ANZIAM J. 48(2007), 315–325

FACILITY LOCATION VIA CONTINUOUS OPTIMIZATION WITH DISCONTINUOUS OBJECTIVE FUNCTIONS

J. UGON[✉ 1], S. KOUHBOR[1], M. MAMMADOV[1], A. RUBINOV[12] and A. KRUGER[1]

(Received 13 September, 2006; revised 21 November, 2006)

Abstract

Facility location problems are one of the most common applications of optimization methods. Continuous formulations are usually more accurate, but often result in complex problems that cannot be solved using traditional optimization methods. This paper examines the use of a global optimization method—AGOP—for solving location problems where the objective function is discontinuous. This approach is motivated by a real-world application in wireless networks design.

2000 *Mathematics subject classification*: primary 90B85; secondary 90B18.
Keywords and phrases: global optimization, discontinuous optimization, location problems.

1. Introduction

Location problems play an important part in optimization, as they have a very broad area of practical applications: [4] lists over 3400 references on facility location and related problems.

As a rule, location problems are generally tackled using combinatorial optimization, with the set of possible locations for facilities having to be finite. The paper [16] reviews this approach to facility location problems. Its complexity increases when the number of possible locations increases, leading to the development of approximation algorithms.

Many applications, however, do not require such a restriction on the placement of the facilities: these only have to be placed over a given area, not at special locations. This problem configuration occurs for example in telecommunications [15, 17], data analysis [2] and public transportation [3] problems. The classical approach to solve

[1]School of Information and Mathematical Sciences, University of Ballarat, Vic 3353, Australia.
Corresponding author's e-mail: j.ugon@ballarat.edu.au.
[2]Deceased.

https://doi.org/10.1017/S1446181100003515 Published online by Cambridge University Press

these problems is to transform them into combinatorial problems by discretizing the search space [1, 20]. Due to the NP-hardness of the combinatorial location problems, a solvable discretization may result in inaccurate results.

An alternative approach is to formulate the problem as a continuous optimization problem. With such an approach, the complexity of the problem depends on the properties of the cost function, and on the shape of the search space. Often, the cost is represented by a min-type function, which prevents the problem from being convex or smooth.

The aim of this paper is to present an algorithm for solving continuous location problems with discontinuous functions directly. For that purpose we use a modified version of the appropriate global optimization method AGOP developed in [9] and [10].

Our main motivation is an application in the field of optimal wireless networks design. This problem is presented as a case study, and we carry out numerical experiments based on real specifications.

The problem is formulated in Section 2. A modification of AGOP is presented in Section 3. Section 4 is devoted to a case study of a practical application. Sections 5 and 6 present respectively numerical results and the conclusion.

2. Location problem

In general, locating facilities involves several steps:

(1) Find the number of facilities necessary, subject to a given constraint (demand satisfaction).

(2) Find the most suitable placement for these facilities, considering one or several criteria.

In this paper, we focus on the first step. An important part of it consists in verifying the constraint for different numbers of facilities. The constraint verification problem is in itself a facility placement problem.

The problem considered in the current paper can be stated as follows: given a set of customers, find the minimal number of facilities, and their satisfactory placement for covering the demand of the customers.

This problem can be expressed as follows:

$$\text{minimize} \quad n \tag{2.1}$$

$$\text{subject to} \quad \exists [x_i]_1^n : g_j(x_1, \cdots, x_n) \leq 0, \forall j \in \{1, \ldots, J\}, \tag{2.2}$$

where $(x_1, \ldots, x_n) \in \mathbb{R}^{n \times m}$ and $g_j : \mathbb{R}^{n \times m} \to \mathbb{R}$ for all $j \in \{1, \ldots, J\}$. Here J represents the number of customers and m represents the dimension of the geographical area. The inequalities (2.2) represent the constraints verifying that all the demand area is covered. The functions g_j may be discontinuous.

In what follows it is assumed that the solution exists for a relatively small value of n, and therefore an enumeration method can be considered. On the other hand, the problem of finding a feasible solution satisfying (2.2) is a difficult one, due to the nature of the functions g_j.

The discontinuity of the functions g_j may arise in many practical situations: if the area contains obstacles, the service as a function of the distance is likely to be discontinuous.

In the telecommunication problem considered further in this paper, each function g_j corresponds to one user, and measures the coverage of this user. Below a certain threshold, the user is not covered, and the constraint (2.2) is not satisfied.

Besides telecommunications, the model has applications in data analysis. It can also be applied in other covering-type problems, where obstacles have to be taken into account.

Due to the assumption made, problem (2.1)–(2.2) reduces to the feasibility problem: for a fixed n find $[x_i]_1^n$ satisfying (2.2). In turn, the system of inequalities (2.2) can be satisfied if and only if the optimal value of the optimization problem

$$\text{minimize} \quad \sum_{j=1}^{J} \max\left(0, g_j(x_1, \cdots, x_n)\right) \tag{2.3}$$

is zero.

The objective function of this problem is discontinuous, since the functions g_j are. Therefore very few optimization methods can be applied here. Besides, due to the min-type nature of functions g_j, the objective function in (2.3) usually has a large number of local minima. As a result, most methods that are theoretically applicable to such problems will not be successful in solving this problem. Note also that a single evaluation of the objective function can be computationally expensive.

3. Solving the problem

The main problem (2.1)–(2.2) can be solved by solving a sequence of global optimization problems (2.3). We will apply the following scheme:

Main Scheme:

Step 1. Set $n = 1$.

Step 2. Solve the problem (2.3).

Step 3. If the value of the objective function is zero, then *Stop*. Otherwise set $n = n + 1$ and go to Step 2.

Note that the above three-step procedure is just an illustration. When a better lower estimate of the number of APs is known (or can be easily determined) this number can be taken instead of $n = 1$ in Step 1. Similarly, when a "good" upper estimate is known, a binary search algorithm can be used. However, one should keep in mind that to do this one needs to know a really tight upper estimate. Otherwise, due to the computational complexity of the problems, which increases exponentially, the application of the binary search algorithm can be more time consuming than the simple Main Scheme described above.

3.1. Overview of possible solvers The choice of the solver is crucial.

Although the initial problem (2.1)–(2.2) is an integer programming problem, for each n, to verify the feasibilty, it is necessary to solve a continuous optimization problem (2.3).

The objective function of (2.3) is discontinuous. As a result, it is not possible to apply any of the classical methods based on local properties of the functions (such as Newton-based or bundle methods [6, 8] and their derivatives), nor even methods based on Lipschitz continuity (such as branch and bound methods [5]).

To confront the discontinuity of the functions met in telecommunications network designs due to obstacles, a few authors [14, 18, 19] have used genetic algorithms. Due to their heuristic nature, it is quite easy to adapt these methods for solving the type of problem under consideration. However, genetic algorithms are very dependant on the initial population. The complex structure of the functions (these functions have a large number of local minima) results in the necessity of having a large population size. Since even for a simple real-world problem, the evaluation of the objective function is computationally very demanding, it is very impractical to use evolutionary algorithms.

Other heuristic approaches, such as simulated annealing or neural networks, present the same drawback: although these methods can be easily implemented, they would perform poorly or be too slow on the problem at hand.

In contrast to the mentioned approaches, AGOP (Algorithm for Global Optimization Problems, see [9, 10]), considered in the current paper, finds a good solution using a relatively small number of function evaluations. Its operation is explained below.

3.2. Operation of AGOP Consider the problem:

$$\text{minimize} \quad f(x) : \mathbb{R}^n \to \mathbb{R}, \quad \text{s.t. } x \in B, \tag{3.1}$$

where B is a given box constraint. AGOP must first be given a set of points, say $\Omega = x_1, \ldots, x_q \subset \mathbb{R}^n$. Generally, a suitable choice for an initial set of points is generated from the vertices of the box B.

Suppose that $x_\star \in \Omega$ provides the smallest value of the objective function, that is, $f(x_\star) \leq f(x)$ for all $x \in \Omega$. An approach has been developed for finding a descent direction v at the point $x_\star$ (see [9] for details). An inexact line search along this direction provides a new point $\hat{x}_{q+1}$. A local search about $\hat{x}_{q+1}$ is then carried out. This is done using the *local variation* method. This is an efficient local optimization technique that does not explicitly use derivatives and can be applied to nonsmooth functions. A good survey of direct search methods can be found in [7]. Letting x_{q+1} denote the optimal solution of this local search, the set Ω is augmented to include x_{q+1}. Starting with this updated Ω, the whole process can be repeated. The process is terminated when v is approximately 0 or the prescribed bound on the number of iterations is reached. The solution returned is the current x_*, that is, the point in Ω with the smallest cost.

The main part of the algorithm is to determine a possible descent direction v at each iteration. The method used by AGOP for this aim is based on dynamical systems described by non-functional relationships between two scalar variables (see [10] for details). These relationships are defined in terms of influences of the change (increase or decrease) of one variable on the change of the other. The forces acting from one variable on the change of the other variable are defined by the means of influences. This allows us to define a (non-standard) dynamical system, which provides the direction of changes of each variable at any given point. AGOP uses this idea to determine a descent direction v. First, given a set Ω, we define a dynamical system that describes the relationships between the objective function and a particular variable x^i, $i = 1, \dots, n$. This provides a vector $\bar{v} = (\bar{v}^1, \dots, \bar{v}^n)$, calculated at the point x_*, where the coordinate $\bar{v}^i$ is the force acting from x^i on the increase of f. Then the vector $v = -\bar{v}$ is taken as a possible descent direction at the point x_*.

3.3. Modification of AGOP Taking into account the peculiarities of problem (2.3), we introduce the following improvements to the procedure of AGOP.

First, we note that the calculation of the value of the objective function is computationally intensive. Moreover, the value of the objective function is zero if there is a feasible solution to (2.2). Therefore the execution can be stopped once the zero value of the objective function is achieved. In many cases the set of optimizers is quite large, and as a result a function value of zero may be found very early on by the algorithm.

The other improvement is related to the choice of initial sets Ω. In this case we are trying to use the solution obtained in the previous step of the sequence of problems that are being solved during the execution of **Main Scheme**. It also takes into consideration the geographical nature of the problem: a solution x is structured as a set of geographical points $\bar{x} \in \mathbb{R}^m$. The solution reached at iteration $(n-1)$ can be used at iteration n. The set Ω of initial points is constructed as follows: $\Omega = \{x_1, \dots, x_q\}$, where $x_i = (x_0^{*,n-1}, \dots, x_{q-m}^{*,n-1}, y_1, \dots, y_m)$, $x^{*,n-1}$ is the solution reached at iteration $n-1$, and $y \in \mathbb{R}^m$ is an initial point constructed from the boundaries of the geographical area.

This allows us to reduce the initial size of the set Ω, and therefore to accelerate the execution of AGOP. Furthermore, it also generates initial points that are potentially closer to the set of optimizers, which is reached faster.

4. Case study

The model described in Section 2 above is motivated by applications in wireless telecommunications. Consider the following problem: given a building where a wireless network needs to be installed, find the minimal number of antennas (Access Points (APs)) which can cover the total area where users may move.

The building can be divided into three types of areas:

- Areas where APs can be placed and users need to receive (for example offices).
- Areas where APs can be placed, but users do not need to receive (for example, stationery rooms).
- Areas where APs cannot be placed (for example, elevators).

As a result, the area to cover may be quite complex, and the coverage cannot be computed easily. Therefore, this area is discretized: potential users are placed everywhere where a user may need to access the network. This problem can be formulated using (2.1)–(2.2).

The signal emitted by an AP x_i deteriorates before reaching a potential user u_j. This can be measured by the so-called *pathloss* (loss of power between emission and reception, measured in decibels (dB)). Over a certain threshold, the pathloss is too large, and the user cannot receive the signal.

The pathloss can be written as follows [11–13]:

$$p(x_i, u_j) = p_1 + p_2, \tag{4.1}$$

where p_1 and p_2 represent the deterioration of the signal caused by the distance between AP x_i and user u_j and by physical obstacles (such as walls) respectively.

When the user is far enough from the AP, the following formula can be used:

$$p_1(x_i, u_j) = p^0 + 20 \log d(x_i, u_j),$$

where $d(x_i, u_j)$ is the distance between AP x_i and user u_j and p^0 is the pathloss at the reference distance (equal to 1 m). For computational purposes, we will assume that if $d(x_i, u_j) < 1$, the pathloss remains constant (p^0). Since we are interested in ensuring that the user is within reach of the AP, this assumption is reasonable.

The pathloss due to obstacles is calculated as $p_2(x_i, u_j) = \sum_{t=1}^{Q} \delta_t l_t$, where Q is the total number of obstacles, l_t is the loss for crossing obstacle t; $\delta_t = 1$ if the obstacle t is crossed by the signal, and $\delta_t = 0$ otherwise.

In such a case, in the formulation (2.1)–(2.2), we have

$$g_j(x_1, \ldots, x_n) = \min_{1 \le i \le n} p(x_i, u_j) - p_{\max},$$

where $p_{\max}$ represents the maximum allowed pathloss. Above this threshold, the quality of signal is not sufficient.

This particular problem has the following characteristics:

- The number J of functions g_j is equal to the number of users. This number can be quite large, as the final coverage may depend on the density and the distribution of the users in the building.
- The function g_j is the minimum of functions g_j^i, where $1 \le i \le n$. So for each choice of $[x_i]_1^n$ the total number of function evaluations is $n \times J$.
- The functions g_j^i depend on the number of walls separating the user from the access point. When the number of walls in the building is large, evaluating g_j^i is computationally very expensive.

As a result, the evaluation of an objective function is computationally very demanding.

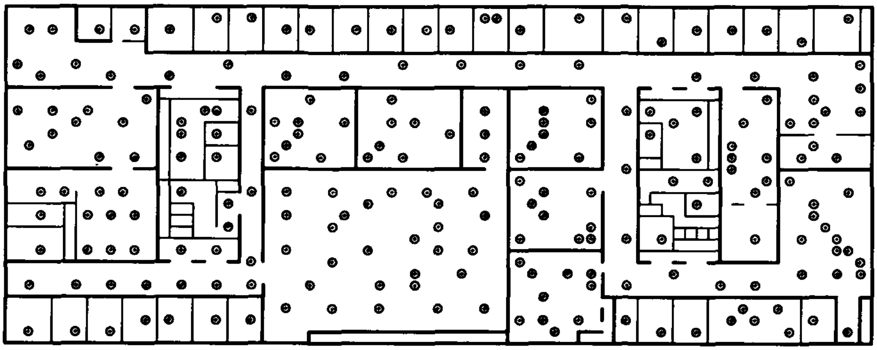

FIGURE 1. Layout of the design area.

5. Testing and results

5.1. Testing Some experiments were carried out in order to verify the effectiveness of the model based on the case study described in Section 4.

Firstly, a real-world situation described in [18] was examined. The building has dimensions 75 m × 30 m and contains 129 walls (obstacles) of pathloss either 3 or 6. In this building, 223 potential users are distributed over the area. Figure 1 shows the building specifications and the distributions of the users.

Another set of experiments with 183 potential users distributed over the area has been conducted. The effect of reducing the number of users is that some rooms no longer contain a potential user. This set of experiments was carried out in order to observe the effect of the distribution of users on the number of APs necessary to cover the area.

The experiments were carried on the VPAC supercomputer "Brecca" [21]. The threshold for the pathloss has been varied. Results are presented in Tables 1 (a) and (b). The following notation is used in the tables: P_{th}—pathloss threshold value; Num. APs—number of Access Points found; t_t—total processing time; and Num. PL. eval.—number of evaluations.

TABLE 1. Results of the experiments for various values of the threshold, for (a) 223 and (b) 183 potential users.

P_{th}	Num. APs	t_t (sec)	Num. PL. eval.
115	1	2.74	5.6×10^5
114	2	12.48	2.4×10^6
105	2	12.78	2.4×10^6
100	2	14.51	2.7×10^6
95	2	20.48	3.5×10^6
92	3	48.57	7.1×10^6
90	3	48.55	6.9×10^6
80	4	125.87	1.3×10^7
75	6	407.27	3.3×10^7
70	8	701.24	5.1×10^7
65	14	2241.29	1.2×10^8

(a)

P_{th}	Num. APs	t_t (sec)	Num. PL. eval.
115	1	1.01	2.1×10^5
114	1	1.45	3.0×10^5
105	2	10.00	1.9×10^6
104	2	11.15	2.1×10^6
100	2	12.78	2.5×10^6
95	2	11.41	2.2×10^6
92	2	11.27	2.1×10^6
90	2	20.64	3.4×10^6
82	3	41.97	6.1×10^6
80	3	41.43	6.1×10^6
75	5	146.45	1.4×10^7
70	6	285.82	2.5×10^7
65	9	777.29	5.4×10^7

(b)

5.2. Results In [18], the same problem was investigated using a different approach, based on genetic algorithms. Application of the model described in the current paper and the modified AGOP method allows us to obtain better results. For instance, in the case of the threshold of 80 dB, as can be seen from Table 1, four APs are enough to cover the area, in contrast to the five suggested in [18].

The solutions were obtained within a reasonable time. When the number of APs becomes larger, the time necessary to solve the problem increases too. From that viewpoint, a few observations can be made.

- The threshold has an influence on the processing time: although for $P_{\mathrm{th}} = 114$ and for $P_{\mathrm{th}} = 95$ the same number of APs were found (meaning the same number of optimization problems of the same dimensions were solved), nearly 1.5 times more Path Loss evaluations were needed to find the solution. This is due to the fact that a larger threshold means that more locations allow full coverage. This shows that it is very efficient to take into account the lower bound of the problem during the AGOP execution, which allows processing to stop as soon as the solution has been found.
- The time complexity is very dependent on the number of APs necessary to cover the area. This is due to the enumeration part of the method: the more APs that are necessary, the more optimization problems that need to be solved. What is more, only the last problem solution can be quickened by the technique described in the previous item.
- The number of users does influence both the result and the running time. However, the running time is only slightly increased when the number of APs is the same. This means that the potential users should cover the area well, otherwise there may be some inaccuracies. In an area with obstacles, the discontinuity of the objective function results in the need for users to be distributed suitably densely, but this distribution also takes into account the possible effect of every obstacle. On the other hand, if the number of users is extremely large, the method may need more time. This observation is particularly true when the threshold is low (many APs are needed): for $P_{\mathrm{th}} = 65$, the number of APs varies from 14 to only 9 when we remove only 40 potential users.

6. Conclusion and further research

In this paper, we have presented a novel approach for solving a particular type of location problem. The problem consists of minimizing the number of facilities necessary to cover a certain demand, where this coverage depends on their location (namely their distance from the customers). No assumption was made on the coverage of the demand, which can be discontinuous as a function of the distance.

To tackle this problem, a sequence of continuous optimization problems with discontinuous objective functions are solved. The global optimization software AGOP has been modified to take into account the particularities of these problems. In particular, the modifications have been devised in order to accelerate the search of a solution, using information obtained in previous steps.

A particular application to this type of problem arising in the design of wireless telecommunication networks has been presented, and numerical experiments have been carried out on a particular instance of a real-world situation.

These experiments have shown that in most cases, the algorithm outperforms other approaches, while solving within an acceptable amount of time. This is due to the very good performance of AGOP for solving the sequence of problems. It is also shown that the number of customers has little influence on the performance of the algorithm, which depends much more on the number of APs needed (that is, on the demand threshold).

A number of points may be improved in the current algorithms:

- The enumeration method is not very efficient when the number of APs is large. Using approximation algorithms to estimate roughly the number of APs needed may accelerate the current method;
- Experiments have shown that a large portion of the computing time is spent on solving problems that are not very interesting: if the solution of the problem is a quite high number of APs, then much time is spent on solving cases for a lower number of APs. The use of the suggestion from the point above may reduce this effect, but it may also be of interest to analyze the problem more deeply, to improve the search method.
- Although the number of users does not have a strong influence on the efficiency of the method, this may still become an issue when the number of APs is larger. Experiments have shown, however, that if the users are not adequately distributed over the area, the results can be highly inaccurate. In order to accelerate the algorithm while still obtaining satisfactory results, it may be interesting to use a method similar to the ε-cleaning procedure presented in [2] for reducing the number of customers, while still reaching accurate results. Such a method would allow one to specify a density of potential users high enough to ensure that every obstacle is taken into account, while having an automatic tool that generates a smaller set of users representative of the problem. This is achieved because each potential user is within a reasonable range (that is, distance, but also considering obstacles) of a representative user.

Acknowledgements

The authors would like to thank the referees for their valuable comments and suggestions.

References

[1] M. D. Adicks, R. E. Billo, B. A. Norman, S. Banerjee, B. O. Nnaji and J. Rajgopal, "Optimization of indoor wireless communication network", *IIE Trans.* **34** (2002) 823–836.

[2] A. M. Bagirov, A. M. Rubinov, N. Soukhoroukova and J. Yearwood, "Supervised and unsupervised data classification via nonsmooth and global optimisation", *TOP* **11** (2003) 1–93.

[3] M. J. Beckmann, "Continuous models of transportation and location", *Sist. Urb.* **3** (1981) 403–413.

[4] T. Hale, http://www.ent.ohiou.edu/~thale/thlocation.html.

[5] E. R. Hansen, *Global optimization using interval analysis* (Dekker, New York, 1992).

[6] K. C. Kiwiel, *Methods of descent for nondifferentiable optimization*, Lecture Notes in Mathematics 1133 (Springer, Berlin, 1985).

[7] T. G. Kolda, R. M. Lewis and V. Torczon, "Optimization by direct search: New perspectives on some classical and modern methods", *SIAM Rev.* **45** (2003) 385–482.

[8] C. Lemaréchal, "Nonsmooth optimization and descent methods", Technical Report RR-78-004, International Institute for Applied System Analysis, Laxemburg, Austria, 1978.

[9] M. Mammadov, "A new global optimization algorithm based on a dynamical systems approach", in *Proceedings International Conference on Optimization: Techniques and Applications—ICOTA6*, (University of Ballarat, Australia, 2004), article index number 198.

[10] M. Mammadov, A. M. Rubinov and J. Yearwood, "Dynamical systems described by relational elasticities with applications", in *Continuous Optimisation: current trends and applications* (eds. V. Jeyakumar and A. M. Rubinov), (Springer, 2005) 365–385.

[11] R. Morrow, *Wireless Network Coexistence* (McGraw-Hill, New York, USA, 2004).

[12] K. Pahlavan and P. Krishnamurthy, *Wireless Communications: Principles and Practice*, 2nd ed. (Prentice Hall, New Jersey, USA, 2002).

[13] M. A. Panjwani, A. L. Abbott and T. S. Rappaport, "Interactive computation of coverage regions for wireless communication in multifloored indoor environments", *IEEE J. Sel. Areas Comm.* **14** (1996) 420–430.

[14] B. S. Park, J. G. Yook and H. K. Park, "The determination of base station placement and transmit power in an inhomogeneous traffic distribution for radio network planning", in *Proceedings of IEEE 56th Vehicular Technology Conference*, Volume 4, (IEEE, Vancouver, 2002), 2051–2055.

[15] R. Rodrigues, G. Mateus and A. Loureiro, "Optimal base station placement and fixed channel assignment applied to wireless local area projects", in *Proceedings of IEEE International Conference on Networks*, (IEEE Computer Society, Brisbane, September 1999), 186–192.

[16] M. P. Scaparra and M. G. Scutellà, "Facilities, locations, customers: Building blocks of location models: a survey", Technical Report TR-01-18, Università di Pisa, Italy, 2001.

[17] H. D. Sherali, C. M. Pendyala and T. S. Rappaport, "Optimal location of transmitters for micro-cellular radio communication system design", *IEEE J. Sel. Areas Comm.* **14** (1996) 662–673.

[18] K. S. Tang, K. F. Man and K. T. Ko, "Wireless LAN design using hierarchical genetic algorithm", in *Proceedings of 7th International Conference on Genetic Algorithm*, (Morgan Haufman, East Lansing, MI, USA, 1997), 629–635.

[19] K. S. Tang, K. F. Man and S. Kwong, "Wireless communication network design in IC factory", *IEEE Trans. Indust. Elect.* **48** (2001) 452–458.

[20] M. Unbehanun and M. Kamenetsky, "On the deployment of picocellular wireless infrastructure", *IEEE Comm. Mag.* **10** (2003) 70–80.

[21] "Victorian partnership for advanced computing", www.vpac.org.