3. Generation and evaluation of C_3 (and C_2)

We must generate all triplets $\{a, b, c\} \subset A$ which form an acute triangle, the circumcentre of which lies in S, and calculate their coverage. We will do this by considering in turn each pair $\{a, b\}$, each time handling all additional points c.

Consider therefore a fixed pair $\{a, b\} \subset A$ $(a \neq b)$ and call their distance $\delta = \|a - b\|$. For any $c \in A \setminus \{a, b\}$ the circumcentre x_c of $\{a, b, c\}$ lies on the mediatrix $\mathrm{med}(a, b)$ while the circumradius r_c is an increasing function of the distance from x_c to the midpoint $m = \frac{1}{2}(a + b)$ between a and b; indeed $r_c = (\delta^2/4 + \|m - x_c\|^2)^{1/2}$. For any x on $\mathrm{med}(a, b)$ denote by $C(x)$ the open disk with center x whose boundary passes through a and b, i.e. $C(x) = iB(x, \|x - b\|)$. Note that $C(x_c) = iB(x_c, r_c)$ and its boundary contains $\{a, b, c\}$.

The line ℓ through a and b defines two closed halfplanes which we will arbitrarily call the upper halfplane Π^+ and the lower halfplane Π^-. Similarly $\mathrm{med}(a, b)$ is split into two corresponding closed halflines Λ^+ and Λ^- with origin m, which we consider both (oppositely) strictly ordered (denoted by $\ll$) starting from m. Fig. 8 illustrates the following easy properties, the proof of which are left to the reader.

Lemma 10. *For any two points $x, y \in \Lambda^+$ with $x \ll y$ we have*

$$\Pi^+ \cap C(x) \subset \Pi^+ \cap C(y) \text{ and}$$
$$\Pi^- \cap C(x) \supset \Pi^- \cap C(y).$$

Lemma 11.

For $c \in \Pi^-$ we have $c \in C(m)$ iff $x_c \in \Lambda^+ \setminus \{m\}$.
For $c \in \Pi^+$ we have $c \in C(m)$ iff $x_c \in \Lambda^- \setminus \{m\}$.

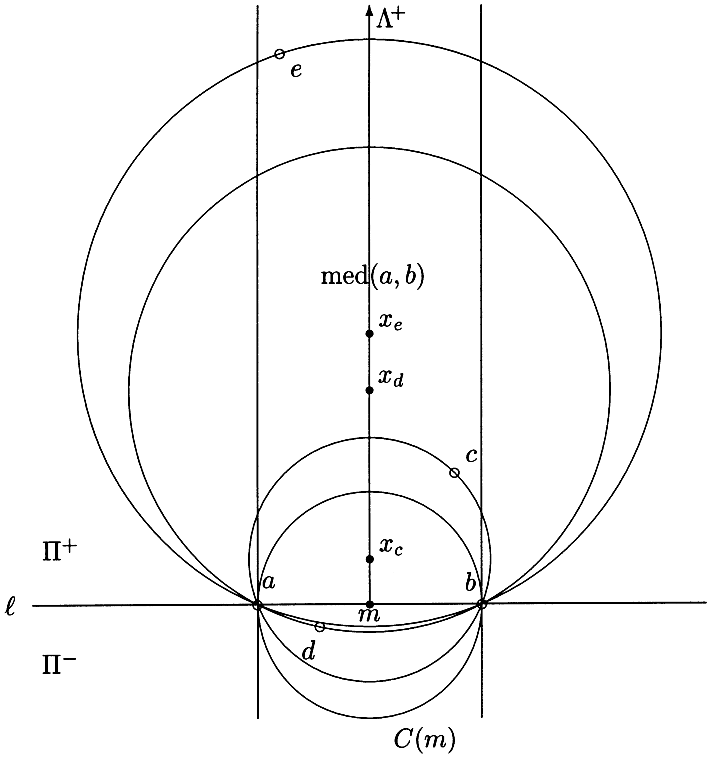

Fig. 8. $x_c \ll x_d \ll x_e$.

Lemma 12. *For $x \in \Lambda^+$ we have*:
1. *for $c \in \Pi^+$ we have $c \in C(x)$ iff $x_c \ll x$*,
2. *for $c \in \Pi^-$ we have $c \in C(x)$ iff $x_c \in \Lambda^+$ and $x \ll x_c$ (Note that by previous lemma this can only happen when $c \in C(m)$).*

By symmetry a similar property is obtained for $x \in \Lambda^-$ by exchanging $+$ and $-$ all over.

These properties allow us to decide whether a point c is covered by some disk $C(x)$ just by comparing the positions of x and x_c along $\text{med}(a,b)$, i.e. according to $\ll$. Hence we will generate all points x_c for $c \in A \setminus \{a,b\}$, and it will be very useful to split them into two lists $\Gamma^+ \subset \Lambda^+$ and $\Gamma^- \subset \Lambda^-$, both ordered by increasing distance from m.

Many of these x_c will also generate candidates, but not all. According to Lemma 7(iii) only triplets forming acute triangles are to be considered. This is easily checked for each c while generating x_c as follows:

- when $c \in C(m)$ triangle $\Delta(a,b,c)$ is always obtuse at c, thus x_c is not a candidate.
- if c and x_c lie on the same side of ℓ (note that then $c \notin C(m)$ by Lemma 11) triangle $\Delta(a,b,c)$ will be obtuse iff c lies outside the open band built orthogonally upon segment $[a,b]$ iff $\langle\, c-a\ ;\ b-a \,\rangle \notin]0,\delta^2[$

In all these cases x_c will be marked as ruled out.

The second condition to be checked on x_c is of course feasibility. This might be done separately for each x_c as part of its generation, but a much better method, with the additional boon of handling the candidate list C_2 simultaneously goes as follows.

First find all points of intersection of $\text{med}(a,b)$ with the boundary of S, yielding $O(s)$ points x_e as described in Section 7.2. (Please note that when S_k has empty interior each intersection point with the boundary of S_k will be generated twice, since the intersected edge appears twice in S's boundary description.)

A point x moving along $\text{med}(a,b)$ will change its status of belonging to S or not, each time x crosses some x_e. Since S is bounded, any x far enough from m in any direction along $\text{med}(a,b)$ will be outside S. The status of m may therefore be obtained by counting the number of points $x_e \in \Lambda^+$, say, and then decide that $m \in S$ iff this number is odd.

The arguments above are correct in the generic situation, where $\text{med}(a,b)$ always simply cuts the boundary. Some degeneracies may however appear.

- First it could happen that $\text{med}(a,b)$ only touches the boundary: the corresponding intersection point x_e is then a vertex. Such an x_e might be a candidate, but the status of x should not change when crossing this x_e; therefore we will call this an *inessential* x_e, while the other x_e are termed *essential*.
- Second $\text{med}(a,b)$ might have a nontrivial segment in common with S's boundary, the endpoints of which yield two points x_e. As explained before only the one farthest from m should be considered as a possible candidate. Normally the segment is part of the boundary of S, hence no points x_c on such a segment should be candidates since they are not interior points of S. Therefore both endpoints may be considered inessential x_e. If however S has a strange shape, it might be that it is described by the union of several pieces, each with its own boundary, but possibly overlapping ones; in such a case it should be checked whether the segment really is part of S's global boundary in order to determine whether each endpoint x_e is essential or not.

We propose to include all the points x_e in the sorted lists Λ^+ and Λ^-, thereby determining whether m is IN or OUT of S. We also calculate as COV the coverage of $C(m)$. This variable will dynamically hold the coverage value of the disk $C(x)$ corresponding to the current x. Then each of these lists will be linearly scanned in the order given by $\ll$, initialising the status (IN or OUT) of the current point to that of m and COV to that of $C(m)$.

At each move to the next x in the list the following steps are taken (here described for Λ^+).

1. If $x = x_e$ then:
   - if status is IN, check the condition of Lemma 6(ii) on x_e. If this is met, store x_e as a candidate for C_2 together with its radius $r = \|a - x_e\|$ and its coverage COV.
   - if x_e is essential then invert the status.
2. If $x = x_c$ and $c \in \Pi^-$, then subtract $w(c)$ from COV by application of Lemma 12(ii). (No candidate can be obtained here because x_c was ruled out, the triangle $\Delta(a,b,c)$ being obtuse.
3. If $x = x_c$ and $c \in \Pi^+$, then:
   - if x_c was not ruled out and the status is IN, then store x_c as a candidate for C_3 together with its radius $r = \|a - x_c\|$ and its coverage COV;
   - anyway, increase COV by $w(c)$.

Two reasons may exist for ruling out some x_c in this last case.

- The first was mentioned above and corresponded to an obtuse triangle.
- The second is connected to the fact that, by a simple repetition of the above over all pairs $\{a,b\}$, each triangle will be checked up to three times. This is avoided using some simple numbering rule for the points of A. Generating the pairs in lexicographic order then allows one to rule out (as already checked) all x_c for c's numbered lower than either a or b.

Building the sorted lists Γ^+ and Γ^- takes $O(n \log n)$ time for each pair $\{a,b\}$ and the further processing explained above is of $O(n)$. Since $O(n^2)$ pairs have to be processed this way, the full generation and evaluation of C_2 and C_3 takes $O(n^3 \log n)$.

Putting together all complexity results of Section 5 and the Sections 7.1 and 7.3 above, we obtain

Theorem 13. *The complete trade-off curve between the coverage and radius of circular disks centered in* S *may be generated in* $O(n^3 \log n)$ *time.*

Theorem 13 gives complexity bounds for solving the two standard single-objective problems associated with (BP), as presented in the following corollaries.

Corollary 14. *Each minimal covering problem can be solved in* $O(\log n)$ *time with a preprocessing time of* $O(n^3 \log n)$.

Proof. For given R, one just needs to identify from the trade-off curve the efficient disk with smallest radius of at least R. Since this curve is defined by the list of $O(n^3)$ efficient disks, which is sorted on radius, this may be done by a binary search. The result then follows from the fact that this task takes $O(\log(n^3)) = O(\log n)$ time. □

Remark that, for a single value for the radius R, our procedure requires in total $O(n^3 \log n)$ time, while the algorithm of Drezner and Wesolowsky (1994) is in $O(n \log n)$. However, for repeated trials of R (as will be rule in any interactive procedure), Drezner and Wesolowsky procedure must start each time from scratch (thus consuming again $O(n \log n)$ time), (it is not known in advance which particular values of R are critical) while our procedure will only consume $O(\log n)$ time, thus enabling the execution of an interactive procedure in real time.

With a similar reasoning for a given coverage one obtains

Corollary 15. *Each maxquantile or largest circle problem can be solved in* $O(\log n)$ *time with a preprocessing time of* $O(n^3 \log n)$.

As will be shown in Section 7.4, better complexity results can be obtained for the unweighted case.

7.4. An improved complexity method for the unweighted case

We show now that a method of $O(n^3)$ exists for generation and evaluation of all candidate disks. This means that the final complexity of $O(n^3 \log n)$ still remains, but only due to the sorting of the candidate list in view of dominated disk deletion, a task which is well-understood and quite efficient in practice. For unweighted problems the overall complexity is then reduced to $O(n^3)$, since the sorting step is unnecessary, which implies in particular that both minimal covering and largest circle problems can be solved, in the unweighted case, in $O(\log n)$ time after an $O(n^3)$ preprocessing time.

As the naive generation of C_1 and C_2 and their coverage values has complexity $O(n^3)$ let us take a different look at the generation and evaluation of C_3. In fact we might as well generate all triplets $B = \{a, b, c\} \subset A$ together with their corresponding disk (x_B, r_B) and coverage $\mathrm{cov}(x_B, r_B)$, where x_B is the circumcentre of B and r_B the circumradius $d_B(x_B)$, if this task may be done efficiently. This will just increase the candidate list, without increasing its $O(n^3)$ length.

For each $B = \{a, b, c\} \subset A$ its circumcentre x_B is a Voronoi vertex in a Voronoi diagram of A of an appropriate order (see Preparata and Shamos, 1985, p. 237). Indeed, for $k = | A(x_B, r_B) |$ the point x_B will be the Voronoi vertex in the order $k+1$ Voronoi diagram of A, at the intersection (in the nondegenerate case where no fourth point is active) of the Voronoi faces corresponding to the subsets of $k+1$ closest points $A(x_B, r_B) \cup \{a\}$, $A(x_B, r_B) \cup \{b\}$ and $A(x_B, r_B) \cup \{c\}$. In fact x_B will also be a vertex on the order $k+2$ Voronoi diagram on the faces of closest points $A(x_B, r_B) \cup \{a, b\}$, $A(x_B, r_B) \cup \{b, c\}$ and $A(x_B, r_B) \cup \{a, c\}$, but this is of no importance to us here.

Our task thus boils down to that of constructing all Voronoi vertices of all orders, hence the Voronoi diagrams of all orders $k = 1, \ldots, n-1$. This task may be done in $O(n^3)$ as explained in Edelsbrunner (1987), and includes the determination for each face of the corresponding set of k closest points. For any vertex x, after checking $x \in S$ in $O(1)$ time, the corresponding set B is obtained by xoring the closest point sets of all three adjacent Voronoi faces. This yields the radius $r_B = d_B(x)$, and the coverage $\mathrm{cov}(x, r_B)$ is the total weight of the intersection of the closest point sets of the adjacent Voronoi faces. All these extra items may also be generated in the course of Edelsbrunner's algorithm.

Note. A. Tamir (1995) suggested to us that even for weighted problems the dominated disk deletion step might be merged with Edelsbrunner's algorithm without increase in complexity, thus yielding an optimal $O(n^3)$ method. This remains, however, an open question.

7.5. Location on a network

In practice the location of the undesirable facility will often be restricted to some existing network in order to be reachable, see e.g. Karkazis and Boffey (1994). Although the general case treated in the previous subsections applies to S being a network, it is interesting to emphasize the fact that in this perhaps at first glance more complicated situation, one obtains in fact a slightly lower complexity by way of the naive basic method.

Indeed, when S is part of a physical network, such as a road, rail and/or waterway network, it may be adequately described by a finite union of linear line segments. Note that since we do not consider an abstract network, but one explicitly embedded in the plane, the edges usually considered will often have to be modeled as unions of several straight line pieces in order to accommodate bends. Even higher precision of the description may be obtained by using elliptical arc pieces by spline approximation, see Ueberhuber (1995) and the references therein.

As such S has no interior points, meaning we may forget about the three active point case described in Lemma 7. Therefore only the generation and evaluation of C_1 and C_2 remain, which were shown in Sections 7.1 and 7.2 to be implementable in a naive way in $O(n^3)$ time. The candidate list produced is then of length $O(n^2)$, hence dominated disk deletion takes only $O(n^2 \log n)$ time, leading to an overall $O(n^3)$ complexity. Observe that the geometrical method developed in Section 7.3 would not lead to a lower complexity since all $O(n^3)$ points x_c would still have to be generated in order to determine the coverage!

8. Extensions and suggestions for further research

8.1. Largest almost empty circles

The method of Section 7.4 may be considered to be an extension of the way Toussaint (1983) handles the largest empty circle problem. One particular order k-Voronoi diagram is not of direct use in our problem, except in the unweighted case – then it may serve to find the largest disk covering exactly $k-1$ points. For the determination of the largest circle covering at most a given weight W of points, one would need an as yet undocumented extension of the order-k Voronoi diagram concept, which we would like to call a knapsack–Voronoi diagram.

Given a finite set of points A in the plane, with corresponding weights w_a, for a fixed W, the *W-knapsack–Voronoi diagram* is defined by the regions of all points x for which the set of closest points of A up to a total weight of W (the weight of the last one possibly less than its full weight) is fixed. The name stems from the analogy with continuous knapsack problems. We feel that this type of diagram has an interest of its own, and possible applications in other fields, in particular for Weber problems with supply surplus as studied in Kaufman and Plastria (1988).

Using currently available techniques, however, it will be necessary to check all order k-Voronoi diagrams for $k = 1, \ldots, K$, where K is chosen in such a way that any subset B of A with $w(B) \leqslant W$ has cardinality less than K. The ideal K is found by solving the knapsack problem $\max\{|B| \mid w(B) \leqslant W\}$, which is easily done by sorting A nondecreasingly and count how many weights taken in this order still sum up to no more than W, or even quicker, without sorting using the linear time method described by Balas and Zemel (1980). Probably the most efficient way to generate all order k-Voronoi diagrams for $k = 1, \ldots, K$ is to use the incremental method, which may be done in $O(K^2 n \log n)$ time, see Preparata and Shamos (1985, p. 244). For small K this may have an even better complexity than the $O(n^3)$ obtained in Section 7.

Checking all order k-Voronoi diagrams for $k = 1, \ldots, K$ in fact generates all information necessary in order to find all efficient disks up to coverage W. Such partial enumeration may be important in practice. Indeed in the context of undesirable facility location the covered weight should normally be kept as small as possible, and certainly below some given threshold, e.g. one which is politically affordable.

8.2. Other feasible regions

The geometrical method of Section 7 should extend quite naturally to nonpolygonal feasible regions.

Two types of special constraints may be suggested here, both leading to feasible regions of a shape satisfying Assumption 8 in Section 5.

The first concerns the presence around the facility to be located of a 'high danger' zone of given radius R, within which no population should lie. The feasible region may then be described as the intersection of some, usually polygonal, region S_p with the complement of all disks $B(a, R)$ for $a \in A$

$$S = S_p \cap \bigcap_{a \in A} (\mathbb{R}^2 \setminus B(a, R)).$$

This means that the supplementary part of S's boundary due to the disk complements consists of arcs of circles, the outer normals of which point towards S's interior. According to Lemma 5 only endpoints of such arcs may remain in C_1, while no point internal to such an arc will satisfy the conditions of Lemma 6, and will thus not have to be retained in C_2.

The other type of additional constraint concerns reachability and is somewhat opposite to the previous one, but might be used in conjunction with it. As suggested by Drezner and Wesolowsky (1980) in connection with largest empty circle problems, the facility should usually be located at some point which is sufficiently reachable from a certain number of given points, often the same ones which should be protected from it. Therefore these authors propose as feasible region the intersection of a set of disks centered at each $a \in A$ and of given, possibly different radii. We suggest extending this towards a more general destination set A' (which may have points in common with A) and accept other distance measures d_b up to point $b \in A'$ than the Euclidean one (e.g., polyhedral, see Plastria, 1995) for the simple reason that here we consider proximity possibly connected with transportation. The feasible region will thus be further restricted to lie within the intersection of a given number of convex sets, the balls for these distance measures d_b. The extra complication these assumptions may imply is the appearance of new pieces of boundary of possibly nonlinear shape, the endpoints of which are new convex vertices. These will of course to be taken into account for the construction of C_1, but also nonvertex points satisfying Lemma 5 may now appear and should be determined. In case $b \in A \cap A'$ and d_b is Euclidean the degenerate special case discussed in Section 5 may appear, possibly calling for the use of a linear search. The whole approach may now be viewed as an extension of the proposal of Rangan and Govindan (1992) for the largest empty circle problem.

8.3. Inflated Euclidean distances

The maxmin model studied by Drezner and Wesolowsky (1980) was in fact slightly more general: distances up to each point in A were measured by *inflated Euclidean distances* $\lambda_a \|a - x\|$ (note that these authors call these inflation factors the weights, which would be rather confusing in our context). Although the use of this generalization for modeling undesirable facility location is disputable (and we do not advocate it: why should the risk to two different centers at the same Euclidean distance be different?), it may be handled in almost the same way as the normal case. Indeed the only difference appearing is that $\mathrm{med}(a, b)$ is no longer a straight line, but a circle (the Appolonius circle). The basic algorithm and its improved geometric version certainly apply, but the Voronoi diagram machinery may also be put into play, provided its multiplicative weighted version is used, see Okabe et al. (1992, p. 129 et seq), although the subject of order k multiplicatively weighted Voronoi diagrams and their constructions seem not to be documented as yet. This would extend the work of Melachrinoudis and MacGregor Smith (1995).

8.4. Other planar distance measures

Distance measures other than the Euclidean one might be investigated. Both Lemmas 3 and 4 carry through without any changes. It is not hard to see that if a disk (x, r) is efficient, then x is a global maximum in S of the minimal distance function $d_{U(x,r)}$ to $U(x, r) = A \setminus A(x, r)$, the set of points of A not covered by $iB(x, r)$. In other words x is then a global solution to some maxmin problem with directional differentiable objective, for which many results are known, e.g. in Chapter 3 of Dem'yanov and Rubinov (1986). These lead to conclusions similar to those in Lemmas 5–7, and in principle the ideas of the basic algorithm should carry through.

However, we consider that in the realm of undesirable facility location the distance measures of interest are the Euclidean, and its modifications due to the presence of winds. For the modeling of airborne pollution it seems that an ellipsoidal gauge, i.e. the Minkowsky distance generated by an ellipse with 'centerpoint' possibly differing from its center of symmetry, might be of interest. As long as this centerpoint remains inside the ellipse we obtain *skewed norms* as defined in Plastria (1992a). If the centerpoint moves out of the ellipse this does not correspond to a usual distance measure anymore. However, the level sets (above a certain level) of the Gaussian plume model (see e.g. Karkazis and Papadimitriou, 1992) for dispersion of airborne pollutants may be approximated by an ellipse for which the point of emission lies outside. Therefore even such a case might be applicable.

Let us therefore consider our model, but replacing the circular disks by such ellipses. More precisely, let T be a regular linear transformation of $\mathbb{R}^2$, and q any vector in $\mathbb{R}^2$. Denote by U the standard Euclidean unit disk, hence $B(x,r) = x + rU$. We assume now that we have as new unit disk the set $C = T(U) + q$, which is an ellipse with symmetry center at the point q. The disk of radius r and center x for this new distance measure is then given by $C(x,r) = x + rC$, which is an ellipse, but the symmetry center of which lies now at $x + rq$.

Our naive algorithm may be adapted as follows to this new situation in the case of polygonal regions. First, for each given point all boundary vertices should be tested for being locally farthest feasible points. Second, for each pair of given points the mediatrix – which is now a branch of a hyperbola in general – should be intersected with the boundary of the feasible region. Third, for each triplet of points the circumellipse should be constructed, thereby determining the radius r, and then it should be checked whether the translate over $-rq$ of its symmetry center is feasible. In this way the candidate solution list is constructed, which should still be evaluated and weeded. All this does not look like a very difficult task. It requires some research however in order to find an efficient way of carrying it out.

8.5. Affected regions

Another very important extension in practice is to consider affected regions instead of affected points. Since most of our analysis and methods heavily relied on the assumption of discrete affected points, it is not clear at all how to extend the methods presented here. In view of the evident practical interest of such models they certainly deserve attention by the research community.

Acknowledgements

The authors thank A. Tamir first, for suggesting that an $O(n^3 \log n)$ method might exist, which led us to developing the geometric approaches in Section 7, and second, for his later conjecture stated in Section 7.4 about an even better complexity technique.

References

Balas, E., Zemel, E., 1980. An algorithm for large zero–one knapsack problems. Operations Research 28, 1130–1145.

Carrizosa, E., Plastria, F., 1995. On minquantile and maxcovering optimisation. Mathematical Programming 71, 101–112.

Carrizosa, E., Plastria, F., 1999. Polynomial algorithms for parametric minquantile and maxcovering planar location problems with locational constraints, TOP, to appear.

Chankong, V., Haimes, Y.Y., 1983. Multiobjective Decision Making: Theory and Methodology, North-Holland, Amsterdam.

Dem'yanov, V., Rubinov, A.M., 1986. Quasidifferential calculus. Optimization Software.

Drezner, Z., Wesolowsky, G., 1980. A maxmin location problem with maximum distance constraints. AIIE Transactions 12, 249–252.

Drezner, Z., Wesolowsky, G., 1994. Finding the circle or rectangle containing the minimum weight of points. Location Science 2, 83–90.

Edelsbrunner, H., 1987. Algorithms in Combinatorial Geometry. Springer, Berlin.

Erkut, E., Neuman, S., 1989. Analytical models for locating undesirable facilities. European Journal of Operational Research 40, 275–291.

Hamacher, H.W., Nickel, S., 1995. Restricted planar location problems and applications. Naval Research Logistics 42, 967–992.

Hansen, P., Peeters, D., Thisse, J.-F., 1981. On the location of an obnoxious facility. Sistemi Urbani 3, 299–317.

Hiriart-Urruty, J.B., Lemarechal, C., 1993. Convex Analysis and Minimization Algorithms. Springer, Berlin.

Karkazis, J., Boffey, B., 1994. Modelling pollution spread. Studies in Locational Analysis 7, 91–104.

Karkazis, J., Papadimitriou, C., 1992. Optimal location of facilities causing atmospheric pollution. European Journal of Operational Research 58, 363–373.

Kaufman, L., Plastria, F., 1988. The Weber problem with supply surplus. Belgian Journal of Operations Research Statistics and Computer Science 28, 15–31.

Kleindorfer, P.R., Kunreuther, H.C., 1994. Siting of hazardous facilities. In: Pollock, S.M., et al. (Eds.), Operations Research and the Public Sector. Handbooks in OR&MS, vol. 6. Elsevier, Amsterdam, pp. 403–440.

Melachrinoudis, E., Cullinane, T.P., 1986. Locating an undesirable facility with a minimax criterion. European Journal of Operational Research 24, 239–246.

Melachrinoudis, E., Macgregor Smith, J., 1995. An $O(mn^2)$ algorithm for the maximin problem in E^2. Operations Research Letters 18, 25–30.

Okabe, A., Boots, B., Sugihara, K., 1992. Spatial Tesselations: Concepts and Applications of Voronoi Diagrams. Wiley, New York.

Plastria, F., 1992a. On destination optimality in asymmetric distance Fermat–Weber problems. Annals of Operations Research 40, 355–369.

Plastria, F., 1992b. The generalized big square small square method (GBSSS) for planar single facility location. European Journal of Operational Research 62, 163–174.

Plastria, F., 1995. Continuous location problems. In: Drezner, Z. (Ed.), Facility Location: A Survey of Applications and Methods. Springer, Berlin, pp. 225–262.

Plastria, F., 1996. Optimal location of undesirable facilities: a selective overview. JORBEL 36, 109–127.

Preparata, F.P., Shamos, M.I., 1985. Computational Geometry: An Introduction. Springer, Berlin.

Rangan, C.P., Govindan, R., 1992. An $O(n \log n)$ algorithm for a maxmin location problem. Discrete Applied Mathematics 36, 203–205.

Tamir, A., 1995. Private communication.

Toussaint, G.T., 1983. Computing largest empty circles with location constraints. International Journal of Computers and Information Science 12, 347–358.

Ueberhuber, C., 1995. Computer-Numerik, vols. 1-2 (in German). Springer, Berlin.

Vincke, P., Gassner, M., Roy, B., 1992. Multicriteria Decision-aid. Wiley, New York.