International Series in
Operations Research & Management Science

H. A. Eiselt
Vladimir Marianov *Editors*

Contributions to Location Analysis

In Honor of Zvi Drezner's 75th Birthday

Springer

International Series in Operations Research & Management Science

Volume 281

More information about this series at http://www.springer.com/series/6161

H. A. Eiselt • Vladimir Marianov
Editors

Contributions to Location Analysis

In Honor of Zvi Drezner's 75th Birthday

Springer

Editors
H. A. Eiselt
Faculty of Management
University of New Brunswick
Fredericton, NB, Canada

Vladimir Marianov
Department of Electrical Engineering
Pontificia Universidad Católica de Chile
Complex Engineering Systems Institute
Santiago, Chile

ISSN 0884-8289 ISSN 2214-7934 (electronic)
International Series in Operations Research & Management Science
ISBN 978-3-030-19110-8 ISBN 978-3-030-19111-5 (eBook)
https://doi.org/10.1007/978-3-030-19111-5

This Springer imprint is published by the registered company Springer Nature Switzerland AG.
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

Tammy & Zvi Drezner © 2008 H.A. Eiselt

Few researchers can be considered as influential in several fields as Zvi Drezner. In his main field of contributions, location theory, he is probably one of the most prolific authors. In his 75th year, some of his coauthors and friends considered him deserving of a book that synthesizes some of his contributions and highlight their importance for different fields. Not less relevant is his work with colleagues and students. He has collaborated with most researchers in the facility location field and has taught numerous students that now follow his path. As a matter of fact, following the nomenclature for the followers of the famous mathematician, who obtain the so-called Erdös numbers, we would like to introduce "Drezner numbers." A Drezner number is simply the number of connections needed to connect oneself with Zvi, given that each connection is a scientific paper. In other words, someone who coauthored a paper with Zvi will have a Drezner number of 1, somebody who wrote a paper with someone who is among Zvi's coauthors, has a Drezner number of 2, and so forth. The two editors of this book are both proud to have a Drezner number of 1. Then again, given the "six degrees of separation" and the fact that Zvi is a very prolific writer, this may not be too surprising.

Other than Zvi's academic achievements, which we will discuss below, the birthday boy is also a character, something rare in this increasingly standardized world. Some years ago, one of us was invited to his and his wife Tammy's home. They were very nice to pick us up at our hotel and get us back. The way back turned out to be a lot longer. The reason was not the usual Los Angeles traffic, but when they got us to our hotel, we sat for at least an hour in the car enjoying their repertoire of Jewish jokes. That alone made for a wonderful evening. We also learned why Zvi is so prolific. As most of his work involve heuristic algorithms and computational testing, he had five computers running in parallel on different problems, crushing numbers, and preparing new publications. One of those computers was in the trunk of the car, making the daily commute a lot more productive than most of us will ever be. That's something to think about. We did another piece of education the next day. Zvi and Tammy took us to a Neiman-Marcus store, where they showed us a purse, suggesting this would be a great gift for my wife. I had to pass, though, as $5000 purses are not on my list of things to get. Only in California, I guess.

Back to computations and heuristics. Apparently, humor is in the blood in all of Zvi's family. I distinctly recall the European Working Group on Location meeting in Barcelona in 2000, when Zvi's wife, Tammy, presented a work by both of them on genetic algorithms. They had introduced different genders in their algorithm, which apparently worked out well computationally. Tammy commented that with the words (a truism, I guess) that "apparently, it makes sense to have sex." There is nothing we can add to that.

Another amazing fact is Zvi's ability to pull "well-known" facts from geometry, trigonometry, and other fields out of a hat, whenever needed. I remember telling Zvi about one of the problems I was working on. He listened, and when I was done, he said something like "oh, that!" This reminded me of the well-publicized episode of young G. B. Dantzig, who told John von Neumann about his results on linear programming duality, which apparently sparked the same response.

It would be exceedingly difficult to pinpoint Zvi's major contribution; also, this is a highly personal assessment. To us, his 1982 paper in *Regional Science and Urban Economics*, probably the first step toward competitive location models by an operations researcher, will certainly be very close to the top of the list. The merging process for the quadratic assignment problem in his 2003 paper in *INFORMS Journal on Computing* was a breakthrough in developing heuristics for quadratic assignment and other location problems. The 2004 paper on the "Big Triangle Small Triangle" method coauthored with Atsuo Suzuki in *Operations Research* had many follow-ups.

It would be impossible to include all Zvi's work or the impact he has had on other researchers in a single book.

We know that Zvi is knowledgeable in many fields, but we thought it would be an excellent idea if we let him write about the topic he must know most about:

his own achievements. He has summarized them in the first chapter of this book. Starting with the story of how he received his degree, we learn about his first paper in 1962 (many of those who have followed in his footsteps and are now active in the field were not even born then!) in the Astronomer's Bulletin, a story vaguely reminiscent of Varignon and his gnomonics. This is followed by Zvi's *really* impressive curriculum vitae and some inspiring thoughts regarding teaching the subject.

Richard L. Church has a fresh look at the Weber problem. He does away with the usual view of Weber's contribution, which looks not so much as Weber's book but as Georg Pick's mathematical appendix, based on Launhard's work. It turns out that while Weber describes a very simple prototype of a location problem that goes little beyond the usual Fermat and Torricelli models, his actual discussion is about much more complex problems, a number of which sometimes are referred to by some of his latter-day successors as "generalizations of the standard Weber problem," even though they were already described by Weber.

Jack Brimberg and Said Salhi investigate p-median problems in continuous space. Based on Cooper's original location—allocation method—they first review a number of heuristic algorithms based on Cooper's contribution, followed by the introduction of a number of refinements. New algorithms, such as a depth-first search method and a decomposition technique, are described, and computational results are offered.

Atsuo Suzuki's main interest concerns Weber problems, and his contribution in this book is a Weber problem with a new twist. In particular, the piece discusses a Weber problem on the sphere. The applications of such problems are found on locations on a global scale. The formulation of this problem turns out to have a nonconvex objective. As a result, the usual Weiszfeld algorithm will not necessarily find an optimal solution. He employs a version of the big triangle-small triangle method, developed by Zvi Drezner and the author.

Anita Schoebel's work is on well-known continuous locations but with the additional twist that the single new facility has to have integer coordinates. The applications of this problem are found in robust decision-making. Her paper delineates finite dominating sets for problems with ℓ_1, ℓ_2, and ℓ_∞ distances.

The chapter by George Wesolowsky also looks at the Weber problem with particular focus on the many papers he and Zvi have written together on the subject. This includes the usual minisum problem, continuous location problems with minimax and maximin objectives (the latter in case of undesirable facilities), covering models, hub problems, as well as different solution approaches, including trajectory methods and a Demjanov-style algorithm.

Mozart Menezes's work deal with voting methods, "wisdom of the crowd," and Condorcet points. In location and voting problems, he explores relations between Weber solutions and Condorcet points and conducts numerical experiments regarding the computation of Condorcet points for problems with Euclidean distances. He uses the notion of a "benevolent dictator" to examine the relative efficiency of Condorcet solutions.

Taly Drezner, a scientist in biogeography, discusses heuristics in her paper with a focus on genetic algorithms. Her piece on biological principles in genetic algorithms with male and female subsets is an original work. She subsequently describes some statistical tests developed to verify computational results.

In her paper, Sibel Alumur reviews a variety of hub location and related problems under special consideration of Zvi Drezner's contributions to the individual field. Besides hub location problems, she includes round trip location problems, transfer point location problems, and collection depot location problems. Each of the problems is formulated, and the impact of Zvi's work on the specific field is highlighted.

Tammy Drezner, best known for her work on attraction functions, presents thoughts and results in her contribution to this book. Defining the concept of attractiveness, she first addresses distance decay and different attraction functions. She then explores different extensions, e.g., budget constraints, lost demand (nonessential goods), and solution concepts (e.g., von Stackelberg's famed leader-follower model) along with properties, such as the consistency of rules. Some results concerning solution techniques finish her chapter.

Pawel Kalczynski's contribution concerns competitive location models in the plane that are based on the concept of covering. He delineates spheres of influence, defines purchasing power, and formulates a competitive model based on covering ideas. Leader-follower algorithms and heuristics are offered alongside with bounds and computational experience.

Xin and Wang examine a location problem with a median objective, in which the weights, representing the customers, are random. The authors consider problems, in which a "value-at-risk" constraint has been added. They develop an algorithm for the problem, which they illustrate by means of a numerical example.

The last chapter in this book by Dawit Zerom is different from the others, in that it deals with a new statistical distribution, viz., the bivariate exponential distribution and its properties, which he and Zvi developed together. We are glad to include this work, as it is one of the many examples in which Zvi has gone beyond location models and shown that he is, indeed, a true Renaissance man.

Last, but certainly not least, we would like to express our thanks to the Springer staff, who were great as usual: Camille Price for snuckering us into putting together another book; Matthew Amboy, Neil Levine, and Faith Su for supporting us all the way and facilitating the process; as well as Aparajita Srivastava and Christine Crigler in production. All that is left for us to do is to wave to Zvi (and, of course, Tammy) and wish them many happy returns and many more papers.

Professor Marianov acknowledges the support by grant FONDECYT 1160025 and by the Complex Engineering Systems Institute through grant CONICYT PIA FB0816.

Fredericton, NB, Canada H. A. Eiselt
Santiago, Chile Vladimir Marianov
February 2019

Contents

Chapter 1
My Career and Contributions

Zvi Drezner

1.1 Early Career

I started my career earning a BA in Mathematics in 1965 from the Technion, Israel Institute of Technology. I then served a mandatory service in the military for an extended period of time so I could serve at the newly established computer center as a computer programmer. The first computer, Philco 212, had a 32K memory and its machine language was TAC. Later on the computer language ALTAC (Algebraic TAC) was designed. Some years later ALTAC was modified to FORTRAN (FORmula TRANslator) which is still widely used today. Most of my programs nowadays I code in Fortran.

I registered for a master's degree in the Technion. One of the projects that I was assigned in the military computer center was the planning of a new military base. There were 57 buildings/units planned and there were many requirements that certain pairs of buildings should be close to one another (with a weighted importance) and some should be far from one another (for example, the intelligence office and the cafeteria or the entrance) and had negative weights. I was told that the project was given to an architectural firm and they could not find a satisfactory layout.

At the time I was not familiar with Operations Research and did not know anything about layout algorithms. The approach I used for its solution was based

Z. Drezner (✉)
Steven G. Mihaylo College of Business and Economics, California State University-Fullerton, Fullerton, CA, USA
e-mail: zdrezner@fullerton.edu

H. A. Eiselt, V. Marianov (eds.), *Contributions to Location Analysis*, International Series in Operations Research & Management Science 281,
https://doi.org/10.1007/978-3-030-19111-5_1

on Physics concepts. I imagined each building to be a circle (some are bigger than others) and the weights being springs connecting pairs of circles with their strength proportional to the weight. After a few attempts using this approach, I put all the circles at one point and "exploded" them while connected with springs. The system of circles was spread out, so I stopped the explosion at some point and let the circles move back maintaining the springs and requiring that they do not overlap. When the architectural firm got my layout they could not believe their eyes. All pairs of buildings that needed to be close to one another were indeed close and those that needed to be far from one another were indeed far. My military friend David Levenbaum, who later did a Ph.D. in Physics, coined the name "The Big Bang" method.

I decided to propose this approach as the thesis for my master's degree. My advisor Prof. Ilan Amith came up with the name "Dispersion-Concentration," in short—DISCON. Believing that I am going to get my master's degree I applied and was accepted for a Ph.D. at Tel-Aviv University.

When I defended my thesis, the committee decided that this thesis is qualified for a Ph.D. The committee considered my defense as a defense of a Ph.D. proposal. They said, "come back in a year, submit it and get a Ph.D." They found two professors in the corridor (for a Ph.D. defense you need five faculty rather than three) who signed my defense. The criterion for a Ph.D. is that the committee members consider the dissertation suitable for publication in the top journal of the field. Indeed, the main part of the dissertation was published in Operations Research (Drezner 1980). Another part of the dissertation was published much later (Drezner 2010a). Follow-up papers include Drezner (1987a) where an alternative approach based on the eigenvectors of the weights matrix is used. Marcoulides and Drezner (1993) applied this approach to convert an n-dimensional data into a two dimensional plot. This conversion method was used in Marcoulides and Drezner (1999) for cluster analysis and in Drezner and Marcoulides (2006) for illustrating the convergence of genetic algorithms.

I called Tammy and told her what happened. I said that I am not sure I want to do it because I was accepted for a Ph.D. at Tel-Aviv university with a TAship. Also, I will not have a master's degree. Tammy said that I "must be out of my mind, take it!" I talked to the faculty at Tel-Aviv University and they were very nice about it and even offered me to keep my TAship even though I withdrew from the Ph.D. program. The committee asked me whether I wouldn't mind to get my degree in Computer Science because they just founded the department. So, I am the first Ph.D. in Computer Science from the Technion (Drezner 1975).

Scholar.Google reminded me of a refereed paper I co-authored during my military service (Almogy et al. 1968). The paper simulates the operation of a fleet. The simulation comprises two major parts: (a) the schedule generator, which assigns schedules to the fleet elements by considering cargo generations at the different ports and determining the most profitable route for a particular ship within the limits of the policy constraints; (b) the voyage generator, which moves each ship along the route assigned to it by the schedule generator. The simulation was written for the Philco 212 unit of the Computer Center of the Israeli Ministry of Defense.

I thought that Almogy et al. (1968) was my first refereed paper until a colleague Micha Hofri reminded me of a manuscript written in Hebrew (Drezner 1962), when I was 19, on methods for calculating a satellite orbit. I translated it to English and summarized some highlights below.

1.1.1 Highlights of Drezner (1962)

Drezner (1962) appeared in four parts in the "The Stars in Their Month" Bulletin of the Israeli Amateur Astronomers Society: Volume 9 (1962) No. 12, pages 136–141; Volume 10 (1963) No. 3, 29–37, No. 4, 49–53, and No. 5, 65–70. It was collected by the Society and published as one paper of 27 pages. I thank Shimon Malin for his constructive comments and the extensive work he invested in preparing the manuscript for publication.

In Drezner (1962) I show methods for calculating the time to complete an orbit, the precession velocity, the satellite inclination to the earth equator, measuring the satellite's altitude, its entering and exiting the sun's shade and calculating its average height. I applied each method on my actual observations of the satellite Echo-1 launched on Aug 12, 1960. Echo-1 was a 100 ft (30.5 m) diameter metal balloon reflecting the sun's rays that was easily observed by the naked eye. Echo-1 re-entered Earth's atmosphere, burning up on May 24, 1968.

A satellite orbit can be described as an ellipse with the earth as one of its foci (Kepler's first law, Russell 1964). However, this ellipse is not stationary in space. While moving around the earth, its plane is slowly rotating so that the line perpendicular to its plane depicts a cone originating at the earth's center (this is termed precession).

If the ellipse was not rotating, the satellite would have moved in a stationary ellipse in space. When observing the satellite in a particular moment we ignore, for the purpose of the mathematical formulation, the ellipse's precession and discuss the ellipse that the satellite is moving in at that moment. This ellipse is termed "the belt." We then correct the position of the satellite for the precession motion. We can say, for example, that the belt is passing overhead at a certain moment, if the zenith is one of the belt's points, even if the satellite is not there. If, for example, we observe the satellite west of the zenith, we can conclude, most of the time, that the belt passed the zenith earlier. The velocity of the precession's rotation is the velocity of the belt movement in space.

Summary of Notation

Symbol	Units	Description
R	km	Earth radius
H	km	Satellite's distance from earth's center
h	km	Altitude of the satellite above earth ($H - R$)
ξ	deg	Satellite's angle above the horizon
x	deg	The angle observed from the earth center between us and the satellite
γ	deg	The angle between the belt and the equatorial plane
ϕ	deg	Northern latitude of the observer (my home was at $32°03'$)
T_0	h	Precession time after 24-T_0 hours (see Sect. 1.1.2)
T	min	Satellite's time to complete one orbit (the period)
y	min	The time difference that the satellite is observed earlier next day
n	days	Number of days that the satellite moves to a next orbit (Sect. 1.1.5)
K		Integer number of complete orbits per day (rounded down)
δ	deg	Sun's inclination (northern)$+74'$ (minutes of degrees)
$\overline{H}$	km	Half the larger axis of the ellipse

We also define: the great circle connecting west to east through the zenith as "the line"; the point where the satellite crosses our latitude ϕ is termed "the true point" (see Sect. 1.1.4).

1.1.2 Measuring the Precession Rate

Suppose that in a particular day the satellite was overhead at 19:00. Next day it will probably not pass overhead. There are two reasons for that: (1) the belt will not pass again overhead at 19:00 because of the precession, and (2) the satellite will not necessarily be at that point in the belt. Suppose that after 5 days the satellite passes overhead at 18:00. This means that the belt is passing overhead $\frac{60}{5} = 12$ min earlier every day. To find T_0 we wait for a number of days until the satellite passes overhead again and divide the difference by the number of days.

1.1.2.1 Measuring the Precession of Echo-1

In order to measure the daily precession we need to measure at least twice the time the belt was at the zenith. It is sufficient to find its passing "the line" near the zenith and estimate the passing overhead. In Table 1.1 we list actual measurements of Echo-1 and the passing at the zenith corrected by (1.13) below.

We show how to calculate the precession based on the first two observations. The time between July 5, 1961, and April 21, 1962, is 290 days. Echo-1 completed

Table 1.1 Observations of Echo-1 passage

Date	Time	At "the line"	Estimated zenith time
July 5, 1961	3 h 44 min 30 s	5° W of zenith	3.66 h
April 21, 1962	18 h 19 min 45 s	At the zenith	18.33 h
June 15, 1962	3 h 8 min 37 s	2.5° W of zenith	3.10 h
November 26, 1962	4 h 41 min 50 s	1° W of zenith	4.68 h

Table 1.2 Measuring the precession rate

Period	No. of days	"Satellite days"	Time	T_0
July 5, 1961–April 21, 1962	290	290 + 3 + 1	81.33	0.2766
April 21, 1962–June 15, 1962	55	55	15.23	0.2769
June 15, 1962–November 26, 1962	164	164 + 2	46.42	0.2796

three full precession rounds and therefore the cumulative precession time is 81.33 h ($3 \times 24 = 72$ h + 18.33–3.66 h). The precession rate T_0 is the precession time of the orbit after a "satellite day," which is $24 - T_0$ hours. We therefore add 3 days because there were three full precession rounds and one more because the change from morning to evening. In total, there were 294 satellite days and thus $T_0 = \frac{81.33}{294} = 0.2766$ h. In Table 1.2 we depict the three calculations.

I speculate that the relative large change in the third period might have been caused by a nuclear test by the USA at high altitude of 300 km over Johnston island on July 9, 1962, at 11:00 Israel time. I calculated that Echo-1 was close to that location and might have been jolted by that explosion.

1.1.3 Measuring the Inclination of the Belt to the Equatorial Plane

Suppose that our latitude is smaller than the satellite's inclination to the equatorial plane. Suppose that the satellite passes overhead northward and after t_1 hours moves overhead southward. Consider Fig. 1.1. O is the earth's center, points C, D, B are on the satellite's orbit, L is the center of the circle (inside the earth) of latitude ϕ. The angle between LB and LC is $15t_1$ (in degrees) and therefore the angle between LB and LA is half of it. We get

$$\cos(7.5t_1) = \frac{AL}{LB} = \frac{AL}{LO} \times \frac{LO}{LB} = \cot\gamma \tan\phi. \tag{1.1}$$

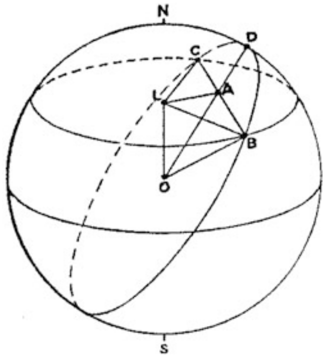


Fig. 1.1 Proof of Eq. (1.1)

1.1.3.1 Measuring the Inclination of Echo-1 to the Equatorial Plane

Two measurements are applied for this calculation:

(1) On April 21, 1962, Echo-1 passed the zenith in 18 h 19 min 45 s from the south-west to the north-east.
(2) On May 5, 1962, it passed at 21 h 41 min 39 s 2° west of the zenith going from the north-west to the south-east. By the time correction for precession the belt passed the zenith 2×0.016 h earlier, i.e., at the time 21.662 h.

In the first measurement Echo-1 passed at the zenith. We use $T_0 = 0.2769$. Fourteen days have passed, therefore, the first point passed the zenith $14 \times 0.2769 = 3.877$ h earlier, i.e., at $18.329 - 3.877 = 14.452$. The difference is $21.662 - 14.452 = 7.210$. However, formula (1.1) does not consider the precession. We need to incorporate the precession movement during these 7.210 h. During this time the belt progressed $7.210\frac{T_0}{24-T_0} = 7.210\frac{0.2769}{24-0.2769} = 0.084$ h. Therefore, we use in (1.1) $t_1 = 7.210 + 0.084 = 7.294$ h. Using $\phi = 32°03'$ we get $\gamma = 47°17' \pm 2'$.

1.1.4 Measuring the Period (First Method)

In order to determine the time it takes for the satellite to complete one orbit, we need to find a point in the orbit which is easy to determine when the satellite went through it. We select the point it crosses our latitude ϕ. If it happens to be observed at the zenith, it is clearly above ϕ. If not, it looks like it passes over ϕ when it crosses the line west-zenith-east which we call “the line.” However, this is not accurate. In Fig. 1.2 it crosses the line at point A but is above ϕ at point B which is *north* of the line. This point is termed the “true point.” We find the time ΔT between the satellite being at A and at B.

See Fig. 1.3. We first find the relationship between x and latitude ϕ_1 where the satellite is. The angle ϕ is between OA and OL. The circle $QABP$ is “the line” whose radius is OP. B is the satellite location. Consider the planes perpendicular

Fig. 1.2 Proof of Eq. (1.4)

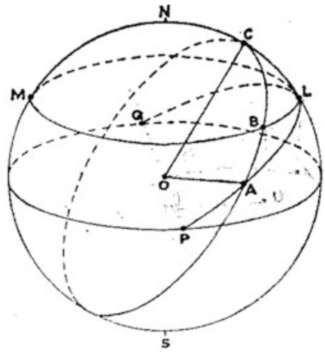
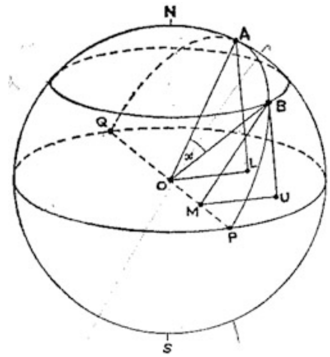


Fig. 1.3 Proof of Eq. (1.2)



to OP passing through A and B. The triangles AOL and BMU are similar, thus $\frac{BU}{AL} = \frac{BM}{AO} = \frac{BM}{R}$ yielding $\frac{BU}{R} = \frac{AL}{R} \times \frac{BM}{R}$. Therefore,

$$\sin\phi_1 = \sin\phi\sin(90^\circ - x) = \cos x \sin\phi. \tag{1.2}$$

Now consider Fig. 1.2. If we are at the northernmost point of the orbit C with latitude γ and the satellite is passing overhead, it will rise in the west and set in the east (because when it crosses our latitude it is at the northernmost point of its orbit and thus crossing the longitude line in a right angle). Its orbit is along "the line." Let the angle between OC and OB be θ_1 and the angle between OC and OA be θ_2. By Eq. (1.2) applied twice:

$$\cos\theta_1 = \frac{\sin\phi}{\sin\gamma}; \quad \cos\theta_2 = \frac{\sin\phi_1}{\sin\gamma} = \cos x\frac{\sin\phi}{\sin\gamma}. \tag{1.3}$$

The angle between "the line" and the "true point" is $\theta_2 - \theta_1$. The time ΔT it takes the satellite to travel between these two points satisfies $\theta_2 - \theta_1 = 2\pi\frac{\Delta T}{T}$. Thus:

$$\sin\left(2\pi\frac{\Delta T}{T}\right) = \sin(\theta_2 - \theta_1) = \sin\theta_2\cos\theta_1 - \sin\theta_2\cos\theta_1$$

$$= \sqrt{1 - \cos^2 x \frac{\sin^2\phi}{\sin^2\gamma}} \times \frac{\sin\phi}{\sin\gamma} - \cos x \frac{\sin\phi}{\sin\gamma}\sqrt{1 - \frac{\sin^2\phi}{\sin^2\gamma}}$$

$$= \frac{\cos x \sin\phi}{\sin^2\gamma}\left(\sqrt{\frac{\sin^2\gamma}{\cos^2 x} - \sin^2\phi} - \sqrt{\sin^2\gamma - \sin^2\phi}\right)$$

$$= \frac{\sin\phi \sin x \tan x}{\sqrt{\sin^2\gamma - \sin^2\phi} + \sqrt{\frac{\sin^2\gamma}{\cos^2 x} - \sin^2\phi}}. \tag{1.4}$$

The angle x is small; therefore, we can use the approximation:

$$\frac{\Delta T}{T} \approx \frac{\sin\phi \sin^2 x}{4\pi\sqrt{\sin^2\gamma - \sin^2\phi}}. \tag{1.5}$$

In order to apply (1.4) or (1.5) we need to determine the value of x. We measure ξ, the angle of the satellite above the horizon, and we have

$$\cos(x + \xi) = \frac{R}{H}\cos\xi. \tag{1.6}$$

Ways to estimate H are discussed in Sects. 1.1.6 and 1.1.7.

In order to estimate the time of one orbit, we determine its passing over latitude ϕ in different days. We measure it passing "the line" and correct for passing over latitude ϕ using the correction by Eq. (1.4) or (1.5).

1.1.4.1 Measuring the Period of Echo-1 by the First Method

The two measurements of Echo-1 crossing "the line" were:

(1) On May 5, 1962, it passed at 21 h 41 min 39 s in the west at 88° above the horizon.
(2) On May 10, 1962, it passed at 21 h 46 min 2 s in the west moving south at 27° above the horizon.

In order to know when Echo-1 passed the "true point," we correct these times by ΔT calculated by Eqs. (1.5) and (1.6). The correction for the first measurement is negligible and the correction for the second measurement is 0.90 min. Echo-1 moved southward; therefore, we subtract 0.9 min or 54 s yielding the time 21 h 45 min 8 s. The time between the two measurements was 5 days, 3 min, 29 s or 7203.48 min. Every day Echo-1 makes about 12.4 orbits; therefore, it completed 62 orbits and its period is

$$T = \frac{7203.48}{62} = 116.185\,\text{min}. \tag{1.7}$$

To calculate the time it appears after one day (see Sect. 1.1.5), if its period was 120 min, it will show up at the same time again after 24 h. Between these 2 days, in 12 orbits it will show up $y = 12(120 - 116.185) = 45.78$ min earlier the next day.

1.1.5 *Measuring the Period (Second Method)*

Ignore for a moment the precession assuming that the satellite moves in a stationary ellipse in space. Because of the earth's rotation, this ellipse will be observed rotating. If today, for example, the satellite is passing a certain point such as the zenith, it will pass in a "subsequent orbit," meaning in its next orbit, far from that point. Suppose now that its period is exactly 2 h (or any other time dividing to a whole number in 24 h). In this discussion we ignore the rotation of the earth around the sun. In such a case it will pass the same point after 24 h. However, usually the period does not divide exactly 24 h. Suppose that the period is 119 min, then the satellite will pass the same point 12 min earlier next day. In about 10 days the "subsequent" orbit will pass at the same point.

In general, suppose that the satellite is behind y minutes per day. How many days will it take for the subsequent orbit to pass through the same point? Without considering the precession, the time of one cycle is $\frac{1440-y}{K}$ minutes. The cumulative delay in n days is ny, the consecutive orbit will return to the same point if $ny = \frac{1440-y}{K}$. If the precession per day is T_0, then the condition is $n(y - T_0) = \frac{1440-y}{K}$ leading to

$$n = \frac{1440 - y}{K(y - T_0)} \quad \text{or} \quad y = T_0 + \frac{1440 - T_0}{Kn + 1}. \tag{1.8}$$

We suggest the following approach based on this idea. If n is very close to an integer number, we measure the passing time of the satellite at the same point n days apart and divide the time difference by $Kn + 1$. If n is not close to an integer we can select a multiple of n which is close to an integer. For example, if $n = 3.48$ we can measure the times 7 days apart.

1.1.5.1 Measuring the Period of Echo-1 by the Second Method

I measured twice the time Echo-1 crossed a certain wall (see Sect. 1.1.5). On April 29, 1962, it crossed it at 22 h 19 min 48 s and on May 11, 1962, it crossed the same wall at 18 h 59 min 7 s. The accuracy of the measurements is ±3 s. The time between the two measurements is 12 days minus 200.7 min or 17,079.3 min.

During this period the number of orbits per day was about 12.4. Therefore, Echo-1 completed 147 orbits and thus

$$T = \frac{17{,}079.3}{147} = 116.186\,\text{min} \tag{1.9}$$

which is very close to (1.7).

The belt passes $y = 120(120 - 116.186) = 45.77$ min earlier each day. To calculate n (the number of days the consecutive orbit will show at the same time) we apply in (1.8) $T_0 = 0.2769$ h $= 16.61$ min and get $n = 3.984$ days.

We investigate the error in measuring T or y because $n = 3.984$ is less than 4. We experimented with another case where the difference between n and a whole number is larger. We calculate the error assuming that it is proportional to the difference from a whole number.

The next measurement was at May 10, 1962. Echo-1 crossed the same wall at 19 h 46 min 54 s. 15,687.1 min passed since April 29, 1962. Echo-1 had $12 \times 11 + 3 = 135$ full orbits, meaning that it passed 3 times during 11 days to the subsequent orbit. The calculated period is $T_1 = \frac{15{,}687.1}{135} = 116.201$ min and $y_1 = 1440 - 12 \times 116.201 = 45.59$ min.

These calculations were done assuming that when Echo-1 crossed the wall on May 10 it completed indeed $12 \times 11 + 3 = 135$ complete orbits, i.e., it passed 3 consecutive orbits in 11 days, and thus $n = \frac{11}{3} = 3.667$. We got $y - y_1 = 0.18$ min. Assuming that the difference is proportional to the deviation of n from an integer number, then: $\frac{0.18}{3.984 - 3.667} = \frac{\Delta y}{4 - 3.984}$ yielding $\Delta y = 0.009$ min. Therefore, the corrected y is $45.77 - 0.009 = 45.76$ min. The value of T can be corrected accordingly. We can see that the error is negligible.

1.1.6 Measuring the Satellite's Altitude

We observe the satellite in "true points" on two occasions when one of the observations is at the zenith. In the second observation it passes the true point at angle ξ. The time difference between the two observations considering the precession T_0 (see Sect. 1.1.2) is t_1. t is defined by

$$\sin(7.5t) = \sin t_1 \cos \phi. \tag{1.10}$$

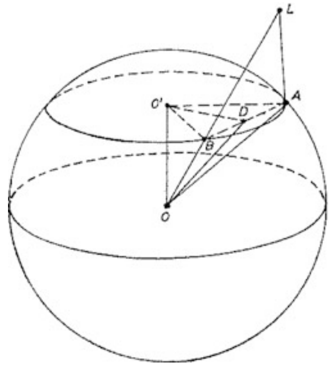
The angle $15t$ is the angle originated at the earth's center between the two observations A and B (see Fig. 1.4). D is at the center between A and B and therefore the angle between OB and OD is half that angle, $7.5t$.

L is the satellite's location, H is the length of OL, and R is the length of $OA = OB$. Applying the sinuses theorem on the triangle OLA, we get

$$\frac{R}{H} = \frac{\cos(15t + \xi)}{\cos \xi}. \tag{1.11}$$

The relation between the angle ξ of the satellite above the horizon and the angle x seen from the earth's center is given by Eq. (1.6). This formula can be written as

Fig. 1.4 Proof of Eq. (1.10)



$$\tan \xi = \frac{\cos x - \frac{R}{H}}{\sin x} \tag{1.12}$$

Note that it can also be measured by (1.6) directly: we can find ξ when the satellite is at its lowest point in the north. This happens when the northernmost point in the satellite orbit is to our north. In such an occasion $x = \gamma - \phi$. Let Δx be the distance between two observers seeing the satellite at the same time. Suppose also that the satellite is located in the plane determined by the two observers and the earth's center. What will be the difference between the two angles ξ that they measure?

In order to investigate the relationship between small differences in x to the resulting differences in ξ we find the derivatives of the two sides of (1.12) by x: $(1 + \tan^2 \xi)\frac{d\xi}{dx} = \frac{-\sin^2 x - \cos^2 x + \cos x \frac{R}{H}}{\sin^2 x} = \frac{-1 + \cos x \frac{R}{H}}{\sin^2 x}$. Substituting $\tan \xi$ by (1.12) we get: $\frac{d\xi}{dx} = \frac{\frac{R}{H}\cos x - 1}{1 - 2\frac{R}{H}\cos x + \left(\frac{R}{H}\right)^2} = -\frac{1}{2} - \frac{\frac{1}{2}\left[1 - \left(\frac{R}{H}\right)^2\right]}{1 + \left(\frac{R}{H}\right)^2 - 2\frac{R}{H}\cos x}$, yielding

$$\Delta\xi = -\left(\frac{1}{2} + \frac{\frac{1}{4}\left(\frac{H}{R} - \frac{R}{H}\right)}{\frac{1}{2}\left(\frac{H}{R} + \frac{R}{H}\right) - \cos x}\right)\Delta x. \tag{1.13}$$

The negative sign means that when x increases ξ declines.

When two observers cooperate, it is possible to estimate $\frac{d\xi}{dx}$ and get H. Also, this formula can also be used when we measure ξ in a location different from the one we usually use, and adjust ξ to the value that would be obtained at the usual location.

1.1.6.1 Measuring Echo-1 Altitude

The calculation of the height of Echo-1 on May 8, 1962, by Eqs. (1.10) and (1.11) is based on two observations:

(1) It passed at 18° in the east moving south at 18 h 25.9 min.
(2) The belt passed in the zenith that day at 20.84 h.

In the first measurement $\xi = 18°$; thus, its deviation from "the line" is by (1.4) $\Delta T = 1.5$ min. The "true point" is always north of "the line" (see Sect. 1.1.4); therefore, it was at the "true point" at 19 h 25.9 min − 1.5 min = 19 h 24.4 min.

In Fig. 1.7 we plotted the height difference between the height at "the line" ξ and the height at the "true point." Since the true point is north of the line and it moved south we find on the graph a difference of 6.5°. Therefore, its height at the "true point" is $18° + 6.5° = 24.5°$. In summary, it passed the "true point" at 19.41 h, which is 1.43 h before the belt crossed the zenith, at an angle of 24.5° above the horizon.

For the second data point we need to adjust the time for the precession. We use $T_0 = 0.277$ h for a satellite day and therefore the correction for the precession is $\frac{1.43 \times 0.277}{24 - 0.277} = 0.02$ h. If there was no precession the belt would have passed the zenith at $20.84 + 0.02 = 20.86$ h.

We use $t_1 = 20.86 - 19.41 = 1.45$ h; $\phi = 32°03'$ in (1.10) and get $15t = 18.4°$. By (1.11) $\frac{R}{H} = 0.805$. Since $R = 6370$ km, we get $H = 7910$ km. The altitude of Echo-1 is $h = H - R = 1540 \pm 60$ km. The error was calculated assuming an error of $\pm 1°$ in measuring ξ.

This measurement and three others are summarized in Table 1.3. The first three measures were done on consecutive days. Their average is $h = 1500 \pm 40$. On July 4th Echo-1 was 190 km higher at the same "true point." It means that the perigee and apogee are not stable in space. They complete a full cycle in about 110 days.

1.1.7 Measuring the Satellite's Altitude by Its Entering or Exiting the Sun's Shade

Let δ be the northern inclination of the sun plus 74′ (with the appropriate sign). The 74′ is the refraction of the sun rays by the atmosphere. When the satellite disappears (or appears) above latitude ϕ and the satellite's location is t hours before or after midnight (at midnight the sun is lowest below the horizon), then:

$$\frac{R}{H} = \sqrt{1 - (\cos\phi\cos\delta\cos(15t) - \sin\phi\sin\delta)^2}. \tag{1.14}$$

Table 1.3 Measuring the altitude of Echo-1

Date	Passing "the line"		At the "true point"		Belt at (time in h)		t_1	h
(1962)	$\xi°$	Time (h m s)	$\xi°$	Time (h)	Zenith	Corrected	(h)	(km)
May 8	18 E	19 23 9	24.5	19.41	20.84	20.86	1.45	1540
May 9	16.5 W	22 32 23	12	22.51	20.56	20.54	1.97	1430
May 10	27 W	21 46 2	23.5	21.75	20.29	20.27	1.48	1530
July 4	34 E	4 1 53	38	04.02	05.07	05.08	1.06	1690

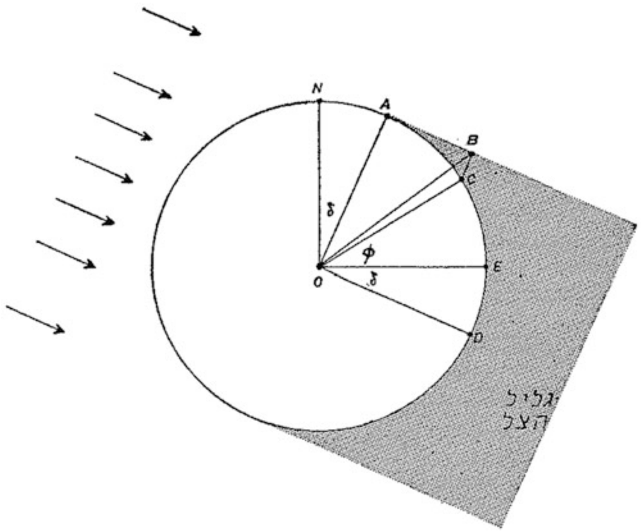


Fig. 1.5 Proof of Eq. (1.15)

We first prove the formula for the case $t = 0$ (see Fig. 1.5). The satellite enters (or leaves) the sun's shade at point B so that the point under it is at midnight. OA is the earth's radius R and OB is the distance between the satellite and the earth's center H. The angle between CO and CB is $90° + \xi$.

$$\frac{R}{H} = \frac{OA}{OB} = \cos\{90° - (\phi + \delta)\} = \sin(\phi + \delta)$$
$$= \sqrt{(1 - \cos^2(\phi + \delta)} = \sqrt{1 - (\cos\phi\cos\delta - \sin\phi\sin\delta)^2} \qquad (1.15)$$

For the general case consider Fig. 1.6. The satellite will enter the at above the same point B so that $AB = R$. The distance H satisfies: $\frac{R}{H} = \frac{AB}{OB} = \frac{\sqrt{OB^2 - OA^2}}{OB} = \sqrt{1 - \left(\frac{OA}{OB}\right)^2}$. OA is the projection of OB in the direction OD. In order to calculate OA consider the vector OB as a sum of two vectors that are the projections of (1) on the earth axis and (2) on the plane perpendicular to this axis. We calculate separately the projection of each vector on OD:

(1) its length is $OB\sin\phi$ and its projection is $-OB\sin\phi\sin\delta$,
(2) its length is $OB\cos\phi$. Its projection on OE is $OB\cos\phi\cos(15t)$ and its projection on OD is $OB\cos\phi\cos(15t)\cos\delta$.

Therefore,

$$\frac{OA}{OB} = \cos\phi\cos(15t)\cos\delta - \sin\phi\sin\delta$$

yielding Eq. (1.14).

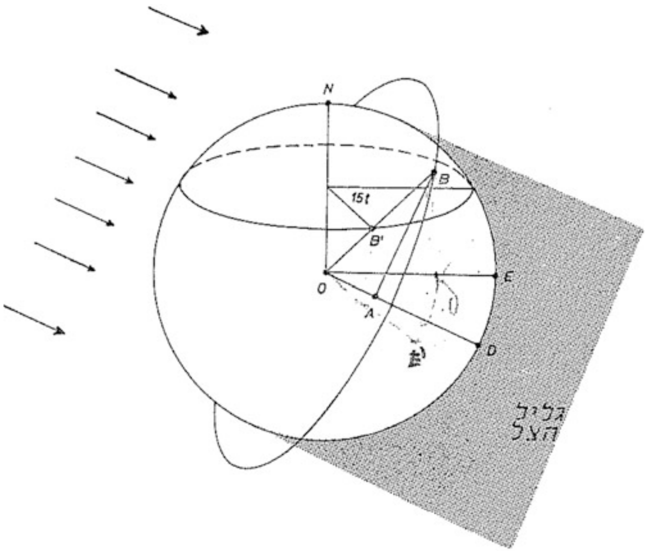


Fig. 1.6 Proof of Eq. (1.14)

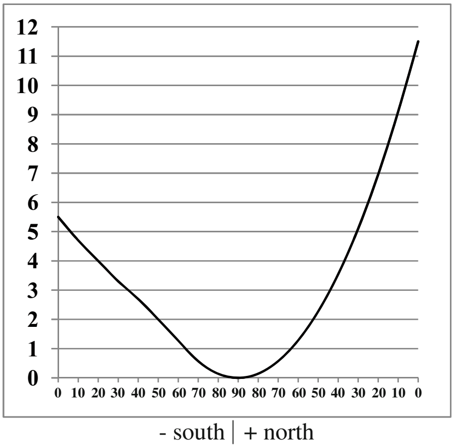


Fig. 1.7 Correction between "the line" and the "true point"

To simplify the calculation of (1.14) define u by: $\cos u = \cos\phi \cos(15t) \cos\delta - \sin\phi \sin\delta$ and then $\frac{R}{H} = \sin u$ (Eq. 1.14).

The error in calculating the height directly by this formula is quite small. However, I found a convenient way to use it when the satellite is exactly over the "true point" (above our latitude ϕ). The value of t used in (1.14) is the time that passed since midnight until the satellite passed overhead when it disappeared. If we ignore the precession, then the time between midnight and the belt passing the zenith is the same in all places with the same latitude. Therefore, in order to obtain t we can find the time that the belt passed overhead and take into account T_0.

Another method for measuring the height is to find its angle over the horizon when it disappears (or appears) exactly to our north or south, i.e., at the same longitude, at midnight. The formula is

$$\frac{R}{H} = \frac{1}{\sqrt{1 + \left(\frac{\cos\xi}{\sin(90^\circ - \xi - \phi - \delta)} - \cot(90^\circ + \xi - \phi - \delta)\right)^2}}. \tag{1.16}$$

To prove this formula consider Fig. 1.5. Let the angle between OA and OB be $90^\circ - u$, then $\frac{R}{H} = \sin u$. The angle between OC and OA is $90^\circ - \phi - \delta$. Therefore, the angle between OC and OB is $u - \phi - \delta$. Since the angles of a triangle add up to 180° we get that the angle between BC and BO is $90^\circ - \xi - u + \phi + \delta$. By the sine theorem

$$\frac{BO}{\sin(90^\circ + \xi)} = \frac{CO}{\sin(90^\circ + \xi + u - \phi - \delta)}$$

$$\frac{1}{\sin u} = \frac{H}{R} = \frac{BO}{AO} = \frac{BO}{CO} = \frac{\sin(90^\circ + \xi)}{\sin(90^\circ + \xi + u - \phi - \delta)}$$

$$\sin u \cos\xi = \sin(90^\circ + \xi + u - \phi - \delta)\cos u + \sin u \cos(90^\circ + \xi + u - \phi - \delta)$$

$$\sin u[\cos\xi - \cos(90^\circ + \xi + u - \phi - \delta)] = \sin(90^\circ + \xi + u - \phi - \delta)\cos u$$

$$\begin{aligned}\cot u &= \frac{\cos\xi - \cos(90^\circ + \xi + u - \phi - \delta)}{\sin(90^\circ + \xi + u - \phi - \delta)} \\ &= \frac{\cos\xi}{\sin(90^\circ - \xi - \phi - \delta)} - \cot(90^\circ + \xi - \phi - \delta)\end{aligned} \tag{1.17}$$

therefore,

$$\frac{R}{H} = \sin u = \frac{1}{\sqrt{1 + \cot^2 u}}$$

which is (1.16). For convenience we can find $\cot u$ by (1.17) and then $\frac{R}{H} = \sin u$.

1.1.7.1 Measuring the Period by Echo-1 Entering or Exiting the Sun's Shade

In Sect. 1.1.7 we showed how to use Eq. (1.14) if the satellite enters the shade at the "true point." In such a case the difference between midnight and the passage of the belt overhead at that evening is t in (1.14). We show how we concluded that on

May 4, 1962, at about 20:00 Echo-1 entered the shade at the "true point" (over our latitude).

We rely on the following two measurements:

(1) On May 4, 1962, Echo-1 disappeared because it entered the shade at 22 h 27 min 29 s.
(2) On May 5, 1962, it disappeared at 21 h 42 min 23 s.

On May 5 I also had the time it crossed "the line" and it passed the "true point" on 21 h 41 min 39 s. On May 4 I could not measure its passing the line because it disappeared before reaching it. However, since the belt passes every day 45 min 46 s earlier, I could figure out that the satellite passed the "true point" on May 4 at 21 h 41 min 39 s + 45 min 46 s = 22 h 27 min 25 s. Therefore, it disappeared on May 4 4 s after passing the true point (before getting to the line) and on May 5 44 s after passing it. The shade retreated by 40 s in 23.2 h or 4 s in 2.3 h. The shade border was about 4 s from the true point at around 22:30; therefore, at around 20:00 it was at the true point.

We can substitute for t in (1.14) the difference between the astronomical midnight (that was at 23.63 that day) and the belt passing above us at 21.94. Correcting for the precession (see Sect. 1.1.6.1) we add 0.02 h and get 21.96 as the time of the belt passing overhead.

To summarize, on May 4, 1962: $t = 23.63 - 21.96 = 1.67$h; $\delta = 15°59' + 74' = 17°13'$ ($74'$ is the correction of the sun's rays refraction); $\phi = 32°03'$.

By (1.14) we get $\frac{R}{H} = 0.817$. Since $R = 6370$ km we get $H = 7800$ km and $h = H - R = 1430 \pm 10$ km.

1.1.8 The Satellite's Entrance to a "Long Day"

The satellite enters a "long day" when its orbit does not enter the sun's shade. There are usually two solutions to (1.16) one when the satellite enters the shade and one when it exits it. A "long day" starts or ends when the enter and exit points are the same.

Let $p = \sqrt{1 - \left(\frac{R}{H}\right)^2}$. Solving (1.14) for $\cos(15t)$ yields

$$\cos(15t) = \frac{\sin\phi \sin\delta + p}{\cos\phi \cos\delta}. \tag{1.18}$$

The belt in the precession motion creates a 3D surface. Equation (1.18) represents all the points of the different belts in which the satellite enters the sun's shade. This is, therefore, the line formula of the intersection between the surface and the shade perimeter.

The belt formula in a particular moment is given by Eq. (1.1): $\cos(7.5t_1) = \tan\phi \cot\gamma$. The time between the satellite passing at its northernmost point and latitude ϕ is $\frac{t_1}{2}$. t in Eq. (1.18) is the time from midnight and passing over latitude ϕ. Therefore, $t - t_0 = \frac{t_1}{2}$, where t_0 is the time between midnight and its passing the northernmost point. By (1.1)

$$\cos(15(t - t_0)) = \tan\phi \cot\gamma. \tag{1.19}$$

The derivatives of (1.18) and (1.19) are

$$-15\sin(15t)\frac{dt}{d\phi} = \frac{\sin\delta\cos^2\phi\cos\delta + \cos\delta\sin\phi(p + \sin\phi\sin\delta)}{\cos^2\phi\cos^2\delta}$$
$$-15\sin[15(t - t_0)]\frac{dt}{d\phi} = \frac{\cot\gamma}{\cos^2\phi}.$$

When the satellite enters a "long day" its orbit is tangent to the surface of the sun's shade. The belt (1.19) and the intersection curve (1.18) are tangent to one another. Therefore, $\frac{dt}{d\phi}$ is the same in the two equations. Using $\sin(15t)$ from (1.18) and $\sin(15(t - t_0))$ from (1.19), solve for $\frac{dt}{d\phi}$ in both and equate them, we get

$$\sqrt{\cos^2\phi - \sin^2\phi\cot^2\gamma}(\sin\delta + p\sin\phi)$$
$$= \cot\gamma\sqrt{\cos^2\phi\cos^\delta - p^2 - 2p\sin\phi\sin\delta - \sin^2\phi\sin^2\delta}. \tag{1.20}$$

Extensive algebraic and trigonometric derivations using, for example, the identity $\cos^2\phi - \sin^2\phi\cot^2\gamma = \frac{\cos^2\phi}{\sin^2\gamma} - \cot^2\gamma$, lead to the quadratic equation for $\sin\phi$:

$$p^2\sin^2\phi + 2p\sin\delta\sin\phi + \sin^2\delta - cos^2\gamma\left(\frac{R}{H}\right)^2 = 0$$

whose solutions are

$$\sin\phi = \frac{\pm\cos\gamma\frac{R}{H} - \sin\delta}{\sqrt{1 - \left(\frac{R}{H}\right)^2}}. \tag{1.21}$$

Using (1.14) we get (1.18). Equation (1.18) indicates that the satellite will enter a "long day" when the belt passes over latitude ϕ t hours after midnight at the point under it.

1.1.8.1 Measuring Echo-1 Entering a Long Day

We first determine by (1.21) the latitude ϕ_1 at which Echo-1 enters the "long day." We calculate it for the end of June near the summer solstice. At this time $\delta = 23.2° + 1.2° = 24.4°$. Using $\frac{R}{H} = 0.82$, $\gamma = 47.3°$ we get $\phi_1 = 14.5°$.

We now calculate t by (1.18). t is the time difference between midnight at the point with latitude ϕ_1 and the time the belt passes over it. When the satellite enters a "long day," this difference is the same for all points at latitude ϕ_1 if we ignore the precession. We therefore select the point with latitude ϕ_1 at our longitude. We get by (1.18) $t = 2.66$ h. The precession for this time period is $\frac{2.66 \times 0.277}{24 - 0.277} = 0.03$. We therefore use $t = 2.63$. In conclusion, the condition for Echo-1 entering a "long day" is that it passes over latitude 14.5° and our longitude 2.63 h before midnight, which is at 23.70 h, meaning at 21.07 in the evening.

Equation (1.1) determines t_1 which is the time the belt crosses latitude ϕ_1 until the satellite gets to its northernmost point. For $\phi = 14.5°$ and $\gamma = 47.3°$ we get $\frac{t_1}{2} = 5.08$ h. The precession correction is 0.06 h so we use $\frac{t_1}{2} = 5.02$.

Rather than finding when the belt passes $\phi_1 = 14.5°$ at 21.07, Echo-1 enters the "long day" when the northernmost point crosses our longitude at $21.07 + 5.02 = 26.09$ (2.09 on next day morning). The times the northernmost point passed to our north are:

June 28, 1962 (evening) 26.86 which is June 29, 1962, 02.86 morning.
June 29, 1962 (evening) 26.59 which is June 30, 1962, 02.59 morning.
June 30, 1962 (evening) 26.31 which is July 1, 1962, 02.31 morning.
July 1, 1962 (evening) 26.03 which is July 2, 1962, 02.03 morning.
July 2, 1962 (evening) 25.76 which is July 3, 1962, 01.76 morning.

In conclusion, Echo-1 was expected to enter a "long day" on July 1, 1962. This was confirmed by observations.

1.1.9 Determining the Average Height of a Satellite

In order to find the average distance from the earth center $\overline{H}$, we apply the third Kepler's law (Russell 1964)

$$\frac{\overline{H}^3}{T^2} = c, \tag{1.22}$$

where c is a constant for satellites rotating the earth. We can determine c by these values for the moon. The moon's mass cannot be ignored compared with the earth's mass. Therefore, the moon's mass should be considered. We apply the moon's mass being $\frac{1}{81}$ of the earth's mass, the moon's period is 39,343.2 min, and its average distance is 384,393 km. We get $\log \sqrt[3]{c} = 2.519755$.

$$\overline{H} = \sqrt[3]{c}\sqrt[3]{T^2} = 330.94\sqrt[3]{T^2} \tag{1.23}$$

In order to determine the change in the average height $\Delta\overline{H}$ by the change in the period ΔT we get by the Taylor series $\frac{\Delta\overline{H}}{\overline{H}} = \frac{2}{3}\frac{\Delta T}{T} - \frac{1}{9}\left(\frac{\Delta T}{T}\right)^2 + \frac{4}{81}\left(\frac{\Delta T}{T}\right)^3 - + \ldots\ldots$ For small changes it is sufficient to use the approximation

$$\Delta\overline{H} = \frac{2\overline{H}}{3T}\Delta T = \frac{\frac{2}{3}\sqrt[3]{c}}{\sqrt[3]{T}}\Delta T = \frac{220.63\Delta T}{\sqrt[3]{T}}. \tag{1.24}$$

1.1.9.1 Measuring the Average Height of Echo-1

On May 5, 1962, Echo-1 retreated 45.76 min per day and on May 12, 1962, it retreated 45.82 min per day. The retreat increased by 0.06 min in 7 days and each day it completes about 12 periods and therefore the period decreased in 7 days by $\frac{0.06}{12} = 5 \times 10^{-3}$ min. The period decreased each orbit compared with the previous one by $\frac{5\times10^{-3}}{84} = 6 \times 10^{-5}$ min.

Using $\overline{H} = 7.9 \times 10^6$ meters and $T = 116$ min, by Eq. (1.24):

$$\Delta\overline{H} = \frac{2 \times 7.9 \times 10^6}{2 \times 116} \times 6 \times 10^{-5} = 3\,\text{m}.$$

During this time period Echo-1 went down 3 m each orbit.

Extensive observations and calculating the decline per period over almost 3 years are depicted in Fig. 1.8. The graph is not so smooth in 1962. This may be due to the improvement in the accuracy of the measurements. It is surprising that in some

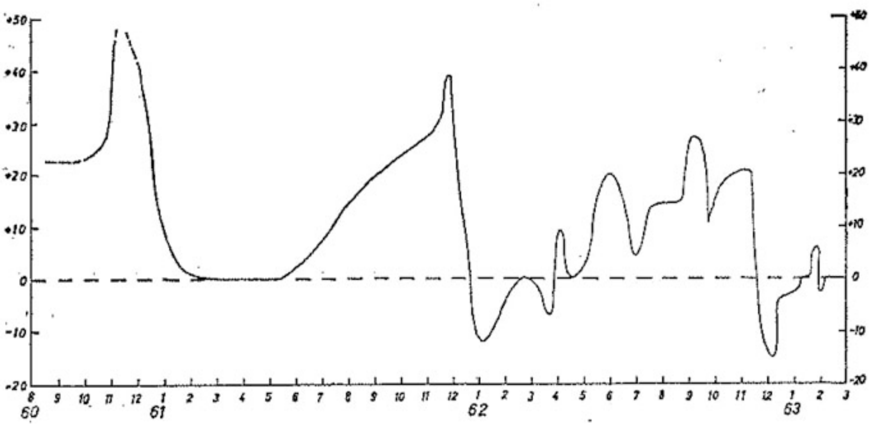


Fig. 1.8 Satellite's decline in meters/period

months Echo-1 gained altitude. Is it related to the number of meteors? or may be to sun activity?

1.2 Post-Doctoral Activity

I applied for Post-Doc positions and accepted an offer from George Wesolowsky at McMaster University, Hamilton, Canada. The 2 years working with George were my most influential and constructive educational experience. George was an outstanding advisor. He taught me how to design research projects and report them in papers. Following his wonderful advice and support I was prepared to continue and do independent research, and write papers on my own. During those 2 years George and I were able to produce about 15 papers, many appeared later on. George had a wealth of ideas mostly on location models.

1.3 Non-location Research

Due to my training in mathematics I was able to design algorithms and solve problems in a variety of fields. I summarize some of these papers especially if they are useful to a general audience and not mentioned in other chapters.

1.3.1 The Fading Universe Theory

My early interest in Astronomy triggered an idea which replaces the "Big Bang Theory" with an alternative explanation. The red shift phenomenon which is the impetus for the theory can be explained by declining speed of light over the years (Drezner 1984a). Calculations based on available data suggest that a decrease in the speed of light by 1 km/s every 60,000 years will cause the red shift that we observe. It is interesting that the same decline in the speed of light also explains the rate of increase of the length of the day.

1.3.2 Statistics Topics

1.3.2.1 Multivariate Normal Probabilities

In one of my first papers with George Wesolowsky (Drezner and Wesolowsky 1981) we needed the calculation of bivariate Normal probabilities. In Drezner (1978) the probabilities were calculated by Gaussian Quadrature formulas (Abramowitz and

Stegun 1972). George refused to add his name as a co-author claiming that he did not contribute to it. A follow-up paper with George (Drezner and Wesolowsky 1990) was published later. Papers on the calculation of trivariate and multivariate Normal probabilities are Drezner (1992a, 1994). Anklasaria and Drezner (1986) applied the calculation of the multivariate Normal probabilities to estimate the completion time of a project consisting of a list of activities. Common approaches consider only the sequence of activities leading to the largest expected completion time. However, since completion times are normally distributed, sequences with lower expected completion times may take the longest time to complete.

1.3.2.2 Selecting Variables in Multiple Regression

Commonly used statistical approaches are forward selection, backward elimination, and step-wise selection. Said Salhi suggested tabu search for the selection process (Drezner et al. 1999). Tabu search handily beat the commonly used approaches. There are follow-up papers on variable selection with variations on this idea: applying simulated annealing (Drezner and Marcoulides 1999); applying ant colony optimization (Marcoulides and Drezner 2003); solving financial applications with the variable selection method (Drezner et al. 2001); adding resource constraints to the variable selection problem (Marcoulides and Drezner 2004).

1.3.2.3 Multirelation

Drezner (1995c) defines a multirelation between several variables as $1 - \lambda(R)$, where $\lambda(R)$ is the smallest eigenvalue of the correlation matrix. The multirelation is between 0 and 1. For two variables the two eigenvalues are $1 - \rho$ and $1 + \rho$. The smaller eigenvalue is $1 - |\rho|$. Therefore, the multirelation is equal to $|\rho|$. Dear and Drezner (1997) analyzed the significance level of the multirelation.

1.3.2.4 Normality Test

Drezner et al. (2010b) improved the Kolmogorov–Smirnov test for normality (Massey Jr. 1951). In the current implementation of the Kolmogorov–Smirnov test, given data are compared with a Normal distribution that uses the sample mean and the sample variance. Drezner et al. (2010b) proposed to select the mean and variance of the Normal distribution that provide the best fit to the data.

Drezner and Turel (2011) analyzed the level of a too-frequent value in data (such as many zeros), for which transformation to normality that passes tests for normality is impossible. Analysts and researchers are often concerned with the question: should we bother transforming a variable to normality, or should we revert to other approaches not requiring a Normal distribution? Drezner and Turel (2011) found the critical number of the frequency of a single value for which there is no feasible transformation to normality within a given α of the Kolmogorov–Smirnov test.

1.3.2.5 Discretizing Continuous Distributions

Drezner and Zerom (2016) proposed a generally applicable discretization method to approximate a continuous distribution with a discrete one. The method adopts a criterion which is shown to be flexible in approximating higher order features of the underlying continuous distribution while preserving its mean and variance. Discretizing bivariate distributions is proposed in Drezner et al. (2018e).

1.3.2.6 Correlated Binomial and Poisson Distributions

A colleague, Herb Rutemiller, made a passing comment in the corridor. The distribution of students' exam grades seems to be bi-modal and does not follow a binomial distribution. Nick Farnum and I thought of a "generalized binomial distribution" (GBD). Consider a sequence of Bernoulli experiments where the initial probability of success is p but the probability changes as successive experiments are done. If the first experiment is a success, the probability that the second experiment is a success increases, and if it is a failure it decreases (Drezner and Farnum 1993). If students take a multiple choice exam, this is exactly what happens because of skill. We found the same distribution for the number of wins of baseball teams at the end of the season. Not all teams are "equally skilled." The distribution of the number of wins is not binomial. There are 162 games in a season and the probability of a win is $p = 0.5$. The mean of the distribution is 81 but the distribution itself is not binomial and its variance is greater than $np(1-p) = 40.5$.

An association factor similar to the correlation coefficient θ is given. $\theta = 0$ yields the "standard" binomial distribution. Suppose that in n experiments, the number of successes is r. The probability of success in the next experiment is $(1-\theta)p + \theta\frac{r}{n}$. The mean of the distribution is np, the same as the binomial distribution, but the variance is $p(1-p)\frac{n-\frac{1}{B(n,2\theta)}}{1-2\theta}$. We found that in baseball games $\theta = 0.397$. For complete details see Drezner and Farnum (1993).

In a follow-up paper (Drezner 2006b) it is proven that for $\theta \leq 0.5$ the limit to the GBD, as the number of trials increases to infinity is the Normal distribution. In other cases it can be bi-modal. It was also found, by analyzing real data, that the grade distribution of 1023 exams yielded $\theta = 0.5921$ and the number of wins of NBA teams at the end of the season yielded $\theta = 0.5765$.

In an interesting recent paper, the success of mutual funds over the years is analyzed by the generalized binomial distribution. It is shown that the performance is not random, i.e., it is skill and not luck (Bhootra et al. 2015).

For $p \to 0$ and $\lambda = np$, the generalized binomial distribution defines the correlated Poisson distribution (Drezner and Farnum 1994). The mean of the correlated Poisson distribution is λ but the variance is $\frac{\lambda}{1-2\theta}$ for $\theta < 0.5$ and is infinite for $\theta \geq 0.5$. As examples Drezner and Farnum (1994) used the expected number of demand points which are local optima for the Weber location problem on the sphere using the data in Drezner (1989b). For this data set $\theta = 0.455$. The distribution of 402 sow bugs (Consul 1989) yielded $\theta = 0.5533$ which means that the bugs tend to cluster on the skin of a sow.

1.3.3 Queueing

In Drezner (1999) exact formulas for parameters of the queuing system when the arrival process is not Poisson but correlated Poisson (Drezner and Farnum 1994) are given. In a correlated Poisson distribution the variance of the arrival rate, σ_λ^2, can be different from λ. It was shown that the L_q (expected length of the queue) is

$$L_q = \frac{\lambda^2\sigma^2 + \left(\frac{\lambda}{\mu}\right)^2 + \frac{\sigma_\lambda^2 - \lambda}{\mu}}{2\left(1 - \frac{\lambda}{\mu}\right)}$$

which is the Pollaczek's formula (Gelenbe et al. 1998) with a correction for σ_λ^2, the variance of the arrival rate. In Poisson arrival $\sigma_\lambda^2 = \lambda$ and the formula reduces to Pollaczek's formula. For details see Drezner (1999).

Drezner and Zerom (2018) investigated a tandem queuing system of two single servers with correlated service times. The effect of positively correlated service times on system performance is examined. Using an intuitive dependence model for service times, a new analytically tractable formula for the total expected waiting time in the system is found. Positive correlation improves system performance due to a reduction in the expected waiting time in queue for the second server.

Drezner et al. (1990) presented a queueing-location problem where a location of a service station has to be determined. The objective is to minimize the sum of the total travel time plus delay at the service station. The two main results of this article are a convexity proof for general distances and a theorem that limits the area in the plane where the solution can lie. Some solution procedures are proposed. Follow-up papers that consider multiple servers are (Berman and Drezner 2007; Aboolian et al. 2009; Drezner and Drezner 2011a).

An efficient method for calculating the parameters of an M/M/k queueing system is presented in Pasternack and Drezner (1998). It is especially effective when the probability of zero customers in the system, P_0, is very small. Floating point representation may exceed the limit of a negative power and be calculated as 0. The expressions in standard formulas have a P_0 multiplier (Gelenbe et al. 1998; Gnedenko and Kovalenko 1989) and fail to find the desired results. The approach is to calculate v_k for a given k, λ, and μ by the following sequence: $v_1 = \frac{\mu}{\lambda}$ and $v_{i+1} = \frac{(i+1)\mu}{\lambda}(v_i + 1)$ obtaining v_k. The expected length of the queue is

$$L_q = \frac{\lambda k\mu}{(k\mu - \lambda)[(k\mu - \lambda)v_k + k\mu]} = \frac{\rho}{(1-\rho)[(1-\rho)v_k + 1]},$$

where $\rho = \frac{\lambda}{k\mu} < 1$. The rest of the M/M/k queue parameters can be calculated.

1.3.4 Finding Whether a Point is in a Polygon

Drezner (1998a) designed an algorithm to check whether a point is inside a polygon or not. The polygon does not have to be convex. It is represented by a sequence of consecutive vertices ending at the first vertex. The angles between successive vertices are calculated. If the sum of the angles is $\pm 2\pi$, the point is inside the polygon. If the sum is zero, it is not. It is interesting that the proof is based on complex number theory (Carathéodory 2001). Every point on the plane (x, y) is represented as a complex number $z = x + iy$ and the integral $\oint \frac{1}{iz}dz$ over the circumference of an area, such as a polygon, is equal to $\pm 2\pi$ if the "pole" (0, 0) is inside the area and zero if it is not. I do not think that there is another operations research analysis based on complex number theory. For complete details see Drezner (1998a).

1.3.5 Information Dissemination on a Network

Amnon Barak, one of my computer science colleagues, described the following situation. Many (can be millions) computers are connected in a network and one computer wishes to disseminate some information to all of them. He proposed that the originating computer selects another computer at random and sends it the message. Every time unit it selects another computer at random and sends it the information. Every computer that received the information repeats the process by randomly selecting a computer every time unit, sends it the message, and so on. How long will it take to get the information to all computers? Even today, computer scientists use a tree network and every computer sends the information to two preassigned computers rather than random dissemination. For n computers, it takes $\log_2 n$ time units by applying a tree configuration. Drezner and Barak (1986) proved that with high probability all computers will get it in about 1.7 times this value. However, if some computers are inactive, random dissemination is not affected while a tree configuration is. To handle the possibility of failed computers, computer scientists use a complicated acknowledgment system to verify that the target computer indeed got the message which more than doubles the time because a computer waits for some pre-specified time to be "sure" that there is no acknowledgment coming. Amnon has difficulty selling this approach to computer scientists because they expect 100% guarantee and if there is, say, 10^{-16} probability that not all active computers got it, it is not acceptable to them. We have many follow-up papers with his colleagues and Ph.D. students (Hoefler et al. 2017; Barak et al. 1986; Amar et al. 2009; Barak et al. 2015).

1.3.6 Genetic Algorithms

In the late 1990s Said Salhi visited us in California and introduced me to metaheuristics, mainly tabu search (Glover and Laguna 1997; Glover 1977, 1986), simulated annealing (Kirkpatrick et al. 1983), and genetic algorithms (Holland 1975; Goldberg 2006). I was especially fascinated by genetic algorithms because the design of the specific algorithm depends on the problem to be solved.

A population of solutions is maintained. A starting population of members is established. In the basic genetic algorithm the following is repeated for a given number of generations:

1. Two population members are selected as parents.
2. The two parents are "merged" (mate) and produce an offspring.
3. If the offspring is better than the worst population member and is not identical to another population member, it is added to the population and the population member with the worst objective value is removed. Otherwise, the population remains unchanged.

The best member of the final population is the result of the algorithm.

The first problem that I tried to solve was the one-way two-way network design suggested by George Wesolowsky (Drezner and Wesolowsky 1997b). A network of two-way roads is given. Some of the roads can be designed as one-way increasing the speed on these roads. What is the best combination of one-way roads for a given traffic flow between nodes? The most crucial part of a genetic algorithm is the merging process. Selecting at random links from each parent does not exploit the special structure of this problem. Drezner and Salhi (2002) suggested to draw a virtual line and take the structure of one parent from one side of the line and the structure of the other parent for the other side. If we are "lucky," each half is a good configuration for its links. Once an offspring is generated this way, a descent algorithm is applied to fix the border links between the two parents. This intuitive approach proved to be very successful and similar schemes were applied for the quadratic assignment problem (Drezner 2003) and for the p-median on a network (Alp et al. 2003).

I published with co-authors many modifications of the basic genetic algorithm over the years. Drezner and Marcoulides (2003) suggested that rather than randomly selecting two parents to mate, only the first parent is randomly selected. K (a parameter) potential mates are then randomly selected and the one who is the most dissimilar to the first parent is selected as a second parent. Drezner and Drezner (2006) suggested to partition the population to males and females and mate one male and one female. Drezner and Drezner (2018c) suggested to randomly select two population members and select the better one with a pre-specified probability as the first parent. The second parent is then randomly selected. Drezner and Drezner (2019d) proposed the alpha male parents selection rule. A pre-specified number of alpha males is selected and the rest of the population are females. In a generation, each female selects an alpha male and produces one offspring. Drezner (2005b) suggested a different rule for removing population members. Drezner (2005a) proposed to create small populations, evolve them by applying a genetic algorithm on each one separately, and then combine them into one population and continue to run the genetic algorithm. Drezner and Misevicius (2013) suggested to apply every generation an improvement algorithm on a randomly selected population member. If the population member is improved, it is replaced by the improved one. Then, proceed with selecting two parents, etc.

1.3.7 Inventory and Supply Chain

My most cited paper (well over 2000 citations) is about the bullwhip effect (Chen et al. 2000) in supply chain. David Simchi-Levi approached me in one of the conferences and asked if I can develop an analytic expression for the ratio between two variances which is termed the bullwhip effect. I came up with a closed form expression for the ratio. David and his Ph.D. student Jeniffer Ryan compared it to simulation results and found a perfect fit. Other inventory related papers include: inventory models with two products that one can substitute for the other if the need arises (Pasternack and Drezner 1991; Drezner et al. 1995); the videotape rental model (Pasternack and Drezner 1999); models that combine inventory control and location analysis (Drezner et al. 2003; Drezner and Scott 2013).

1.3.8 Robot Design

Shimon Nof suggested the problem of a robot arm that needs to pick items from bins and place them in a set of destinations. The objective is to minimize the total distance going back and forth by the robot's arm. This can be modeled as a traveling salesman problem with a stipulation that the salesman must alternate between bins and destinations. All pairs of bins and all pairs of destinations are assigned large distances between them to force the salesman to alternate between bins and destinations (Drezner and Nof 1984). If each part in a bin must be placed in a particular destination, the distance from a bin to other destinations is set to a large number as well. Follow-up papers are Nof and Drezner (1986, 1993).

1.3.9 Linear Inequalities

Drezner (1983b) considered the problem of finding a feasible solution to a set of linear inequalities. There is a ball that can be determined a priori from the problem data with the property that it contains a feasible solution if there is one. The relaxation method (Agmon 1954; Motzkin and Schoenberg 1954) constructs a sequence of shrinking balls that contain a feasible solution if there is one. It is proven in Drezner (1983b) that if one of the smaller balls is nested inside the first ball, then there is no feasible solution to the problem. This principle proved to be superior to the existing stopping criterion. It is also shown that the principle cannot be extended to the Russian method for linear programming (Khachian 1979).

1.3.10 The Repeated Partition Scheduling Problem

This scheduling problem was motivated by the conference site in Oberwolfach, Germany. There are several tables in the dining room and the seating arrangement is changed every meal so that every attendant meets all other attendants during the conference. nk people are partitioned d times into k groups of n people each (Drezner 1998b). It is required that each person meets every other person (being in the same group) at least once. The objective is that each person meets with other participants about the same number of times. There are other applications such as arranging a golf tournament, personnel assignment, testing drugs, TV commercials, etc. Goldstein and Drezner (2007) solved the repeated partition scheduling problem by three metaheuristics: simulated annealing, tabu search, and genetic algorithms.

1.3.11 Solving Non-convex Programs by Sequential Linear Programs

Drezner and Kalczynski (2018) found that non-convex programs such as the maximization of a convex objective subject to constraints which are outside of convex regions can be heuristically solved by a multi-start approach based on solving a sequence of linear programs (termed MSLP). The concept of "Sequential Linear Programming" (SLP) was introduced as early as the 1960s (Courtillot 1962) but most attempts report convergence issues necessitating the ad hoc establishment of moving limits and were not that efficient (John et al. 1987; Chen 1993). It is interesting that this general approach is suitable for solving non-convex problems but is not suitable for solving convex problems. Many non-convex problems are solved by global optimization methods which exploit the special structure of the problem and usually require excessive processing times that restrict the size of the problems that can be practically solved. This iterative approach is faster than a direct multi-start approach and provided better results on four test problems (p-dispersion or circle packing in a square, a newly defined p-dispersion covering problem, and their three dimensional variants) for a total of 116 instances. A follow-up paper is Kalczynski and Drezner (2019c). When the functions of a general non-linear program, both the objective function and all the constraints, can be expressed as a difference between two convex functions, one of the convex functions can be replaced by a tangent plane at the present iteration point. The modified problem is convex. The optimal solution to this modified program is the next iteration point. The procedure is demonstrated on the special case when all functions are quadratic (not necessarily convex).

Kalczynski and Drezner (2019a) propose to apply the MSLP algorithm (Drezner and Kalczynski 2018) for the solution of two multiple obnoxious facility problems on the plane. One problem considers obnoxious facilities and the other obnoxious demand points. For obnoxious facilities the objective is maximizing the total distance between demand points and their closest facility. For obnoxious demand points the objective is maximizing the total distance between facilities and their closest demand point. The algorithm is compared with the interior point and SNOPT solvers in Matlab. The interior point solver performed quite poorly. The solutions by SNOPT are of comparable quality to the MSLP algorithm but the MSLP algorithm required much shorter computer times. The largest problem of locating 20 facilities among 1000 demand points was solved by MSLP in 2 min compared with 40 min by SNOPT.

1.3.12 Contributions to Education

The examples listed below can be used, and some were used, in class to demonstrate various concepts.

In an advanced class I created an example of building a global FEDEX-like hub. The 20 largest metropolitan areas in the world are given in a table. If we want to deliver packages from each city to the other 19 cities we need a fleet of 380 planes. However, if we find a hub so that all planes fly to the hub exchange packages and fly back, we need only 20 planes. The objective is to minimize the maximum distance to the hub so that the total time from source to destination is minimized. I created an Excel file that solves the problem, then I switch to Google Maps or MapQuest and find the location by entering the latitude and longitude of the solution. The hub (that happens to be in the Mediterranean) is shown on the map. We also find the location that minimizes the weighted sum of distances which minimizes total cost and its actual location, happens to be in Russia.

To demonstrate the concepts of convexity and local optima I created an Excel file (with VBA) that solves ten times packing six circles in a square with the maximum possible radius. The program shows the randomly generated initial solution (usually small circles with a diameter equal to the minimum distance between the pairs of points). The program waits for 3 s and the final configuration shows up. After 10 starting solutions, the optimal solution is usually found about 3 times out of 10.

Estrada and Drezner (2006) simulates the spread of a virus by an Excel program and VBA. I suggested to my student Jay Estrada to submit it to the *INFORMS Journal of Education* and we got very flattering reviews about the "neatness" of the demonstration.

I show in class how to solve the PERT/CPM by hand and then I had an Excel file that takes the data and finds "instantly" the solution (no VBA). A student, Kevin Quinn, constructed a much nicer Excel file that basically produces the same results but the data does not have to be organized in a certain way. It is published in Quinn and Drezner (2017).

Drezner and Erickson (1997) showed statistically, based on real data, that the common belief that "if the market goes up in January, it will go up for the whole year" is indeed likely. Since January is part of the year, the two events are correlated and even by random chance it happens 60% of the time. However, January also "predicts" what will happen in the next 11 months as well. The p-value is 0.002. It turns out that April and August are also good predictors (p-value $= 0.04$). In October: if the market goes in one direction, it will go the opposite direction in the next 11 months (p-value $= 0.06$ which is borderline).

Drezner (2001) calculated the maximum possible EVPI for any probability distribution for the states of the world. This maximum EVPI is an upper bound for the EVPI with given probabilities and thus an upper bound for any partial information about the likelihood of each state of the world.

Drezner and Minh (2002) considered the problem of planning a mix of products with a limited budget constraint. This problem is an extension of the well-known product mix problem. The problem is easily formulated as a linear programming problem. The optimal solution is found by an explicit formula without applying any linear programming solution method. Therefore, very large problems can be easily solved.

1.3.13 Programming "Tricks"

1.3.13.1 Breaking Ties

When coding descent algorithms or Tabu search, a list of solutions with corresponding objective functions is generated and the "best" one selected. Which one should be selected if there are ties? It is best to select one at random not to bias the search by always selecting the first one or the last one. For years I was saving the list of all tying solutions and eventually randomly selecting one from the list.

I found an efficient way of doing it: The kth tying move replaces the selected move with probability $\frac{1}{k}$. By this rule, the first move (or a new best found move) is always selected ($k = 1$). When a tying move is found, it replaces the selected move with probability $\frac{1}{2}$, and so on. When the process ends and K moves are tied for the best one, each of the tying moves is selected with probability $\frac{1}{K}$ (Drezner 2010b).

I thought that it must be known. I met Fred Glover at an INFORMS conference and casually told him about it. He really liked it. He said that he struggled with this problem for years. I was convinced to write it down and submit it.

1.3.13.2 Calculating Mean and Variance

Another programming "problem" is the calculation of the mean and variance of many values (may be even millions). If we calculate the sum of the values and their sum of squares, the formula for the variance may yield a large and unacceptable rounding error. So I came up with the following idea which is easy to implement. Let x_k be the kth value, m_k the average of all values $x_1, \ldots, x_k$, and V_k be the variance of these values.

1. Set $m_0 = V_0 = 0$.
2. For each $k = 1, \ldots$ calculate,
   (a) $\Delta_k = \frac{1}{k}(x_k - m_{k-1})$.
   (b) $m_k = m_{k-1} + \Delta_k$.
   (c) $V_k = \frac{k-1}{k}\left(V_{k-1} + k\Delta_k^2\right)$.

Note that for every k the mean and variance of all the values so far are known and the process can be stopped at any point. This is especially useful if a pre-specified standard error is required.

Step 2c can be adjusted to calculate directly the standard error. Let S_k be the standard error of the first k values. Since $V_k = kS_k^2$, $kS_k^2 = \frac{k-1}{k}\left[(k-1)S_{k-1}^2 + k\Delta_k^2\right]$ which is

$$S_k^2 = \frac{k-1}{k}\left[\frac{k-1}{k}S_{k-1}^2 + \Delta_k^2\right] .$$

1.3.13.3 Variance Reduction

A simulation of a function $F(X)$ when X is a Normal variable with a mean μ and variance σ^2 is performed. The function $F(X)$ may be a result of an optimization problem with a parameter X and may not have an explicit expression. It may take a relatively long time to calculate each value. The common way to perform such simulation is to generate random Normal variates, calculate $F(X)$ for each, and calculate the mean of these values and their standard error. It is likely that the simulated values have a significant variance. We propose to reduce such variance by the following modification. Generate a random standard Normal variate z, evaluate $F(\mu + z\sigma)$ and $F(\mu - z\sigma)$, rather than generating two standard Normal variates z_1, z_2. This modification reduces significantly the standard error of the mean, thus requires fewer evaluations of $F(X)$ to get the same standard error. The accuracy of the mean is not compromised because if a large number of calculations of $F(X)$ are performed, both z and $-z$ are equally likely to be selected.

1.3.13.4 Generating Random Numbers

Generating random points in a square so that it can be replicated for comparison purposes in other papers can be done by generating sequences for the x-coordinates and y-coordinates by the method proposed in Drezner et al. (2018g, 2019c). A sequence r_k of integer numbers in the open range (0, 100,000) is generated. A starting seed r_1, which is the first number in the sequence, and a multiplier λ which is an odd number not divisible by 5, are selected. We used $\lambda = 12{,}219$. The sequence is generated by the following rule for $k \geq 1$:

$$r_{k+1} = \lambda r_k - \lfloor \frac{\lambda r_k}{100{,}000} \rfloor \times 100{,}000.$$

For example, the x coordinates were generated by $r_1 = 97$ and for the y-coordinates $r_1 = 367$. The coordinates are then divided by 100,000. The sequence of coordinates can be easily constructed in Excel. Enter r_1 into cell A1. In cell A2 enter `= A1*12,219 - INT(A1*0.12219)*100,000` and copy it down in column A. Column A can be replicated in column B for the y coordinates by replacing r_1 in cell B1. These sequences can be used to generate up to 5000 points because $r_{5001} = r_1$.

1.4 Location Papers

I have location papers in many fields of location analysis. Many of these papers are discussed in other chapters. I discuss below classes of papers that are not extensively discussed in other chapters.

1.4.1 Objective Functions

1.4.1.1 Variations on the Weber Problems

The objective of the Weber problem (Weber 1909) is to find the best location X for a facility which satisfies

$$\min_X \left\{ \sum_{i=1}^{n} w_i d_i(X) \right\},$$

where n is the number of demand points, w_i is the weight associated with demand point i, $d_i(X)$ is the distance between demand point i and the facility location X. For a review of the Weber problem see Wesolowsky (1993) and Drezner et al. (2002b).

Drezner et al. (2009c) investigated a model where travel time is not necessarily proportional to the distance. Every trip starts at speed zero, then the vehicle accelerates to a cruising speed, stays at the cruising speed for a portion of the trip, and then decelerates back to a speed of zero. A time equivalent distance which is equal to the travel time multiplied by the cruising speed is defined. It is proved that every demand point is a local minimum for the Weber problem defined by travel time rather than distance.

For example, using a cruising speed of 900 km/h, flying from LA to Las Vegas (about 400 km) should take less than half an hour using standard models (400/900 h). Taking into account acceleration and deceleration the time is almost doubled which is very close to the actual flying time. On the other hand, flying from LA to NY (about 5000 km) is underestimated by existing models by only 8%.

Drezner and Wesolowsky (1989) considered the Weber problem when the distance from point A to point B is not the same as distance from B to A. This is common in rush hour traffic or for flights that in one direction have tail winds and in the opposite direction have head winds. Drezner and Drezner (2018a) analyzed location problems where the distance (time) to get to the destination by air is affected by winds. Two models are proposed: the asymmetric Weber location problem and the round trip Weber location problem.

These observations are especially important when evaluating hub location models (Skorin-Kapov et al. 1996; O'kelly 1987; Campbell 1994; Contreras 2015). When a flight has one or more stopovers, the underestimation of flight time is increased (in addition to the waiting time at the stopover) because there are more acceleration and deceleration periods. Therefore, when using distances rather than flight times, hub location models may underestimate the decline in quality of service due to stopovers at hubs.

Drezner et al. (2016c) considered the Weber problem on a network where demand points can be on nodes of the network or anywhere in the plane off the network. Distances for demand points located on nodes can be either network distance or Euclidean distance, while distances to points off the network are Euclidean. Travel time on the network is slower by a given factor. Applications include building a hospital providing emergency services either by an ambulance using network roads or by a helicopter flying directly to the patient, especially when the patient is off the network. The facility can be located anywhere on the network and the optimal solution is not necessarily on a node. The problem is optimally solved by the "Big Segment Small Segment" global optimization algorithm (Berman et al. 2011a).

Drezner (1979) considered the Weber problem when there is uncertainty in the location of demand points. Each demand point can be located anywhere in a circle. The set of all possible optimal points is found.

Farahani et al. (2009) investigated the Weber problem with multiple relocation opportunities. The weight associated with each demand point is a known function of time. Relocations can take place at predetermined times. The objective function is to minimize the total location and relocation costs.

Drezner and Scott (2010) considered the following problem. A facility needs to be located in the plane to sell goods to a set of demand points. The cost for producing an item and the transportation cost per unit distance are given. The planner needs to determine the best location for the facility, the price charged at the source (mill price), and the transportation rate per unit distance to be charged to customers. Demand by customers is elastic and assumed to decline linearly in the total charge.

Drezner and Scott (2006) considered the location of a facility in the plane when service availability is a convex decreasing function of the distance (distance decay). The total cost of the system consists of three components: (1) the cost of waiting in line for service by the M/M/1 queueing model, (2) the cost of providing the service, and (3) the cost of lost demand. A generalized Weiszfeld algorithm (Drezner 2009) and the Big Triangle Small Triangle (BTST) global optimization technique (Drezner and Suzuki 2004) are applied and tested.

Drezner (1985b) found the sensitivity of the optimal site to the Weber problem to changes in the locations and weights of the demand points. An approximate formula for the set of all optimal sites is found when demand points are restricted to given areas and weights are within given ranges.

Drezner and Goldman (1991) found the smallest set of points that may include at least one optimal solution to the Weber problem for a given set of demand points and any unknown set of weights.

Drezner and Scott (1999) found the set of feasible solution points to the Weber location problem with squared Euclidean distances when the weights are limited to intervals. The result is then used to solve the minimax regret objective when individual scenarios can be any set of weights in a given set of intervals.

David Simchi-Levi posed the question: "can it be shown that when the number of demand points increases, the probability that the Weber solution point is on a demand point converges to 1?" Intuitively, there is no space left between demand points when the number of demand points increases to infinity. Drezner and Simchi-Levi (1992) showed that when n demand points are randomly generated in a unit circle, with randomly distributed weights in [0, 1], then the probability that the Weber solution with Euclidean distances is on a demand point is approximately $\frac{1}{n}$. This counter-intuitive result was verified by simulation. Later on I realized that a similar result holds for Manhattan (ℓ_1) distances (not published). The Weber solution is found by drawing a grid of horizontal and vertical lines through the demand points and the optimal solution is on one of the grid intersection points (Francis et al. 1992; Love et al. 1988). There are n^2 intersection points and n of them are demand points. The probability that the optimal solution is on a demand point is approximately $\frac{1}{n}$.

Drezner (1989b) investigated the Weber objective on a surface of a sphere when the demand points and weights are randomly generated. It is proven that when the number of demand points increases to infinity: (1) the ratio between the maximum value of the objective function and the minimum value converges to one, and (2) the expected number of points that are a local minimum is equal to one.

Drezner and Wesolowsky (1991) considered the possibility that some of the weights in the Weber problem can be negative. This problem is also termed “the Weber problem with attraction and repulsion” (Maranas and Floudas 1993; Chen et al. 1992; Plastria 1991). Drezner and Suzuki (2004) proposed the BTST global optimization algorithm and used this problem to illustrate the procedure.

Drezner et al. (2018c) proposed the Weber obnoxious facility location problem. The facility location is required to be at least a given distance from demand points because it is “obnoxious” to them. A practical example is locating an airport that generates noise and pollution but serves travelers and thus their total travel distance to the airport should be minimized. Since in most applications the nuisance generated by the facility “travels by air,” the analysis deals mainly with the case where the required minimum distance between the facility and demand points is Euclidean. The Weber objective distance can be measured by a different norm. Very efficient branch-and-bound algorithms to optimally solve the single facility problem and an algorithm based on creating a finite candidate set are developed. The algorithms were tested on problems with up to 10,000 demand points using Euclidean, Manhattan, and ℓ_p for $p = 1.78$ norms for the Weber objective. The largest problems were optimally solved in a few seconds of computer time. Many extensions to the basic Weber obnoxious facility location problem are proposed for future research.

Drezner et al. (2003) analyzed the optimal location of a central warehouse, given a fixed number and locations of local warehouses. They investigated whether the solution determined by the traditional model that minimizes total transportation cost differs from the one determined by a model that also takes into account the inventory and service costs. Numerical results show that ignoring inventory costs may lead to inferior location solutions. In a follow-up paper Drezner and Scott (2013) added to the model the consideration of perishable products. The location of the distribution center affects the inventory policy. Brimberg and Drezner (2019) considered the case that the demand points can be partitioned into clusters and the number of facilities allocated to each cluster is found by dynamic programming. They solved both the p-median and p-center problems. The model was inspired by the idea presented in Drezner et al. (2016a).

1.4.1.2 Minimax and Center Objectives

Drezner (2011) reviewed planar center problems and their history. The unweighted one-center problem was proposed and solved by Sylvester (1857, 1860) more than 150 years ago.

Drezner (1981a) considered a variation on the weighted one-center problem (minimizing the maximum weighted distance from the facility). The objective is to find the maximum total weight of demand points within a given distance r from the facility. An optimal procedure of complexity $O(n^2 \log n)$ is proposed.

Drezner (1984c) suggested two heuristics and one optimal algorithm for the solution of the weighted p-center problem. The two heuristic procedures are similar to the location-allocation algorithm (Cooper 1963, 1964). The optimal algorithm is based on the property that the solution to the one-center problem is based on two or three demand points (Elzinga and Hearn 1972). There are $\frac{1}{6}n(n^2 + 5)$ possible maximal sets. Based on a proof in Drezner (1981a), the number of relevant maximal sets is bounded by $n(n-1)$. The proposed optimal algorithm is polynomial in n for a fixed p. The extension of the modified center problem (Drezner 1981a) is also solved in polynomial time.

Drezner (1982b) proved that in a convex minimax optimization problem in k dimensions there exist a subset of $k + 1$ functions such that a solution to the minimax problem with these $k + 1$ functions is a solution to the minimax problem with all functions. Drezner (1987c) developed a procedure which is an extension of the Elzinga and Hearn (1972) algorithm for k-dimensional problems. Drezner and Shelah (1987) constructed a special configuration that shows that the Elzinga and Hearn (1972) algorithm can have a complexity of at least $o(n^2)$ even though in practice the complexity is linear for most tested problems.

Drezner (1984d) optimally solved the two-center and the two-median problems on the plane. It is proved that in the optimal solution there is a line separating the set of demand points into two subsets such that the single facility solution point for each subset yields the optimal solution to the two facilities problem. There are at most $\frac{1}{2}n(n-1)$ such lines. Once all separating lines are constructed, the optimal solution is identified. A lower bound for each separating line reduces the effort required by the algorithm.

Drezner (1987d) considered the (unweighted) p-center problem with Manhattan (ℓ_1) distances. Algorithms that find the optimal solution of the 1-center and 2-center problems in $o(n)$ time and the 3-center problem in $o(n \log n)$ time are proposed. For $p \geq 4$, algorithms polynomial in n for a given p are proposed.

Drezner (1991) considered the problem of locating p facilities among n demand points in an m-dimensional space. For each demand point i a value f_i which is the weighted sum of the distances to all facilities plus a setup cost is calculated. The objective is to minimize the maximum value of f_i among all demand points by the best location of the facilities. For locating a single facility f_i reduces to the weighted distance plus a setup cost.

Drezner (1995e) suggested an iterative procedure to solve convex minimax optimization problems. Each iteration, the functions are approximated by spheres and the spherical minimax problem solved.

Rodríguez-Chía et al. (2010) presented a solution procedure based on a gradient descent method for the p-center problem in the plane using Euclidean distances. The solution approach is based on analytical expressions for the bisector lines separating facilities' locations.

Irawan et al. (2016) solved large-scale unconditional and conditional p-center problems both heuristically and optimally.

1.4.1.3 Obnoxious Facilities

Obnoxious facilities are facilities generating nuisance and the objective is to locate such facilities as far as possible from demand points. Without constraints the solution is at infinity. Therefore, such models require a location of the facility in a finite area such as the convex hull of the demand points. The problem on a network was proposed by Church and Garfinkel (1978).

Berman and Drezner (2000) found the best location for an obnoxious facility on a network such that the least weighted distance to all nodes of the network is maximized.

Drezner and Wesolowsky (1996) found the best location in the interior of a network that is as far as possible from nodes and arcs of the network. Drezner et al. (2009b) changed the objective to the sum of distances from nodes and arcs.

Welch et al. (2006) solved the planar p obnoxious facilities location problem maximizing the minimum distance between facilities and demand points and between facilities.

Drezner et al. (2018g) considered two multiple obnoxious facilities problems. A given number of facilities are to be located in a convex polygon with the objective of maximizing the minimum distance between facilities and a given set of demand points. In one model the facilities have to be farther than a certain distance from one another. In the second model the facilities have to be farther from one another than the minimum distance to demand points multiplied by a given factor. The proposed heuristic solution approach is based on Voronoi diagrams (Suzuki and Okabe 1995; Okabe et al. 2000; Voronoï 1908). It has no random component and needs to be applied only once. A binary linear program, which is applied iteratively, was constructed and solved optimally. The results were compared with a multi-start approach using interior point, genetic algorithm (GA), and sparse non-linear optimizer (SNOPT) solvers in Matlab where each instance is solved using 100 randomly generated starting solutions and selecting the best one. It was found that the heuristic results are much better and were found in a fraction of the computer time. For example, one instance of locating 20 facilities among 1000 demand points in a unit square was heuristically solved in 24 s compared with about 5 h of computer time by Matlab using interior point or SNOPT solvers.

1.4.1.4 Equity Objectives

The minimax objective discussed in Sect. 1.4.1.2 can be viewed as an equity objective. We wish to provide the best possible service to the demand that gets the worst service. In such models we are not concerned with demand that gets excellent service. In equity objectives we aim to minimize the difference between the worst service and the best one. Eiselt and Laporte (1995) list 19 different equity objectives.

Drezner et al. (1986) investigated the location that minimizes the range of the distances to all demand points. Drezner and Drezner (2007) investigated two equity objectives: (1) minimizing the variance of the distances to the facility, and (2) minimizing the range of the distances. These problems were solved optimally using the BTST global optimization algorithm (Drezner and Suzuki 2004).

Drezner et al. (2009a) investigated the location of facilities minimizing the Gini coefficient of the Lorenz curve (Lorenz 1905; Gini 1921) calculated by service distances to the closest facility. An algorithm that finds the optimal location of one facility in a bounded area in the plane is constructed and optimally solved using the BTST procedure (Drezner and Suzuki 2004).

Drezner et al. (2014) analyzed the single facility location problem minimizing the quintile share ratio (Eurostat 2012) for continuous uniform demand in an area such as a disk, a rectangle, and a line; when demand is generated at a finite set of demand points; and when the facility can be located anywhere on a network. The quintile share ratio is a measure of inequity. For income levels in a population it is the ratio of total income received by the 20% of the population with the highest income (top quintile) to that received by the 20% of the population with the lowest income (lowest quintile). The concept is applied to distances rather than income.

Suzuki and Drezner (2009) analyzed the location of p facilities serving continuous area demand. They solved three objectives: (1) the p-center objective, (2) equalizing the load service by the facilities, and (3) minimizing the maximum radius from each point to its closest facility when each facility services the same load.

Berman et al. (2009d) solved the problem of minimizing the maximum load (defined as the sum of weights of demand points closest to the facility) among p facilities on a network.

Drezner and Drezner (2011b) assumed that the set of demand points is partitioned into groups. These groups are not necessarily divided by their locations but by characteristics such as poor neighborhoods and rich neighborhoods. The objective is to provide equitable service to the groups by locating one or more facilities. For example, poor neighborhoods should get comparable service level to rich neighborhoods.

1.4.1.5 Cover Objectives

Cover location models are one of the main branches of location analysis. A demand point is covered by a facility within a certain distance (Church and ReVelle 1974; ReVelle et al. 1976). Facilities need to be located in an area to provide as much cover as possible. Such models are used for cover provided by emergency facilities such as ambulances, police cars, or fire trucks. They are also used to model cover by transmission towers such as cell-phone towers, TV or radio transmission towers, and others. For a review see Plastria (2002), García and Marín (2015), and Snyder (2011).

Drezner (1986b) proposed efficient algorithms to cover as much demand as possible by p facilities The distance norm is Manhattan (ℓ_1).

Drezner and Wesolowsky (1997a) considered the situation where an event may occur anywhere in a planar area or on a linear region such as a route. One or more detectors are to be located within this region with the objective of maximizing the smallest probability of the detection of an event anywhere in the region. In other words, the minimum protection in the region is to be maximized. The probability that an event is detected is a decreasing function of the distance.

Drezner et al. (2004) proposed the gradual cover model with linear decay. Rather than abrupt switch from full cover to no coverage (for example, at distance 2.99 miles there is full cover, while at distance 3.01 miles there is zero cover) it is proposed that cover declines gradually. For example, a demand point is fully covered within a distance r and not covered at all for distance exceeding $R > r$. For distances between r and R the proportion of cover declines linearly. In "standard" cover models $r = R$. The single facility location problem maximizing total cover is optimally solved by BTST (Drezner and Suzuki 2004). Drezner et al. (2010a) modified the linear gradual cover by assuming that r and R are random variables.

Berman et al. (2009c) proposed a covering problem where the covering radius of a facility is controlled by the decision-maker. The cost of achieving a certain covering distance is assumed to be a monotonically increasing function of the covering distance. The problem is to cover all demand points at a minimum cost by finding optimal number, locations, and coverage radii for the facilities.

Berman et al. (2010a) assumed that facilities "cooperate" in providing cover. Each facility emits a signal that declines by distance and a demand point is covered if the total signal exceeds a given threshold. For example, locating light posts in a parking lot so that all points in the lot get a minimum strength of light. The discrete version of the cooperative cover is analyzed in Berman et al. (2011b) and its network variant is analyzed in Berman et al. (2013). Berman et al. (2010b) reviewed the three models mentioned above: gradual cover, variable radius, and cooperative cover.

Berman et al. (2019) analyzed multiple facilities gradual cover models which consider the possibility of partial cover. The issue of joint partial coverage by several facilities in a multiple facilities location model is investigated. Theoretical foundations for the properties of the joint coverage relationship to individual partial covers are established. Models based on these theoretical foundations are developed. The location problems are solved both heuristically and optimally within a pre-specified percentage from the optimal solution. A follow-up paper is Drezner and Drezner (2014). Berman et al. (2019)'s objective is to maximize the total cover, while Drezner and Drezner (2014)'s objective is to maximize the minimum cover.

Drezner et al. (2019) propose, analyze, and test a new rule for calculating the joint cover of a demand point which is partially covered by several facilities. It is reasonable to assume that facilities are "points" because in most applications the facilities occupy a small area compared to the demand area. However, in most applications demand "points" generate demand from an area and each point in the area may have a different distance to the facility. The total number of customers covered by several facilities depends both on the distances of the facilities from the demand area and their directions. For example, if two facilities are located one to the north of the neighborhood and one to the south, one facility may cover customers located at the northern part of the neighborhood, while the other one covers customers in the south and the total number of customers is usually the sum of the two. If the two facilities are located in the same direction, there will be a significant overlap, and the total number of customers covered is less than the sum of the two. The algorithm is tested on a case study of locating cell-phone towers in Orange County, California. The new approach provided better total cover than the cover obtained by existing procedures. Follow-up papers propose and investigate: the maximin objective (Drezner et al. 2019a); the continuous demand (Drezner and Drezner 2019a); solving the models using genetic algorithms (Drezner et al. 2019b).

Kalczynski and Drezner (2019b) solved the problem of packing a given number of ellipses with known shapes in the rectangle with the smallest area. The ellipses can be rotated and do not have to be parallel to the sides of the rectangle.

Drezner et al. (2011, 2012a, 2015a, 2016a) proposed a competitive facilities location models based on cover objectives. A follow-up paper we are working on is Drezner et al. (2018). Rather than using the standard cover objective, the gradual cover objective is applied. For recent reviews of competitive facilities' location problems see Berman et al. (2009a), Drezner (2014), Eiselt et al. (2015), Drezner and Eiselt (2002), and Eiselt (2011).

1.4.1.6 Hub Location and Related Models

The round trip location problem (Drezner 1982a, 1985a; Drezner and Wesolowsky 1982) preceded the hub location problem (Skorin-Kapov et al. 1996; Sasaki et al. 1997; O'kelly 1987; Campbell 1994; Contreras 2015; Drezner and Drezner 2001) and has similar features. n pairs of points are located in an area. The objective function to be minimized is the maximum round trip between the facility to one demand point, its paired demand point, and back.

Another related problem is the transfer point location problem (Berman et al. 2007, 2008, 2005). Demand is generated at a set of demand points who need the services of a central facility. The service provided by the facility is provided through transfer points, i.e., the total distance for a customer is the distance to the transfer point plus the distance to the facility. Each distance may be multiplied by a different weight to convert it to time. Locations for the central facility and the transfer points are sought. Both minisum and minimax objectives can be applied. A similar setting is used in Sasaki et al. (1997) where the transfer points are termed relay points.

Drezner and Wesolowsky (2001) proposed the collection depots location problem. A set of demand points and a set of collection depots are given. A facility that minimizes total travel distance need to be located. Each service consists of a trip to the customer, collecting materials, dropping the materials at one of the available collection depots, and returning to the facility to wait for the next call. Berman et al. (2002) solved the network version of the collection depots problem. Drezner et al. (2019c) investigated the multiple collection depots problem in the plane.

Sasaki et al. (1999) considered the 1-stop multiple allocation p-hub median problem. The problem is formulated as a p-median problem. A branch-and-bound algorithm and a greedy-type heuristic are proposed. Suzuki and Drezner (1997) considered the continuous version of this problem. Drezner and Drezner (2001) considered the hub location problem where customers do not necessarily select the shortest route through a hub but split their selection according to the gravity rule (Huff 1964, 1966; Reilly 1931).

1.4.2 The Environment

1.4.2.1 Location on a Sphere

Companies become increasingly global and demand points may be distributed all over the world. It cannot be assumed that the area of interest is planar. Thus, the importance of spherical models increases over the years.

Drezner and Wesolowsky (1978a) showed that spherical distances are convex up to a distance of $\frac{\pi}{2}$ and if all demand points are in a spherical disk of a radius not exceeding $\frac{\pi}{4}$, then there is a unique local optimum which is the global one.

Drezner (1981b) showed that if all demand points on the sphere lie on a great circle, such as the equator, then the optimal solution to the Weber problem is also on that great circle and is located at one of the demand points.

Drezner (1983a) proposed (in addition to planar problems) a unified approach for solving unweighted maximin and minimax problems on a sphere, subject to constraints which are inside or outside a set of circles. The property that the sum of distances to a point and its antipode on a unit sphere is equal to π is utilized. A constraint outside a circle centered at a point can be converted to a constraint inside the circle centered at its antipode. Maximin objective can be converted to minimax objective by using the antipodes of demand points.

Drezner and Wesolowsky (1983) relied on the property that the sum of distances to a point and its antipode is equal to π to solve the weighted minimax and maximin problems. They developed an algorithm that finds the global optimum of the weighted minimax problem. Since the maximin problem can be converted to an equivalent minimax problem, this algorithm can be used to optimally solve the maximin problem as well. Drezner (1985c) suggested to use the algorithm in Drezner and Wesolowsky (1983) to optimally solve the Weber location problem on a sphere.

A related paper is Drezner et al. (2016a) which deals with a competitive model but can be applied to other models as well. Suppose that the set of demand points can be divided into mutually exclusive clusters. For example, a cluster in New York, a cluster in Tokyo, a cluster in London, and so on. A budget is available for investment in these clusters to improve a well-defined objective. Suppose that, for a particular cluster, the best solution and its objective function value for a given budget allocated to this cluster can be found. The total budget can be divided, for example, to 1000 units, i.e., each unit has 0.1% of the available budget. The allocation in whole units to each cluster to optimize the total value of the objective function can be found by dynamic programming. The same idea can be implemented for other facilities' location objectives. For example, suppose that in a p-median or a p-center model the demand area can be partitioned into mutually exclusive subsets far enough from one another. The total number of facilities is given. The number of facilities in each subset needs to be determined so that the objective function is minimized.

1.4.2.2 Continuous Demand

Drezner and Wesolowsky (1980a) solved the problem of locating a facility among area demands rather than points. They solved the problem with general ℓ_p distances. Drezner and Wesolowsky (1978b) pointed to a mistake in another paper (Bennett and Mirakhor 1974) that found the best location of a facility when demand is generated in areas rather than demand points. Drezner (1997) considered the reverse problem of converting an area to a set of demand points.

Drezner (1986a) considered the Weber and p-median problems with Euclidean and square-Euclidean distances. Both demand points and facilities are assumed to have circular shapes with uniform demand and service origin.

Drezner and Erkut (1995) solved the p-dispersion problem in a square which is equivalent to circle packing in a square (Szabo et al. 2007). p points need to be located in a square so that the minimum distance between the points is maximized. This problem can be generalized to any shape.

Suzuki and Drezner (1996) solved the p-center problem when demand is generated in an area rather than demand points applying Voronoi diagrams (Suzuki and Okabe 1995; Okabe et al. 2000; Voronoï 1908). Drezner and Suzuki (2010) used a similar procedure to find the set of circles of a given radius that cover the most area of a given shape. Suzuki and Drezner (2009) consider the problem of equalizing the loads serviced by each facility which is formulated as minimizing the maximum load serviced by the facilities.

Drezner et al. (2018d) propose and solve a competitive facility location model when demand is continuously distributed in an area and each facility attracts customers within a given distance. This distance is a measure of the facility's attractiveness level which may be different for different facilities. The market share captured by each facility is calculated by two numerical integration methods. These approaches can be used for evaluating functional values in an area for other Operations Research models as well. The single facility location problem is optimally solved by the BTST (Drezner and Suzuki 2004) global optimization algorithm and the multiple facility problem is heuristically solved by the Nelder–Mead algorithm (Nelder and Mead 1965).

1.4.2.3 Location in a Large Area

For covering a large demand area, there are three symmetric grids in the plane: triangular grid, square grid, and hexagonal grid (see Fig. 1.9).

Drezner and Zemel (1992) considered the following question. There is an area with uniform demand and we wish to build facilities in a pattern. There are three symmetric grids in the plane: triangular grid, square grid, and hexagonal grid. Which of these grids protects best against a future competitor? (the leader–follower or Stackelberg equilibrium problem, Stackelberg 1934; Drezner and Drezner 2017). It turns out that a hexagonal grid is best. A future competitor can capture up to 51.27% of the market share of one existing facility. In a square grid a competitor can capture up to 56.25% of the market share and in a triangular grid it can capture $\frac{2}{3}$ of it. This is consistent with many other objectives (Suzuki and Drezner 1996; Drezner and Suzuki 2010; Hilbert and Cohn-Vossen 1932; Szabo et al. 2007). We thought at the time, but never pursued it, that a hexagonal road grid may be better than the commonly used Manhattan (ℓ_1) grid.

Drezner and Drezner (2018b) investigated the total cover area of two, three, four, and many facilities applying the cooperative cover model (Berman et al. 2010a). It is shown that for many facilities located in a symmetric grid, the hexagonal grid is best.

Taillard (1995) proposed the gray pattern problem which is a special case of the quadratic assignment problem (QAP, Drezner 2015b). A square (it can be a rectangle) of m by m points is given. k points out of the m^2 points need to be selected as "black" points (they play the role of facilities). A large area is covered by a square grid with each square having an identical black points distribution. The objective is to have the black points distributed as smoothly as possible so that the area will look as uniform gray. To find the locations for the black points, distances to the eight adjacent squares are considered in the objective function.

Drezner (2006a) proposed an efficient approach to solve the gray pattern problem. Drezner et al. (2015d) improved the algorithm and optimally solved much bigger problems. For example, the instance Tai64c, which is a QAP with $n = 64$ facilities (see the QAPLIB: http://anjos.mgi.polymtl.ca/qaplib/) was optimally solved in 15 s.

Possible follow-up papers may include testing whether it is better to select black points in a hexagon or a triangle forming a hexagonal or triangular grids, see Fig. 1.9. Also, a color pattern structure may be of interest. Rather than black points, several colors (for example, red, blue, and yellow) points at certain proportions need to be selected with the objective of forming uniform patterns for the points of all colors combined in addition to the points of each color separately.

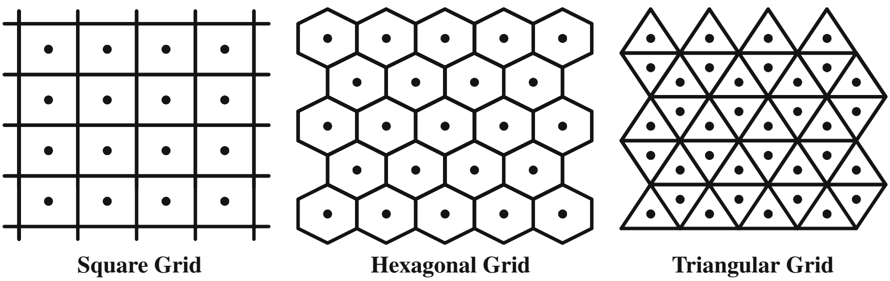


Fig. 1.9 The three symmetric grids

Two models incorporating the concept of creating a structure in one square considering the structure in the adjacent squares were proposed. Drezner and Kalczynski (2018) proposed the p-dispersion covering problem (p-DCP) which is a generalization of the p-dispersion problem in a square (Drezner and Erkut 1995; Kuby 1987). In the p-dispersion problem the objective is to maximize the minimum distance between any two facilities located in a square. In the p-DCP, distances to adjacent squares are included in the objective function so that if an area is covered by many squares, the solution in each square is replicated and the minimum distance between any two facilities is maximized. Drezner and Kalczynski (2017) considered the gray pattern problem in a square when the points can be located anywhere in the square and are not restricted to a pre-specified set of potential locations.

1.4.3 Conditional Location Problems

Conditional location problems are an extension of "standard" location problems (Ogryczak and Zawadzki 2002; Minieka 1980; Chen and Handler 1993; Chen 1988; Berman and Simchi-Levi 1990). Several facilities exist in the area and additional facilities need to be located. Customers get the service from the closest facility, regardless of whether it is existing or new.

Drezner (1989a) showed that conditional p-center problems can be solved by solving $o(\log n)$ p-center problems, where n is the number of demand points. Therefore, once an efficient algorithm exists for the p-center problem (by any metric or on a network), then an efficient one can be constructed for the conditional version of the problem.

Drezner (1995d) analyzed the conditional planar p-median problem when q facilities exist in the area. The $p = q = 1$ case is optimally solved and heuristic algorithms are proposed for the general case.

Berman and Drezner (2008a) proposed a new formulation for the conditional p-median and p-center problems on a network. The new formulation can be used to construct more efficient algorithms than those constructed by the best known formulation.

1.4.4 Unreliable or Uncertain Data

Drezner (1987b) investigated the p-median and p-center problems when facilities are unreliable and may become inactive. Demand points get their services from the closest active facility.

Berman et al. (2003) assumed that the probability that a facility cannot provide satisfactory service increases with the distance from the facility. They solved the p-median problem on a network when failure to provide satisfactory service is considered.

Drezner and Drezner (2017) reviewed various objectives in location problems that are modeled as leader–follower also termed the Stackelberg equilibrium problem (Stackelberg 1934).

Berman and Drezner (2003) considered the 1-center problem on a network when the speeds on links are stochastic rather than deterministic. The objective is to find the location for a facility such that the probability that all nodes are reached within a given time threshold is maximized.

Berman et al. (2009b) considered the p-median problem on a network when one of the links may become unusable either due to a terrorist attack or a natural disaster. They formulated it as a leader–follower problem. The follower's problem is to select the most damaging link and the leader's problem is to minimize total cost following a link becoming unusable.

Berman and Drezner (2008b) proposed the p-median problem on a network under uncertainty. p facilities are to be located but it is possible that up to q additional facilities will have to be located in the future. The objective is to find locations of p facilities such that the expected value of the objective function in the future is minimized. A related problem is investigated in Drezner (1995a). Demand is changing over time and additional facilities are built in the future to serve an increasing demand. The objective is to determine the establishment times and locations for future facilities.

Drezner and Guyse (1999) considered a location problem with future uncertainties about the data. Several possible scenarios about the future values of the parameters are postulated. Four rules utilized in decision theory are examined: the expected value rule, the optimistic rule, the pessimistic rule, and the minimax regret rule.

Demand at various demand points follows a multivariate distribution. There is a desired threshold to be met. Drezner et al. (2002a) considered the objective of minimizing the probability that the market share captured by a new facility does not meet a given threshold. Drezner and Drezner (2011c) considered the same model for the Weber objective not being below a certain threshold.

Many location problems can be expressed as ordered median objective (Nickel and Puerto 2005). Drezner et al. (2012b) investigated the ordered median objective when the demand points are generated in a circle. The mean and variance of the kth distance from the center of the circle and the correlation matrix between all pairs of ordered distances are found. By applying these values, the mean and variance of any ordered median objective and the correlation coefficient between two ordered median objectives can be calculated.

Drezner et al. (2018e) extend the gravity model by allowing attractiveness of facilities to be randomly distributed. It was found that when facilities' attractiveness are random, facilities tend to lose some competitive advantage. The decrease in attractiveness for a given mean is approximately proportional to the variance of the attractiveness distribution. Drezner et al. (2018f) observed that more attractive facilities attract shoppers from larger distances. They propose to estimate attractiveness by the distances shoppers travel rather than by opinion surveys.

1.4.5 Other Models

Drezner (1988a) considered the problem of locating satellites in orbits to maximize their coverage of the sphere (earth). Each satellite covers a spherical disk at any point in time but constantly moves in orbit.

The problem of maximizing the sight angles of shapes from a point located anywhere in the plane is analyzed in Drezner (1988b). There are n points on the plane that are to be observed from a location to be found. We wish to find the observation point that has the best possible view of the points in the sense that if we draw lines of sight from the observation point to the given points, the smallest angle between the lines is maximized. Applications include an art display or surveillance point when an object in the line of sight to a farther object may block its view. It is a follow-up to Drezner and Wesolowsky (1980b) where the location of the facility is restricted to arcs (or the whole circumference) on a circle surrounding all points.

Drezner and Wesolowsky (1998) considered the minisum (Weber) and minimax location problem with rectilinear distances. However, the axes can be rotated to provide the minimum value of the objective function. This problem can be used in the planning phase of constructing roads or aisles with rectilinear distance but not necessarily in north–south and east–west directions.

Drezner and Drezner (2013) suggested overlapping Voronoi diagrams. The Voronoi regions do not partition the plane into disjoint regions. Rather, points may belong to several regions as long as the distance to a Voronoi point does not exceed a certain percentage over the shortest distance to all Voronoi points. The concept is illustrated on a case study of delineating overlapping service areas for public universities. If two campuses are at about the same distance from a potential student, students can register to either of them.

Berman and Drezner (2007) introduced the multiple server location problem. A given number of servers are to be assigned to nodes of a network. A subset of nodes need to be selected for locating one or more servers in each. Each customer at a node selects the closest server. The objective is to minimize the sum of the travel time and the average time spent at the server calculated by the M/M/k queueing formulas. An efficient method for calculating total time in M/M/k queues is presented in Pasternack and Drezner (1998). Follow-up papers are (Aboolian et al. 2009; Drezner and Drezner 2011a).

Drezner (2004, 2007a) introduced the casualty collection points (CCPs) location problem. The CCPs are employed in cases of mass casualty incidents that require delivering emergency medical care to a large number of victims. The underlying assumption is that in a catastrophic event, such as a major earthquake, the area's civil infrastructure (freeways, roads, communications, emergency medical services, etc.) will not be operational. Hospitals may themselves become victims or otherwise inundated or inaccessible. Drezner et al. (2006) proposed a minimax regret multi-objective formulation for the CCP location problem. The objective is to minimize the maximum percent deviation of individual objectives from their best possible objective function value. Five objectives are included in the model: (1) p-median,

(2) p-center, (3) the p-Max Cover objective within two miles (walking distance), (4) the p-Max Cover objective within three miles, and (5) the minimum variance as a measure of equity.

Kalsch and Drezner (2010) considered the simultaneous scheduling-location single machine model in the plane. The model combines both the location of the machine and the scheduling of the jobs executed by the machine. Two objectives are analyzed: the make span and the total completion time.

Drezner and Menezes (2016) compared the Condorcet voting procedure to the Weber location problem. They analyzed the difference in location solutions as well as the value of one objective at the solution point of the other. Menezes et al. (2016) analyzed also approval voting in comparison with the cover objective.

Drezner and Brimberg (2014) considered the problem of fitting a given number of concentric circles to a given set of points. This is a generalization of the problem of fitting a set of points to one circle (Drezner et al. 2002c). Three objectives, to be minimized, are considered: the least squares of distances from the circles, the maximum distance from the circles, and the sum of the distances from the circles. In a follow-up paper, Brimberg and Drezner (2015) considered a continuous location problem for a given number of concentric circles serving a given set of demand points. Each demand point is serviced by the closest circle. The objective is to minimize the sum of weighted distances between demand points and their closest circle.

Berman and Drezner (2006) investigated the problem of locating a given number of facilities on a network. Demand generated at a node is distance dependent, i.e., it decreases when the distance increases. The facilities can serve no more than a given number of customers; thus, they are capacitated and congested when they reach that limit. The objective function is to maximize the demand satisfied by the system given these constraints.

Drezner and Drezner (2019b) assume that a budget is available for expansion of chain facilities. The part of the budget invested in improving an existing facility or constructing a new one is an integer multiple of a basic value such as 0.1% of the available budget. The gravity model (Huff 1964, 1966) is used to estimate market share captured by chain facilities. Both improvement of existing facilities and construction of new facilities are considered in the model. The problem is optimally solved by a branch-and-bound algorithm when the set of possible locations for the new facilities is finite.

1.5 Solution Methods

1.5.1 Global Optimization Algorithms

Many non-convex planar single facility problems can be optimally solved by the global optimization algorithms "Big Square Small Square" (BSSS, Hansen et al.

1981) and “Big Triangle Small Triangle (BTST, Drezner and Suzuki 2004). Methods that used this idea for other environments are Big Region Small Region (BRSR, Hansen et al. 1995) for location on a sphere, Big Cube Small Cube (BCSC, Schöbel and Scholz 2010) for higher dimensional space, Big Segment Small Segment (BSSS, Berman et al. 2011a) for optimization on a network, Big Arc Small Arc (BASA, Drezner et al. 2018c) for optimization on circumferences of disks.

The BTST algorithm starts with a Delaunay triangulation (Lee and Schachter 1980) creating a list of triangles. The vertices are the feasible demand points and the vertices of the feasible region if there is one. Upper bound and lower bounds for each triangle in the list are calculated and $\overline{UB}$ is the best upper bound. Many of the triangles in the list for which $LB \geq \overline{UB}(1-\epsilon)$ for a given relative accuracy ϵ are removed from the list. The process continues by selecting a triangle in the list with the smallest LB as a “big triangle” and splitting it into four “small triangles.” The best upper bound $\overline{UB}$ may be updated. The big triangle is removed from the list and small triangles for which $LB \geq \overline{UB}(1-\epsilon)$ are ignored. Small triangles which do not satisfy this condition are added to the list. The process continues until the list of triangles is empty. For complete details see Drezner and Suzuki (2004).

The BSSS algorithm starts with a list consisting of a “big square” enclosing the feasible region. The upper bound in the big square, $\overline{UB}$, is the best upper bound found so far. The remainder of the process is very similar to the BTST algorithm. A selected square in the list is divided into four small squares by connecting the centers of the sides of the big square.

If the solution is restricted to the convex hull of the demand points, or any convex polygon, the triangulation by BTST takes care of it automatically, while BSSS requires an extra check whether the solution point is feasible or not. The lower and upper bounds required for applying the BSSS algorithm may also be affected by the feasibility issue.

Suzuki and Drezner (2013) extended the BTST (Drezner and Suzuki 2004) algorithm to multiple facility location (in particular the location of two facilities). Drezner (2007b) proposed a general approach for constructing bounds required for the BTST algorithm for the solution of planar location problems. Optimization problems, which constitute a sum of individual functions, each a function of the Euclidean distance to a demand point, are analyzed and solved. The bounds are based on expressing each of the individual functions in the sum as a difference between two convex functions of the *distance*, which is not the same as convex functions of the location. Drezner and Nickel (2009b,a) applied the method proposed in Drezner (2007b) to a general ordered median formulation (Nickel and Puerto 2005). For a review of these approaches see Drezner (2013).

1.5.2 Improvements to the Weiszfeld Algorithm

Several improvements on the Weiszfeld algorithm (Weiszfeld 1936; Weiszfeld and Plastria 2009) were proposed. Drezner (1992b) proposed to multiply the change between two consecutive iterations by a factor of 1.8. An "ideal" multiplier that needs to be computed every iteration was also proposed. Drezner (1996) proposed to accelerate the procedure by estimating the limit of a geometric series by the ratio of the changes of two consecutive iterations. Drezner (2015a) proposed an algorithm based on the values of the objective function at 9 points (the present iteration and eight points around it on the vertices and sides' centers of a square) and fitting a paraboloid to these nine values by a least squares quadratic regression. The next iterate is the minimum point of the paraboloid. We give here a convenient way to calculate the next iterate in the fortified method (Drezner 2015a).

The present iterate is (x_0, y_0). For a given Δ, 8 values $f = \{f_1, f_2, \ldots, f_8\}$ are defined. These are the differences in the value of the objective function between the eight points on the periphery of the square and the squares' center (x_0, y_0). The points are depicted in Fig. 1.10:

Two vectors are defined: $\alpha = \{1, 1, -1, -1, 0, 0, 1, -1\}$ and $\beta = \{1, -1, 1, -1, 1, -1, 0, 0\}$. Then, $f_i = f(x_0 + \alpha_i\Delta, y_0 + \beta_i\Delta) - f(x_0, y_0)$. Calculate $\gamma_i = f_{2i-1} + f_{2i}$; $\delta_i = f_{2i-1} - f_{2i}$ for $i = 1, \ldots, 4$. Then: $v_1 = 4(\gamma_1+\gamma_2-2\gamma_3+3\gamma_4)$; $v_2 = 5(\delta_1-\delta_2)$; $v_3 = v_1+20(\gamma_3-\gamma_4)$; $v_4 = \gamma_1-\gamma_2+\delta_4$; $v_5 = \delta_1 + \delta_2 + \delta_3$.

The next iterate $(\overline{x}, \overline{y})$ is

$$\overline{x} = x_0 + \Delta\frac{v_2v_5 - v_3v_4}{0.3(v_1v_3 - v_2^2)}; \qquad \overline{y} = y_0 + \Delta\frac{v_2v_4 - v_1v_5}{0.3(v_1v_3 - v_2^2)}. \tag{1.25}$$

Drezner (2009) considered Weber-like location problems where the objective function is a sum of terms, each a function of the Euclidean distance from a demand point. It is proved that a Weiszfeld-like iterative procedure, termed generalized Weiszfeld, for the solution of such problems converges to a local minimum (or a saddle point) when three sufficient conditions are met: the functions are twice differentiable (except, possibly at the demand points), concave, and monotonically non-decreasing with the distance.

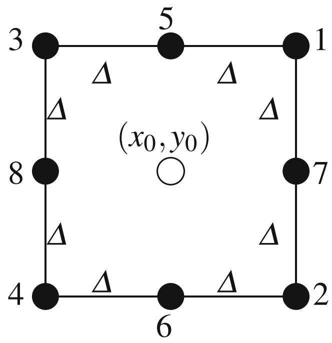


Fig. 1.10 The square configuration for the fortified algorithm

1.5.3 Heuristic Methods for Solving the *p*-Median Problem

Several papers suggested heuristic algorithms for the solution of the planar p-median problem. Brimberg and Drezner (2013) proposed to find a good starting solution by applying metaheuristics such as descent, tabu search, simulated annealing on a grid. The location-allocation algorithm (Cooper 1963, 1964) is then applied for finding the final solution. Brimberg et al. (2014) applied variable neighborhood search on the grid starting solution. Drezner et al. (2015c) applied variable neighborhood search and concentric tabu search (Drezner 2002) on the starting solution and Drezner et al. (2015b) applied a specially designed genetic algorithm. These methods were refined by Drezner and Salhi (2017). Brimberg et al. (2017) and Drezner et al. (2016b) applied the reformulation local search which switches between the continuous model and a discrete relaxation in order to expand the search. In each iteration new points obtained in the continuous phase are added to the discrete formulation. The best results to date are reported in Drezner and Drezner (2018c). A genetic algorithm for solving the network p-median is proposed in (Alp et al. 2003).

1.5.4 The Demjanov Algorithm

The Demjanov algorithm was proposed by Demjanov (1968) and applied in Drezner and Wesolowsky (1997a, 1985). It is designed to solve the minimization of $f(X) = \max\limits_{1\le i\le n} \{f_i(X)\}$, where X is a vector of p variables. It can be used to solve heuristically, for example, the location of p facilities among n demand points such as the p-center problem. For such an application the vector $X = (x_j, y_j)$ for $j = 1, \ldots, p$ is a vector of locations which is a vector of $2p$ variables.

Random locations for the p facilities are selected and the solution is improved by moving in the direction of steepest descent. The objective function is optimized on the ray of steepest descent using a one dimensional optimization procedure such as the golden section search (Zangwill 1969).

Calculation of the gradient is done as follows. A tolerance δ is selected defining the set $I(X)$ of demand points:

$$I(X) = \{\, i \mid f_i(X) \ge f(X) - \delta \,\}.$$

The set $I(X)$ is the set of "binding" demand points. If the function $f_i(X)$ is reduced for all binding demand points when X is changed infinitesimally, then the objective function is reduced. Consider changing location (x_j, y_j) by $(\Delta x_j, \Delta y_j)$ for $j = 1, \ldots, p$. The steepest descent direction of $f(X)$ is obtained by solving the following positively definite quadratic programming problem:

$$\min \sum_{j=1}^{p} \left\{ [\Delta x_j]^2 + [\Delta y_j]^2 \right\}$$

$$\text{subject to: } \sum_{j=1}^{p} \left\{ \frac{\partial f_i(X)}{\partial x_j} \Delta x_j + \frac{\partial f_i(X)}{\partial y_j} \Delta y_j \right\} \leq -1 \quad \text{for } i \in I(X). \quad (1.26)$$

The Demjanov algorithm can also be used to heuristically solve maximin (obnoxious) location problems. The definition of $I(X)$ and the constraints in (1.26) are reversed.

1.5.5 Other Solution Methods

Drezner and Gavish (1985) solved the weighted minimax k-dimensional 1-center problem with n demand points within ϵ accuracy in $o(n \log \epsilon)$ time for a fixed dimensionality k using a variant of the "Russian method for solving linear programming" (Khachian 1979).

Drezner (2019) proposes to capitalize on symmetry that exists in some quadratic assignment instances when employing genetic or hybrid genetic algorithms to solve such instances. Such symmetry usually exists when the potential sites for the facilities are arranged in a rectangle. A simple and effective approach to identify equivalent solutions if such symmetry exists is designed. Three variants of this modification are proposed. Extensive computational experiments (performing 21,000 runs on instances with up to 150 facilities) show that the modified approach performed significantly better without increasing the run times.

Drezner and Drezner (2019c) consider solving various location problems using a trajectory solution approach (Drezner and Wesolowsky 1978c, 1982). For example, in the gravity model (Huff 1964, 1966) with a distance decay function $e^{-\lambda d}$, it is shown that when $\lambda \to 0$ the solution for locating a new facility is the solution to the Weber problem (Drezner et al. 2002b). Then, the trajectory of the solution points from $\lambda = 0$ to the required λ is found by numerically solving a set of differential equations by the Runge–Kutta method (Runge 1895; Ince 1926; Abramowitz and Stegun 1972).

1.6 The Quadratic Assignment Problem

An early review of the quadratic assignment problem (QAP) is Drezner et al. (2005). A recent book chapter, Drezner (2015b), summarizes many solution techniques of the QAP. My Ph.D. dissertation (Drezner 1975) which concentrated on the layout problem, and was published in Drezner (1980), is closely related to the QAP. I briefly summarize some of my contributions to this field.

- Drezner (1984b) found the optimal solution to the QAP Nug15 test problem (Nugent et al. 1968).
- Resende et al. (1995) and Drezner (1995b) analyzed lower bounds for the QAP based on linear programming.
- A new merging procedure in the genetic algorithm designed for the solution of the QAP was introduced in Drezner (2003).
- A new tabu search termed "concentric tabu search" was introduced in Drezner (2002). The concentric tabu search was applied on each offspring before considering it for inclusion into the population. The concentric tabu search was modified to the "extended concentric tabu" in Drezner (2005c). Drezner (2008a) proposed the "simple tabu" and performed extensive experiments comparing it to other genetic algorithms.

Recently, de Carvalho Jr. and Rahmann (2006) introduced a new class of quadratic assignment instances that are extremely difficult to solve. There are 14 instances in this set. Seven of them are Border Length Minimization which are denoted by BL followed by the number of facilities, and seven of them are Conflict Index Minimization denoted by CI followed by the number of facilities. In Table 1.4 we show the progression of the quality of the solutions (percentage above the best known solutions) in published papers.

Table 1.4 Results for de Carvalho Jr. and Rahmann (2006) instances

Instance	Best known	(1)	(2)	(3)	(4)	(5)	(6)	(7)
BL36	3296	1.699%	0%	–	0%	0%	0%	0%
BL49	4548	2.463%	0.352%	–	0%	0%	0%	0%
BL64	5988	3.540%	1.002%	–	0%	0%	0%	0%
BL81	7532	4.886%	1.487%	–	0.053%	0%	0%	0%
BL100	9256	4.624%	1.901%	–	0.173%	0.086%	0.086%	0%
BL121	11,396	5.581%	2.141%	–	0.140%	0.035%	0.035%	0%
BL144	13,432	5.688%	2.978%	–	0.298%	0.208%	0%	0.179%
CI36	168,611,971	0.779%	0.240%	0.055%	0%	0%	0%	0%
CI49	236,355,034	1.060%	0.306%	0%	0%	0%	0%	0%
CI64	325,671,035	0.645%	0.315%	0.178%	0%	0%	0%	0%
CI81	427,447,820	1.607%	0.289%	0.095%	0%	0%	0%	0%
CI100	523,146,366	1.802%	0.431%	–	0%	0%	0%	0%
CI121	653,409,588	1.642%	0.751%	–	0.001%	0%	0%	0%
CI144	794,811,636	2.304%	1.078%	–	0.025%	0%	0%	0%

(1) de Carvalho Jr. and Rahmann (2006); (2) Rodriguez et al. (2004); (3) Pelikan et al. (2007); (4) Drezner (2008b); Drezner and Marcoulides (2009); (5) Drezner and Misevičius (2013); (6) Drezner and Drezner (2018c); (7) Drezner and Drezner (2019d)

1.7 My Record and Some Reflections

Recently, an interview as a luminary discussing my career was posted on the INFORMS website: https://www.informs.org/Explore/History-of-O.R.-Excellence/Biographical-Profiles/Drezner-Zvi#oral_hist. I took a snapshot of my record on my 75th birthday on February 23, 2018. I put my 75th birthday vita on my website http://mihaylofaculty.fullerton.edu/sites/zdrezner/.

On that date I had 304 refereed publications: two books, 284 refereed journal articles, and 18 peer reviewed book chapters. I had a total of 95 co-authors, see Fig. 1.11, many contributing to this book. Two of my co-authors, Amnon Barak and Saharon Shelah, have co-authored with Paul Erdos and are Erdos-1 which makes me Erdos-2 and all my co-authors are Erdos-3 or higher. A list of Erdos-2 authors is available in http://www.oakland.edu/enp/thedata.html. According to Scholar.Google, one of my papers was cited 2189 times, 27 papers were cited at least 100 times, 61 papers were cited at least 50 times, and 204 papers were cited at least 10 times for a total of 14,468 citations of these 204 papers. I am deeply indebted to my co-authors without whom I could not have compiled such a record.

1. Aboolian, Robert
2. Agsari, Nasrin
3. Akella, Ram
4. Almogy, Yoram
5. Alp, Osman
6. Amar, Lior
7. Anklesaria, K.P.
8. Averbakh, Igor
9. Balakrishnan, N.
10. Barak, Amnon
11. Berman, Oded
12. Bhootra, Ajay
13. Brimberg, Jack
14. Chen, Frank
15. Choi D.
16. Cung V.D.
17. da Silviera, G.J.C.
18. Dear, Roger
19. Drezner, Tammy
20. Drezner, Taly Dawn
21. Eiselt, H.A.
22. Erikson, John
23. Erkut, Erhan
24. Espejo, I.
25. Estrada, J.
26. Farahani, Reza
27. Farnum, Nicholas
28. Gavish, Ben
29. Goldman, Alan J.
30. Goldstein, Zvi
31. Gurevich, Yuri
32. Gurnani, Haresh
33. Guyse, Jeffrey
34. Hahn, Peter M.
35. Hamacher, Horst
36. Hoefler, Torsten
37. Hulliger, Beat
38. Irawan, C.A.
39. Jefferson, Thomas
40. Kalczynski, Pawel
41. Kalsch, Marcel T.
42. Klamroth, Kathryn
43. Krass, Dmitry
44. Levin, Chaim
45. Levy, Ely
46. Lieber, M.
47. Marcoulides, G.A.
48. Marianov, Vladimir
49. Mehrez, Abraham
50. Menezes, Mozart B.C.
51. Minh, Paul
52. Misevicius, Alfonsas
53. Mladenovic, Nenad
54. Nakai M.
55. Nickel, Stefan
56. Nof, Shimon Y.
57. Okun, M.
58. Palubeckis, G.
59. Pasternack, B.A.
60. Plastria F.
61. Quinn, Kevin
62. Ramakrishnan, K.G.
63. Resende, M.G.C.
64. Rodriguez-Chia, A.M.
65. Ryan, Jeniffer
66. Salhi, Said
67. Sasaki, Mihiro
68. Schaible, Siegfried
69. Schöbel, Anita
70. Schumaker, R.E.
71. Schwarz, C.
72. Scott, Carlton H.
73. Shelah, Saharon
74. Shiloh, Amnon
75. Shiode, Shogo
76. Simchi-Levi, David
77. Song, Jeanette S.
78. Steiner, George
79. Steiner, Stefan
80. Stohs, Mark
81. Suzuki, Atsuo
82. Szendrovits, A.Z,
83. Taillard, Eric D.
84. Tamir, Arie
85. Thisse, Jacques-F,
86. Turel, Ofir
87. Turner, John
88. Wang, Qian
89. Wang, Jiamin
90. Welch, S. B.
91. Wesolowsky, George
92. Wiesner, W.
93. Zemel, Eitan
94. Zerom, Dawit
95. Ziegler, Hans Peter

Fig. 1.11 Co-authors list

1.7.1 Reflections

Traditionally we instruct our graduate students to first study the literature in order to know and understand available solution methods. I am not convinced this is the best approach for generating novel ideas and developing new methods. This promotes small, incremental changes or modifications rather than larger ideas. Sometimes, the most important new ideas come from an "outsider" who has different perspectives and background.

For example, the field of genetic algorithms can benefit from contributions of researchers who are not familiar with Operations Research. Mimicking natural processes is at the core of developing such algorithms (Drezner and Drezner 2005). The founder of genetic algorithms, John Henry Holland, received a B.S. degree in Physics from the Massachusetts Institute of Technology, then received an M.A. in Mathematics and was the first Computer Science Ph.D. from the University of Michigan. He did not get a "traditional" Operations Research training. The ideas of mimicking natural animal behavior of an alpha male or female's choice of a mate in genetic algorithms were suggested by my daughter Taly, who is an ecologist with no Operations Research training.

Many of my "off-location" papers have originated by talking to colleagues who described their projects. My solution approaches were, in many cases, different from "standard" approaches. Since I was thinking "outside the box," at least some of these ideas were superior to existing methods.

As mentors, we often give our students a problem and expect them to become good technicians and solve it using existing approaches rather than be innovative. When finding a new solution approach, one should then study the relevant literature on the subject and not the other way around. My experience as a Ph.D. student was different from what most Ph.D. students experience today. When I was given the project that eventually became my Ph.D. dissertation, I was not familiar with Operations Research and did not know anything about layout algorithms. Had I known the literature on facility layout, I would have probably used one of the available algorithms. My "out of the box" algorithm performed better than available algorithms. Throughout my career there were many times that I was not sure whether my ideas were already known or not. In most cases they were not.

In many of the papers we see today the author(s) take a problem, add some constraints or variables, produce an "impressive" non-linear and/or integer program, solve it with canned available software programs, report computational experiments, and produce a paper. Sometimes the paper looks like a summary of a dissertation. A competent modeler can create this formulation, then load it to available canned software and get results. What is the contribution of such an exercise to our knowledge-base or future work? This is the way we train most students who believe that this is what research is all about.

Let me illustrate my point using a paper that Pawel Kalczynski, Said Salhi, and I worked on (Drezner et al. 2018g). We investigate the multiple obnoxious facility location problem. We analyze two versions of the problem and I illustrate one of them here. Suppose that $n = 1000$ communities are located in a 100 by 100 miles square. $p = 20$ noisy or polluting factories or landfills need to be located in the area. These facilities are required to be at least $D = 16$ miles from one another to avoid cumulative nuisance to the communities. The objective is to maximize the minimum distance between facilities and the surrounding communities.

The problem is formulated as a non-convex non-linear optimization problem. The solutions are obtained by the interior point method in Matlab. The solution process is repeated 100 times for each instance from randomly generated solutions and the best one selected. Extensive computational experiments are reported for many values of n and p. For the instance of 20 facilities among 1000 demand points, a minimum distance of 0.38 miles between facilities and demand points is found. As part of the presentation we drew a three dimensional graph of the surface of the shortest distance to 100 demand points. See Fig. 1.12.

The three of us looked at the picture in amazement. There are 202 hilltops. It seems to be a reasonable approach to find solutions where facilities are located on hilltops! Recall that the objective is to maximize the "height" of the lowest point where facilities are located. We realize that these hilltops are Voronoi points that can be easily found by Mathematica. So, we formulated a problem of choosing the p tallest hilltops that are at least 16 miles from one another. This leads to a binary linear program (BLP). When these locations are found by optimally solving the BLP, a minimum distance of 4.02 miles is obtained for $p = 20$, $n = 1000$.

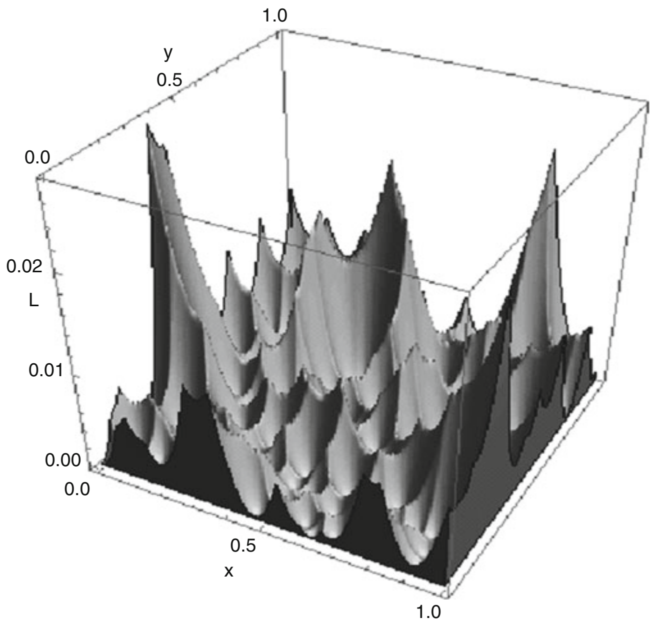


Fig. 1.12 Surface of distances to the closest community

This objective value is more than ten times better than the one found by Matlab! In addition, we solve the problem with the Voronoi heuristic in 24 s, while it takes over 5 h by interior point or SNOPT in Matlab.

We had difficulties in getting this paper published. It was finally accepted for publication in the third journal. I think that had we stopped before generating the 3D diagram, the paper would have fit the mold of the type of papers I described. The reviewers would have been happy with a solution of 0.38 miles. One referee said that SNOPT is the "state of the art" for this type of problem and recommended rejection. We tried SNOPT and got a solution of 1.6 miles. I am sure that if the original paper (using the interior point in Matlab) had been published, a follow-up paper reporting a fourfold improvement of the solution from 0.38 miles to 1.6 miles could be published as well. Nobody would even know that there is a solution of more than 4 miles.

References

Aboolian, R., Berman, O., & Drezner, Z. (2009). The multiple server center location problem. *Annals of Operations Research, 167*, 337–352.

Abramowitz, M., & Stegun, I. (1972). *Handbook of mathematical functions*. New York, NY: Dover Publications.

Agmon, S. (1954). The relaxation method for linear inequalities. *Canadian Journal of Mathematics, 6(3)*, 382–392.

Almogy, Y., Levin, C., & Drezner, Z. (1968). Simulation of the operations of a fleet. *Israel Journal of Technology, 6*, 283–287.

Alp, O., Drezner, Z., & Erkut, E. (2003). An efficient genetic algorithm for the p-median problem. *Annals of Operations Research, 122*, 21–42.

Amar, L., Barak, A., Drezner, Z., & Okun, M. (2009). Randomized gossip algorithms for maintaining a distributed bulletin board with guaranteed age properties. *Concurrency and Computation: Practice and Experience, 21*, 1907–1927.

Anklasaria, K. P., & Drezner, Z. (1986). A multivariate approach to estimating the completion time for PERT networks. *The Journal of the Operational Research Society, 37*, 811–815.

Barak, A., Drezner, Z., & Gurevich, Y. (1986). On the number of active nodes in a multicomputer system. *Networks, 16*, 275–282.

Barak, A., Drezner, Z., Levy, E., Lieber, M., & Shiloh, A. (2015). Dual-layer gossip algorithms for online management of exascale clusters. *Concurrency and Computation: Practice and Experience, 27*, 4797–4818.

Bennett, C. D., & Mirakhor, A. (1974). Optimal facility location with respect to several regions. *Journal of Regional Science, 14*, 131–136.

Berman, O., Drezner, T., Drezner, Z., & Krass, D. (2009a). Modeling competitive facility location problems: New approaches and results. In Oskoorouchi, M. (Ed.), *TutORials in operations research* (pp. 156–181). San Diego, CA: INFORMS.

Berman, O., Drezner, T., Drezner, Z., & Wesolowsky, G. O. (2009b). A defensive maximal covering problem on a network. *International Transactions on Operational Research, 16*, 69–86.

Berman, O., & Drezner, Z. (2000). A note on the location of an obnoxious facility on a network. *European Journal of Operational Research, 120*, 215–217.

Berman, O., & Drezner, Z. (2003). A probabilistic one-center location problem on a network. *Journal of the Operational Research Society, 54*, 871–877.

Berman, O., & Drezner, Z. (2006). Location of congested capacitated facilities with distance sensitive demand. *IIE Transactions, 38*, 213–221.

Berman, O., & Drezner, Z. (2007). The multiple server location problem. *Journal of the Operational Research Society, 58*, 91–99.

Berman, O., & Drezner, Z. (2008a). A new formulation for the conditional p-median and p-center problems. *Operations Research Letters, 36*, 481–483.

Berman, O., & Drezner, Z. (2008b). The p-median problem under uncertainty. *European Journal of Operational Research, 189*, 19–30.

Berman, O., Drezner, Z., & Krass, D. (2010a). Cooperative cover location problems: The planar case. *IIE Transactions, 42*, 232–246.

Berman, O., Drezner, Z., & Krass, D. (2010b). Generalized coverage: New developments in covering location models. *Computers & Operations Research, 37*, 1675–1687.

Berman, O., Drezner, Z., & Krass, D. (2011a). Big segment small segment global optimization algorithm on networks. *Networks, 58*, 1–11.

Berman, O., Drezner, Z., & Krass, D. (2011b). Discrete cooperative covering problems. *Journal of the Operational Research Society, 62*, 2002–2012.

Berman, O., Drezner, Z., & Krass, D. (2013). Continuous covering and cooperative covering problems with a general decay function on networks. *Journal of the Operational Research Society, 64*, 1644–1653.

Berman, O., Drezner, Z., & Krass, D. (2019). The multiple gradual cover location problem. *Journal of the Operational Research Society, 70*(6), 931–940.

Berman, O., Drezner, Z., Krass, D., & Wesolowsky, G. O. (2009c). The variable radius covering problem. *European Journal of Operational Research, 196*, 516–525.

Berman, O., Drezner, Z., Tamir, A., & Wesolowsky, G. O. (2009d). Optimal location with equitable loads. *Annals of Operations Research, 167*, 307–325.

Berman, O., Drezner, Z., & Wesolowsky, G. O. (2002). The collection depots location problem on networks. *Naval Research Logistics, 49*, 15–24.

Berman, O., Drezner, Z., & Wesolowsky, G. O. (2003). Locating service facilities whose reliability is distance dependent. *Computers & Operations Research, 30*, 1683–1695.

Berman, O., Drezner, Z., & Wesolowsky, G. O. (2005). The facility and transfer points location problem. *International Transactions in Operational Research, 12*, 387–402.

Berman, O., Drezner, Z., & Wesolowsky, G. O. (2007). The transfer point location problem. *European Journal of Operational Research, 179*, 978–989.

Berman, O., Drezner, Z., & Wesolowsky, G. O. (2008). The multiple location of transfer points. *Journal of the Operational Research Society, 59*, 805–811.

Berman, O., & Simchi-Levi, D. (1990). The conditional location problem on networks. *Transportation Science, 24*, 77–78.

Bhootra, A., Drezner, Z., Schwarz, C., & Stohs, M. H. (2015). Mutual fund performance: Luck or skill? *International Journal of Business, 20*, 52–63.

Brimberg, J., & Drezner, Z. (2013). A new heuristic for solving the p-median problem in the plane. *Computers & Operations Research, 40*, 427–437.

Brimberg, J., & Drezner, Z. (2015). A location-allocation problem with concentric circles. *IIE Transactions, 47*, 1397–1406.

Brimberg, J., & Drezner, Z. (2019). Solving multiple facilities location problems with separated clusters. In review.

Brimberg, J., Drezner, Z., Mladenovic, N., & Salhi, S. (2014). A new local search for continuous location problems. *European Journal of Operational Research, 232*, 256–265.

Brimberg, J., Drezner, Z., Mladenovic, N., & Salhi, S. (2017). Using injection points in reformulation local search for solving continuous location problems. *Yugoslav Journal of Operations Research, 27*, 291–300.

Campbell, J. F. (1994). Integer programming formulations of discrete hub location problems. *European Journal of Operational Research, 72*, 387–405.

Carathéodory, C. (2001). *Theory of functions of a complex variable*. Providence, RI: American Mathematical Society.

Chen, F., Drezner, Z., Ryan, J. K., & Simchi-Levi, D. (2000). Quantifying the bullwhip effect in a simple supply chain: The impact of forecasting, lead times, and information. *Management Science, 46*, 436–443.
Chen, P., Hansen, P., Jaumard, B., & Tuy, H. (1992). Weber's problem with attraction and repulsion. *Journal of Regional Science, 32*, 467–486.
Chen, R. (1988). Conditional minisum and minimax location-allocation problems in Euclidean space. *Transportation Science, 22*, 157–160.
Chen, R., & Handler, G. Y. (1993). The conditional p-center in the plane. *Naval Research Logistics, 40*, 117–127.
Chen, T.-Y. (1993). Calculation of the move limits for the sequential linear programming method. *International Journal for Numerical Methods in Engineering, 36*, 2661–2679.
Church, R. L., & Garfinkel, R. S. (1978). Locating an obnoxious facility on a network. *Transportation Science, 12*, 107–118.
Church, R. L., & ReVelle, C. S. (1974). The maximal covering location problem. *Papers of the Regional Science Association, 32*, 101–118.
Consul, P. C. (1989). *Generalized Poisson distribution: properties and applications*. New York, NY: Decker.
Contreras, I. (2015). Hub location problems. In Laporte, G., Nickel, S., & da Gama, F. S. (Eds.), *Location science* (pp. 311–344). Heidelberg: Springer.
Cooper, L. (1963). Location-allocation problems. *Operations Research, 11*, 331–343.
Cooper, L. (1964). Heuristic methods for location-allocation problems. *SIAM Review, 6*, 37–53.
Courtillot, M. (1962). New methods in mathematical programming-on varying all the parameters in a linear-programming problem and sequential solution of a linear-programming problem. *Operations Research, 10*, 471–475.
de Carvalho Jr., S. A., & Rahmann, S. (2006). Microarray layout as a quadratic assignment problem. In Huson, D., Kohlbacher, O., Lupas, A., Nieselt, K., & Zell, A. (Eds.), *Proceedings of the German Conference on Bioinformatics* (vol. 83, pp. 11–20). Bonn: Gesellschaft für Informatik.
Dear, R., & Drezner, Z. (1997). On the significance level of the multirelation coefficient. *Journal of Applied Mathematics & Decision Sciences, 1*, 119–131.
Demjanov, V. F. (1968). Algorithms for some minimax problems. *Journal of Computer and System Sciences, 2*, 342–380.
Drezner, T. (2004). Location of casualty collection points. *Environment and Planning C: Government and Policy, 22*, 899–912.
Drezner, T. (2007a). Establishment of casualty collection points for catastrophic events. *California Journal of Operations Management, 5*, 75–83.
Drezner, Z. (2007b). A general global optimization approach for solving location problems in the plane. *Journal of Global Optimization, 37*, 305–319.
Drezner, T. (2014). A review of competitive facility location in the plane. *Logistics Research, 7*, 1–12.
Drezner, T., & Drezner, Z. (2001). A note on applying the gravity rule to the airline hub problem. *Journal of Regional Science, 41*, 67–73.
Drezner, T., & Drezner, Z. (2005). Genetic algorithms: Mimicking evolution and natural selection in optimization models. In Bar-Cohen, Y. (Ed.), *Biomimetics—biologically inspired technologies* (pp. 157–175). Boca Raton, FL: CRC Press.
Drezner, T., & Drezner, Z. (2006). Gender specific genetic algorithms. *INFOR, Information Systems and Operations Research, 44*, 117–127.
Drezner, T., & Drezner, Z. (2007). Equity models in planar location. *Computational Management Science, 4*, 1–16.
Drezner, T., & Drezner, Z. (2011a). The gravity multiple server location problem. *Computers & Operations Research, 38*, 694–701.
Drezner, T., & Drezner, Z. (2011b). A note on equity across groups in facility location. *Naval Research Logistics, 58*, 705–711.

Drezner, T., & Drezner, Z. (2011c). The Weber location problem: The threshold objective. *INFOR: Information Systems and Operational Research, 49*, 212–220.

Drezner, T., & Drezner, Z. (2013). Voronoi diagrams with overlapping regions. *OR Spectrum, 35*, 543–561.

Drezner, T., & Drezner, Z. (2014). The maximin gradual cover location problem. *OR Spectrum, 36*, 903–921.

Drezner, T., & Drezner, Z. (2017). Leader-follower models in facility location. In Mallozzi, L., D'Amato, E., & Pardalos, P. (Eds.), *Spatial interaction models* (pp. 73–104). Cham: Springer.

Drezner, T., & Drezner, Z. (2018a). Asymmetric distance location model. In review.

Drezner, T., & Drezner, Z. (2018b). Cooperative cover of uniform demand. *Networks and Spatial Economics, 2018*, 1–3. https://doi.org/10.1007/s11067-018-9429-7.

Drezner, T., & Drezner, Z. (2019a). A directional approach to gradual cover: The continuous case. In review.

Drezner, T., & Drezner, Z. (2019b). Discrete budget allocation in competitive facility location. In preparation.

Drezner, T., & Drezner, Z. (2019c). The trajectory solution procedure and its application to solving a competitive facility location problem. In preparation.

Drezner, T., Drezner, Z., & Goldstein, Z. (2010a). A stochastic gradual cover location problem. *Naval Research Logistics, 57*, 367–372.

Drezner, T., Drezner, Z., & Guyse, J. (2009a). Equitable service by a facility: Minimizing the Gini coefficient. *Computers & Operations Research, 36*, 3240–3246.

Drezner, T., Drezner, Z., & Hulliger, B. (2014). The quintile share ratio in location analysis. *European Journal of Operational Research, 236*, 166–174.

Drezner, T., Drezner, Z., & Kalczynski, P. (2011). A cover-based competitive location model. *Journal of the Operational Research Society, 62*, 100–113.

Drezner, T., Drezner, Z., & Kalczynski, P. (2012a). Strategic competitive location: Improving existing and establishing new facilities. *Journal of the Operational Research Society, 63*, 1720–1730.

Drezner, T., Drezner, Z., & Kalczynski, P. (2015a). A leader-follower model for discrete competitive facility location. *Computers & Operations Research, 64*, 51–59.

Drezner, T., Drezner, Z., & Kalczynski, P. (2016a). The multiple markets competitive location problem. *Kybernetes, 45*, 854–865.

Drezner, T., Drezner, Z., & Kalczynski, P. (2018b). A gradual cover competitive facilities location model. In review.

Drezner, T., Drezner, Z., & Kalczynski, P. (2019). A directional approach to gradual cover. *TOP, 27*(1), 70–93. https://doi.org/10.1007/s11750-018-00493-y.

Drezner, T., Drezner, Z., & Kalczynski, P. (2019a). Directional approach to gradual cover: A maximin objective. In review.

Drezner, T., Drezner, Z., & Kalczynski, P. (2019b). Extensions to directional approach to gradual cover. In review.

Drezner, T., Drezner, Z., & Kalczynski, P. (2019c). The planar multifacility collection depots location problem. *Computers and Operations Research, 102*, 121–129.

Drezner, T., Drezner, Z., & Salhi, S. (2006). A multi-objective heuristic approach for the casualty collection points location problem. *Journal of the Operational Research Society, 58*, 727–734.

Drezner, T., Drezner, Z., & Schöbel, A. (2018c). The Weber obnoxious facility location model: A Big Arc Small Arc approach. *Computers and Operations Research, 98*, 240–250.

Drezner, T., Drezner, Z., & Scott, C. H. (2009b). Location of a facility minimizing nuisance to or from a planar network. *Computers & Operations Research, 36*, 135–148.

Drezner, T., Drezner, Z., & Shiode, S. (2002a). A threshold satisfying competitive location model. *Journal of Regional Science, 42*, 287–299.

Drezner, T., Drezner, Z., & Suzuki, A. (2018d). A cover based competitive facility location model with continuous demand. In review.

Drezner, T., Drezner, Z., & Zerom, D. (2018e). Competitive facility location with random attractiveness. *Operations Research Letters, 46*, 312–317.

Drezner, T., Drezner, Z., & Zerom, D. (2018f). Facility dependent distance decay in competitive location. In review.
Drezner, T., & Eiselt, H. A. (2002). Consumers in competitive location models. In Drezner, Z., & Hamacher, H. W. (Eds.), *Facility location: Applications and theory* (pp. 151–178). Berlin: Springer.
Drezner, Z. (1962). Methods for calculating a satellite orbit (in Hebrew). *The Stars in their Month, The Israeli Amateur-Astronomers Bulletin, 10*, 1–27. Reprint.
Drezner, Z. (1975). *Problems in non-linear programming (the allocation problem)*. PhD thesis, The Technion, Haifa, Israel.
Drezner, Z. (1978). Computation of the bivariate normal integral. *Mathematics of Computation, 32*, 277–279.
Drezner, Z. (1979). Bounds on the optimal location to the Weber problem under conditions of uncertainty. *Journal of the Operational Research Society, 30*, 923–931.
Drezner, Z. (1980). DISCON—a new method for the layout problem. *Operations Research, 28*, 1375–1384.
Drezner, Z. (1981a). On a modified one-center model. *Management Science, 27*, 848–851.
Drezner, Z. (1981b). On location dominance on spherical surfaces. *Operations Research, 29*, 1218–1219.
Drezner, Z. (1982a). Fast algorithms for the round trip location problem. *IIE Transactions, 14*, 243–248.
Drezner, Z. (1982b). On minimax optimization problems. *Mathematical Programming, 22*, 227–230.
Drezner, Z. (1983a). Constrained location problems in the plane and on a sphere. *IIE Transactions, 15*, 300–304.
Drezner, Z. (1983b). The nested ball principle for the relaxation method. *Operations Research, 31*, 587–590.
Drezner, Z. (1984a). A fading universe theory. *Speculations in Science and Technology, 7*, 125–128.
Drezner, Z. (1984b). On the quadratic assignment problem. *OMEGA, The International Journal of Management Science, 12*, 406–407.
Drezner, Z. (1984c). The p-center problem—heuristic and optimal algorithms. *Journal of the Operational Research Society, 35*, 741–748.
Drezner, Z. (1984d). The planar two-center and two-median problems. *Transportation Science, 18*, 351–361.
Drezner, Z. (1985a). O(nlogn) algorithm for the rectilinear round trip location problem. *Transportation Science, 19*, 91–93.
Drezner, Z. (1985b). Sensitivity analysis of the optimal location of a facility. *Naval Research Logistics Quarterly, 32*, 209–224.
Drezner, Z. (1985c). A solution to the Weber location problem on the sphere. *Journal of the Operational Research Society, 36*, 333–334.
Drezner, Z. (1986a). Location of regional facilities. *Naval Research Logistics Quarterly, 33*, 523–529.
Drezner, Z. (1986b). The p-cover problem. *European Journal of Operational Research, 26*, 312–313.
Drezner, Z. (1987a). A heuristic procedure for the layout of a large number of facilities. *Management Science, 33*, 907–915.
Drezner, Z. (1987b). Heuristic solution methods for two location problems with unreliable facilities. *Journal of the Operational Research Society, 38*, 509–514.
Drezner, Z. (1987c). On the complexity of the exchange algorithm for minimax optimization problems. *Mathematical Programming, 38*, 219–222.
Drezner, Z. (1987d). On the rectangular p-center problem. *Naval Research Logistics Quarterly, 34*, 229–234.
Drezner, Z. (1988a). Location strategies for satellites' orbits. *Naval Research Logistics, 35*, 503–512.

Drezner, Z. (1988b). Maximizing the minimum sight angle of a set of shapes. *IIE Transactions, 20*, 194–200.

Drezner, Z. (1989a). Conditional *p*-center problems. *Transportation Science, 23*, 51–53.

Drezner, Z. (1989b). Stochastic analysis of the Weber problem on the sphere. *Journal of the Operational Research Society, 40*, 1137–1144.

Drezner, Z. (1991). The weighted minimax location problem with set-up costs and extensions. *RAIRO - Operations Research, 25*, 55–64.

Drezner, Z. (1992a). Computation of the multivariate normal integral. *ACM Transactions on Mathematical Software, 18*, 470–480.

Drezner, Z. (1992b). A note on the Weber location problem. *Annals of Operations Research, 40*, 153–161.

Drezner, Z. (1994). Computation of the trivariate normal integral. *Mathematics of Computation, 62*, 289–294.

Drezner, Z. (1995a). Dynamic facility location: The progressive *p*-median problem. *Location Science, 3*, 1–7.

Drezner, Z. (1995b). Lower bounds based on linear programming for the quadratic assignment problem. *Computational Optimization & Application, 4*, 159–165.

Drezner, Z. (1995c). Multirelation—a correlation among more than two variables. *Computational Statistics and Data Analysis, 19*, 283–292.

Drezner, Z. (1995d). On the conditional *p*-median problem. *Computers & Operations Research, 22*, 525–530.

Drezner, Z. (1995e). Spherical approximations for minimax location problems. *Computers & Industrial Engineering, 28*, 893–897.

Drezner, Z. (1996). A note on accelerating the Weiszfeld procedure. *Location Science, 3*, 275–279.

Drezner, Z. (1997). Converting an area to discrete points. *Studies in Locational Analysis, 11*, 1–12.

Drezner, Z. (1998a). Finding whether a point is inside a polygon and its application to forbidden regions. *The Journal of Management Sciences & Regional Development, 1*, 41–48.

Drezner, Z. (1998b). On the repeated partition scheduling problem. *Journal of Business and Management, 5*, 65–77.

Drezner, Z. (1999). On a queue with correlated arrivals. *Journal of Applied Mathematics and Decision Sciences, 3*, 75–84.

Drezner, Z. (2001). The maximum possible EVPI. *INFORMS Transactions on Education, 1*, 117–118.

Drezner, Z. (2002). A new heuristic for the quadratic assignment problem. *Journal of Applied Mathematics and Decision Sciences, 6*, 163–173.

Drezner, Z. (2003). A new genetic algorithm for the quadratic assignment problem. *INFORMS Journal of Computing, 15*, 320–330.

Drezner, Z. (2005a). Compounded genetic algorithms for the quadratic assignment problem. *Operations Research Letters, 33*, 475–480.

Drezner, Z. (2005b). A distance based rule for removing population members in genetic algorithms. *4OR, 3*, 109–116.

Drezner, Z. (2005c). Extended concentric tabu for the quadratic assignment problem. *European Journal of the Operational Research, 160*, 416–422.

Drezner, Z. (2006a). Finding a cluster of points and the grey pattern quadratic assignment problem. *OR Spectrum, 28*, 417–436.

Drezner, Z. (2006b). On the limit of the generalized binomial distribution. *Communications in Statistics: Theory and Methods, 35*, 209–221.

Drezner, Z. (2008a). Extensive experiments with hybrid genetic algorithms for the solution of the quadratic assignment problem. *Computers & Operations Research, 35*, 717–736.

Drezner, Z. (2008b). Tabu search and hybrid genetic algorithms for quadratic assignment problems. In Jaziri, W. (Ed.), *Tabu search* (pp. 89–108). In-Tech. Retrieved from http://books.i-techonline.com.

Drezner, Z. (2009). On the convergence of the generalized Weiszfeld algorithm. *Annals of Operations Research, 167*, 327–336.

Drezner, Z. (2010a). On the unboundedness of facility layout problems. *Mathematical Methods in Operations Research, 72*, 205–216.
Drezner, Z. (2010b). Random selection from a stream of events. *Communications of the ACM, 53*, 158–159.
Drezner, Z. (2011). Continuous center problems. In Eiselt, H. A., & Marianov, V. (Eds.), *Foundations of location analysis* (pp. 63–78). Berlin: Springer.
Drezner, Z. (2013). Solving planar location problems by global optimization. *Logistics Research, 6*, 17–23.
Drezner, Z. (2015a). The fortified Weiszfeld algorithm for solving the Weber problem. *IMA Journal of Management Mathematics, 26*, 1–9.
Drezner, Z. (2015b). The quadratic assignment problem. In Laporte, G., Nickel, S., & da Gama, F. S. (Eds.), *Location science* (pp. 345–363). Cham, Springer.
Drezner, Z. (2019). Taking advantage of symmetry in some quadratic assignment problems. *INFOR: Information Systems and Operational Research*. https://doi.org/10.1080/03155986.2019.1607809.
Drezner, Z., & Barak, A. (1986). An asynchronous algorithm for scattering information between the active nodes of a multicomputer system. *Journal of Parallel and Distributed Computing, 3*, 344–351.
Drezner, Z., & Brimberg, J. (2014). Fitting concentric circles to measurements. *Mathematical Methods of Operations Research, 79*, 119–133.
Drezner, Z., Brimberg, J., Mladenovic, N., & Salhi, S. (2015b). New heuristic algorithms for solving the planar p-median problem. *Computers & Operations Research, 62*, 296–304.
Drezner, Z., Brimberg, J., Mladenovic, N., & Salhi, S. (2015c). Solving the planar p-median problem by variable neighborhood and concentric searches. *Journal of Global Optimization, 63*, 501–514.
Drezner, Z., Brimberg, J., Mladenovic, N., & Salhi, S. (2016b). New local searches for solving the multi-source Weber problem. *Annals of Operations Research, 246*, 181–203.
Drezner, Z., Drezner, T., & Wesolowsky, G. O. (2009c). Location with acceleration-deceleration distance. *European Journal of Operational Research, 198*, 157–164.
Drezner, Z., & Drezner, T. D. (2018c). Biologically inspired parent selection in genetic algorithms. In review.
Drezner, Z., & Drezner, T. D. (2019d). The alpha male genetic algorithm. *IMA Journal of Management Mathematics, 30*, 37–50.
Drezner, Z., & Erickson, J. (1997). On the January indicator. *Journal of Financial Education, 23*, 88–90.
Drezner, Z., & Erkut, E. (1995). Solving the continuous p-dispersion problem using non-linear programming. *Journal of the Operational Research Society, 46*, 516–520.
Drezner, Z., & Farnum, N. (1993). A generalized binomial distribution. *Communications in Statistics-Theory and Methods, 22*, 3051–3063.
Drezner, Z., & Farnum, N. (1994). A correlated Poisson distribution for correlated events. *Communications in Statistics: Theory and Methods, 23*, 841–857.
Drezner, Z., & Gavish, B. (1985). ϵ-approximations for multidimensional weighted location problems. *Operations Research, 33*, 772–783.
Drezner, Z., & Goldman, A. (1991). On the set of optimal points to the Weber problem. *Transportation Science, 25*, 3–8.
Drezner, Z., Gurnani, H., & Pasternack, B. A. (1995). An EOQ model with substitutions between products. *Journal of the Operational Research Society, 46*, 887–891.
Drezner, Z., & Guyse, J. (1999). Application of decision analysis techniques to the Weber facility location problem. *European Journal of Operational Research, 116*, 69–79.
Drezner, Z., Hahn, P. M., & Taillard, E. D. (2005). Recent advances for the quadratic assignment problem with special emphasis on instances that are difficult for meta-heuristic methods. *Annals of Operations Research, 139*, 65–94.
Drezner, Z., & Kalczynski, P. (2017). The continuous grey pattern problem. *Journal of the Operational Research Society, 68*, 469–483.

Drezner, Z., & Kalczynski, P. (2018). Solving non-convex non-linear programs with reverse convex constraints by sequential linear programming. In review.

Drezner, Z., Kalczynski, P., & Salhi, S. (2018g). The multiple obnoxious facilities location problem on the plane: A Voronoi based heuristic. *OMEGA: The International Journal of Management Science*. doi.org/10.1016/j.omega.2018.08.013.

Drezner, Z., Klamroth, K., Schöbel, A., & Wesolowsky, G. O. (2002b). The Weber problem. In Drezner, Z., & Hamacher, H. W. (Eds.), *Facility location: Applications and theory* (pp. 1–36). Berlin: Springer.

Drezner, Z., & Marcoulides, G. A. (1999). Applying simulated annealing for model selection in multiple regression and structural equation modeling. *Multiple Linear Regression Viewpoints, 25*, 1–4.

Drezner, Z., & Marcoulides, G. A. (2003). A distance-based selection of parents in genetic algorithms. In Resende, M. G. C., & de Sousa, J. P. (Eds), *Metaheuristics: Computer decision-making* (pp. 257–278). Boston, MA: Kluwer Academic Publishers.

Drezner, Z., & Marcoulides, G. A. (2006). Mapping the convergence of genetic algorithms. *Journal of Applied Mathematics and Decision Sciences, 11*, 1–16.

Drezner, Z., & Marcoulides, G. A. (2009). On the range of tabu tenure in solving quadratic assignment problems. In *Recent advances in computing and management information systems* (pp. 157–168). Athens: Athens Institute for Education and Research.

Drezner, Z., Marcoulides, G. A., & Salhi, S. (1999). Tabu search model selection in multiple regression analysis. *Communications in Statistics-Simulation and Computation, 28*, 349–367.

Drezner, Z., Marcoulides, G. A., & Stohs, M. H. (2001). Financial applications of a tabu search variable selection model. *Journal of Applied Mathematics and Decision Sciences, 5*, 215–234.

Drezner, Z., & Menezes, M. B. C. (2016). The wisdom of voters: Evaluating the Weber objective in the plane at the Condorcet solution. *Annals of Operations Research, 246*, 205–226.

Drezner, Z., & Minh, D. (2002). On the limited budget problem. *INFORMS Transactions on Education, 3*, 52–57.

Drezner, Z., & Misevicius, A. (2013). Enhancing the performance of hybrid genetic algorithms by differential improvement. *Computers & Operations Research, 40*, 1038–1046.

Drezner, Z., Misevicius, A., & Palubeckis, G. (2015d). Exact algorithms for the solution of the grey pattern quadratic assignment problem. *Mathematical Methods of Operations Research, 82*, 85–105.

Drezner, Z., & Misevičius, A. (2013). Enhancing the performance of hybrid genetic algorithms by differential improvement. *Computers & Operations Research, 40*, 1038–1046.

Drezner, Z., & Nickel, S. (2009a). Constructing a dc decomposition for ordered median problems. *Journal of Global Optimization, 45*, 187–201.

Drezner, Z., & Nickel, S. (2009b). Solving the ordered one-median problem in the plane. *European Journal of Operational Research, 195*, 46–61.

Drezner, Z., Nickel, S., & Ziegler, H.-P. (2012b). Stochastic analysis of ordered median problems. *Journal of the Operational Research Society, 63*, 1578–1588.

Drezner, Z., & Nof, S. Y. (1984). On optimizing bin picking and insertion plans for assembly robots. *IIE Transactions, 16*, 262–270.

Drezner, Z., & Salhi, S. (2002). Using hybrid metaheuristics for the one-way and two-way network design problem. *Naval Research Logistics, 49*, 449–463.

Drezner, Z., & Salhi, S. (2017). Incorporating neighborhood reduction for the solution of the planar p-median problem. *Annals of Operations Research, 258*, 639–654.

Drezner, Z., Schaible, S., & Simchi-Levi, D. (1990). Queuing-location problems on the plane. *Naval Research Logistics, 37*, 929–935.

Drezner, Z., & Scott, C. H. (1999). On the feasible set for the Weber problem and applications. *European Journal of Operational Research, 118*, 620–630.

Drezner, Z., & Scott, C. H. (2006). Locating a service facility with some unserviced demand. *IMA Journal of Management Mathematics, 17*, 359–371.

Drezner, Z., & Scott, C. H. (2010). Optimizing the location of a production firm. *Networks and Spatial Economics, 10*, 411–425.

Drezner, Z., & Scott, C. H. (2013). Location of a distribution center for a perishable product. *Mathematical Methods of Operations Research, 78*, 301–314.

Drezner, Z., Scott, C. H., & Song, J. S. (2003). The central warehouse location problem revisited. *IMA Journal of Management Mathematics, 14*, 321–336.

Drezner, Z., Scott, C. H., & Turner, J. (2016c). Mixed planar and network single-facility location problems. *Networks, 68*, 271–282.

Drezner, Z., & Shelah, S. (1987). On the complexity of the Elzinga-Hearn algorithm for the one-center problem. *Mathematics of Operations Research, 12*, 255–261.

Drezner, Z., & Simchi-Levi, D. (1992). Asymptotic behavior of the Weber location problem on the plane. *Annals of Operations Research, 40*, 163–172.

Drezner, Z., Steiner, S., & Wesolowsky, G. O. (2002c). On the circle closest to a set of points. *Computers & Operations Research, 29*, 637–650.

Drezner, Z., & Suzuki, A. (2004). The big triangle small triangle method for the solution of non-convex facility location problems. *Operations Research, 52*, 128–135.

Drezner, Z., & Suzuki, A. (2010). Covering continuous demand in the plane. *Journal of the Operational Research Society, 61*, 878–881.

Drezner, Z., Thisse, J.-F., & Wesolowsky, G. O. (1986). The minimax-min location problem. *Journal of Regional Science, 26*, 87–101.

Drezner, Z., & Turel, O. (2011). Normalizing variables with too-frequent values using a Kolmogorov-Smirnov test: A practical approach. *Computers & Industrial Engineering, 61*, 1240–1244.

Drezner, Z., Turel, O., & Zerom, D. (2010b). A modified Kolmogorov-Smirnov test for normality. *Communications in Statistics - Simulation and Computation, 39*, 693–704.

Drezner, Z., & Wesolowsky, G. O. (1978a). Facility location on a sphere. *Journal of the Operational Research Society, 29*, 997–1004.

Drezner, Z., & Wesolowsky, G. O. (1978b). A note on optimal facility location with respect to several regions. *Journal of Regional Science, 18*, 303.

Drezner, Z., & Wesolowsky, G. O. (1978c). A trajectory method for the optimization of the multifacility location problem with *lp* distances. *Management Science, 24*, 1507–1514.

Drezner, Z., & Wesolowsky, G. O. (1980a). Optimal location of a facility relative to area demands. *Naval Research Logistics Quarterly, 27*, 199–206.

Drezner, Z., & Wesolowsky, G. O. (1980b). The optimal sight angle problem. *AIIE Transactions, 12*, 332–338.

Drezner, Z., & Wesolowsky, G. O. (1981). Optimum location probabilities in the ℓ_p distance Weber problem. *Transportation Science, 15*, 85–97.

Drezner, Z., & Wesolowsky, G. O. (1982). A trajectory approach to the round trip location problem. *Transportation Science, 16*, 56–66.

Drezner, Z., & Wesolowsky, G. O. (1983). Minimax and maximin facility location problems on a sphere. *Naval Research Logistics Quarterly, 30*, 305–312.

Drezner, Z., & Wesolowsky, G. O. (1985). Layout of facilities with some fixed points. *Computers & Operations Research, 12*, 603–610.

Drezner, Z., & Wesolowsky, G. O. (1989). The asymmetric distance location problem. *Transportation Science, 23*, 201–207.

Drezner, Z., & Wesolowsky, G. O. (1990). On the computation of the bivariate normal integral. *Journal of Statistical Computation and Simulation, 35*, 101–107.

Drezner, Z., & Wesolowsky, G. O. (1991). The Weber problem on the plane with some negative weights. *Information Systems and Operational Research, 29*, 87–99.

Drezner, Z., & Wesolowsky, G. O. (1996). Obnoxious facility location in the interior of a planar network. *Journal of Regional Science, 35*, 675–688.

Drezner, Z., & Wesolowsky, G. O. (1997a). On the best location of signal detectors. *IIE Transactions, 29*, 1007–1015.

Drezner, Z., & Wesolowsky, G. O. (1997b). Selecting an optimum configuration of one-way and two-way routes. *Transportation Science, 31*, 386–394.

Drezner, Z., & Wesolowsky, G. O. (1998). Optimal axis orientation for rectilinear minisum and minimax location. *IIE Transactions, 30*, 981–986.
Drezner, Z., & Wesolowsky, G. O. (2001). On the collection depots location problem. *European Journal of Operational Research, 130*, 510–518.
Drezner, Z., Wesolowsky, G. O., & Drezner, T. (2004). The gradual covering problem. *Naval Research Logistics, 51*, 841–855.
Drezner, Z., & Zemel, E. (1992). Competitive location in the plane. *Annals of Operations Research, 40*, 173–193.
Drezner, Z., & Zerom, D. (2016). A simple and effective discretization of a continuous random variable. *Communications in Statistics - Simulation and Computation, 45*, 3798–3810.
Drezner, Z., & Zerom, D. (2018). A short note on tandem queues with correlated service times. In review.
Eiselt, H. A. (2011). Equilibria in competitive location models. In Eiselt, H. A., & Marianov, V. (Eds.), *Foundations of location analysis* (pp. 139–162). Boston, MA: Springer.
Eiselt, H. A., & Laporte, G. (1995). Objectives in location problems. In Drezner, Z. (Ed.), *Facility location: A survey of applications and methods* (pp. 151–180). New York, NY: Springer.
Eiselt, H. A., Marianov, V., & Drezner, T. (2015). Competitive location models. In Laporte, G., Nickel, S., & da Gama, F. S. (Eds.), *Location science* (pp. 365–398). Berlin: Springer.
Elzinga, J., & Hearn, D. (1972). Geometrical solutions for some minimax location problems. *Transportation Science, 6*, 379–394.
Estrada, J., & Drezner, Z. (2006). Simulating the spread of a virus. *INFORMS Transactions on Education, 6*, 32–36.
Eurostat (2012). Income quintile share ratio (s80/s20) (source: Silc). Eurostat Structural Indicators.
Farahani, R., Drezner, Z., & Asgari, N. (2009). Single facility location and relocation problem with time dependent weights and discrete planning horizon. *Annals of Operations Research, 167*, 353–368.
Francis, R. L., McGinnis Jr., L. F., & White, J. A. (1992). *Facility layout and location: An analytical approach* (2nd ed.). Englewood Cliffs, NJ: Prentice Hall.
García, S., & Marín, A. (2015). Covering location problems. In Laporte, G., Nickel, S., & da Gama, F. S. (Eds.), *Location science* (pp. 93–114). Heidelberg: Springer.
Gelenbe, E., Pujolle, G., & Nelson, J. (1998). *Introduction to queueing networks*. Chichester: Wiley.
Gini, C. (1921). Measurement of inequality and incomes. *The Economic Journal, 31*, 124–126.
Glover, F. (1977). Heuristics for integer programming using surrogate constraints. *Decision Sciences, 8*, 156–166.
Glover, F. (1986). Future paths for integer programming and links to artificial intelligence. *Computers & Operations Research, 13*, 533–549.
Glover, F., & Laguna, M. (1997). *Tabu search*. Boston, MA: Kluwer Academic Publishers.
Gnedenko, B., & Kovalenko, I. (1989). *Introduction to queuing theory. Mathematical modeling*. Boston, MA: Birkhauser.
Goldberg, D. E. (2006). *Genetic algorithms*. Delhi: Pearson Education.
Goldstein, Z., & Drezner, Z. (2007). Solving the repeated partition scheduling problem using metaheuristics. *Journal of Business and Management, 13*, 37–55.
Hansen, P., Jaumard, B., & Krau, S. (1995). An algorithm for weber's problem on the sphere. *Location Science, 3*, 217–237.
Hansen, P., Peeters, D., & Thisse, J.-F. (1981). On the location of an obnoxious facility. *Sistemi Urbani, 3*, 299–317.
Hilbert, D., & Cohn-Vossen, S. (1932). *Anschauliche geometrie*. Berlin: Springer. English translation published by Chelsea Publishing Company, New York (1956): Geometry and the Imagination.
Hoefler, T., Barak, A., Shiloh, A., & Drezner, Z. (2017). Corrected gossip algorithms for fast reliable broadcast on unreliable systems. In *2017 IEEE International Parallel and Distributed Processing Symposium (IPDPS)* (pp. 357–366). Piscataway: IEEE.

Holland, J. H. (1975). *Adaptation in natural and artificial systems*. Ann Arbor, MI: University of Michigan Press.

Huff, D. L. (1964). Defining and estimating a trade area. *Journal of Marketing, 28*, 34–38.

Huff, D. L. (1966). A programmed solution for approximating an optimum retail location. *Land Economics, 42*, 293–303.

Ince, E. L. (1926). *Ordinary differential equations*. Reprinted in 1956 by Dover Publications, USA.

Irawan, C., Salhi, S., & Drezner, Z. (2016). Hybrid meta-heuristics with VNS and exact methods: application to large unconditional and conditional vertex p-centre problems. *Journal of Heuristics, 22*, 507–537.

John, K., Ramakrishnan, C., & Sharma, K. (1987). Minimum weight design of trusses using improved move limit method of sequential linear programming. *Computers & Structures, 27*, 583–591.

Kalczynski, P., & Drezner, Z. (2019a). Locating multiple facilities using the max-sum objective. *Computers and Industrial Engineering, 129*, 136–143. https://doi.org/doi:10.1016/j.cie.2019.01.029.

Kalczynski, P., & Drezner, Z. (2019b). Packing ellipses in a rectangle. In preparation.

Kalczynski, P., & Drezner, Z. (2019c). A sequential heuristic for non-convex quadratically constrained programs. In preparation.

Kalsch, M., & Drezner, Z. (2010). Solving scheduling and location problems in the plane simultaneously. *Computers & Operations Research, 37*, 256–264.

Khachian, L. G. (1979). A polynomial algorithm for linear programming (English translation). *Soviet Mathematics Doklady, 20*, 191–194.

Kirkpatrick, S., Gelat, C. D., & Vecchi, M. P. (1983). Optimization by simulated annealing. *Science, 220*, 671–680.

Kuby, M. (1987). Programming models for facility dispersion: the p-dispersion and maxisum dispersion problems. *Geographical Analysis, 19*(4), 315–329.

Lee, D.-T., & Schachter, B. J. (1980). Two algorithms for constructing a Delaunay triangulation. *International Journal of Computer & Information Sciences, 9*, 219–242.

Lorenz, M. O. (1905). Methods of measuring the concentration of wealth. *Publications of the American Statistical Association, 9*, 209–219.

Love, R. F., Morris, J. G., & Wesolowsky, G. O. (1988). *Facilities location: models & methods*. New York, NY: North Holland.

Maranas, C. D., & Floudas, C. A. (1993). A global optimization method for Weber's problem with attraction and repulsion. In Hager, W. W., Hearn, D. W., & Pardalos, P. M. (Eds.), *Large scale optimization: State of the art* (pp. 259–293). Dordrecht: Kluwer.

Marcoulides, G. A., & Drezner, Z. (1993). A procedure for transforming points in multi-dimensional space to two-dimensional. *Educational and Psychological Measurement, 53*, 933–940.

Marcoulides, G. A., & Drezner, Z. (1999). A procedure for detecting pattern clustering in measurement designs. *Objective Measurement: Theory into Practice, 5*, 261–277.

Marcoulides, G. A., & Drezner, Z. (2003). Model specification searches using ant colony optimization algorithm. *Structural Equation Modeling, 10*, 154–164.

Marcoulides, G. A., & Drezner, Z. (2004). Tabu search variable selection with resource constraints. *Communications in Statistics - Simulation and Computation, 33*, 355–362.

Massey Jr, F. J. (1951). The Kolmogorov-Smirnov test for goodness of fit. *Journal of the American Statistical Association, 46*, 68–78.

Menezes, M. B. C., da Silveira, G., & Drezner, Z. (2016). Democratic elections and centralized decisions: Condorcet and approval voting compared with median and coverage locations. *European Journal of Operational Research, 253*, 195–203.

Minieka, E. (1980). Conditional centers and medians on a graph. *Networks, 10*, 265–272.

Motzkin, T., & Schoenberg, I. (1954). The relaxation method for linear inequalities. *Canadian Journal of Mathematics, 6*, 393–404.

Nelder, J. A., & Mead, R. (1965). A simplex method for function minimization. *The Computer Journal, 7*, 308–313.

Nickel, S., & Puerto, J. (2005). *Facility location—a unified approach*. Berlin: Springer.
Nof, S. Y., & Drezner, Z. (1986). Part flow in the robotic assembly plan problem. *Material Flow, 3*, 197–205.
Nof, S. Y., & Drezner, Z. (1993). The multiple-robot assembly plan problem. *Journal of Intelligent Robotic Systems, 5*, 57–71.
Nugent, C., Vollman, T., & Ruml, T. (1968). An experimental comparison of techniques for the assignment of facilities to locations. *Operations Research, 16*, 150–173.
Ogryczak, W., & Zawadzki, M. (2002). Conditional median: a parametric solution concept for location problems. *Annals of Operations Research, 110*, 167–181.
Okabe, A., Boots, B., Sugihara, K., & Chiu, S. N. (2000). *Spatial tessellations: Concepts and applications of Voronoi diagrams*. Wiley Series in Probability and Statistics. London: Wiley.
O'kelly, M. E. (1987). A quadratic integer program for the location of interacting hub facilities. *European Journal of Operational Research, 32*, 393–404.
Pasternack, B. A., & Drezner, Z. (1991). Optimal inventory policies for substitutable commodities. *Naval Research Logistics, 38*, 221–240.
Pasternack, B. A., & Drezner, Z. (1998). A note on calculating steady state results for an M/M/k queuing system when the ratio of the arrival rate to the service rate is large. *Journal of Applied Mathematics & Decision Sciences, 2*, 133–135.
Pasternack, B. A., & Drezner, Z. (1999). The videotape rental problem. *Journal of Applied Mathematics and Decision Sciences, 3*, 167–174.
Pelikan, M., Tsutsui, S., & Kalapala, R. (2007). Dependency trees, permutations, and quadratic assignment problem. In *Genetic And Evolutionary Computation Conference: Proceedings of the 9 th Annual Conference on Genetic and Evolutionary Computation* (vol. 7, pp. 629–629). Also Technical Report, Missouri Estimation of Distribution Algorithms Laboratory (MEDAL), Report No. 2007003.
Plastria, F. (1991). The effects of majority in Fermat-Weber problems with attraction and repulsion. *Yugoslav Journal of Operations Research, 1*, 141–146.
Plastria, F. (2002). Continuous covering location problems. In Drezner, Z., & Hamacher, H. W. (Eds.), *Facility Location: Applications and Theory* (pp. 39–83). Berlin: Springer.
Quinn, K., & Drezner, Z. (2017). Solving PERT/CPM using excel. *Operations Management Education Review, 11*, 107–114.
Reilly, W. J. (1931). *The law of retail gravitation*. New York, NY: Knickerbocker Press.
Resende, M. G. C., Ramakrishnan, K. G., & Drezner, Z. (1995). Computational experiments with the lower bound for the quadratic assignment problem based on linear programming. *Operations Research, 43*, 781–791.
ReVelle, C., Toregas, C., & Falkson, L. (1976). Applications of the location set covering problem. *Geographical Analysis, 8*, 65–76.
Rodriguez, J. M., MacPhee, F. C., Bonham, D. J., Horton, J. D., & Bhavsar, V. C. (2004). Best permutations for the dynamic plant layout problem. In Dasgupta, A. R., Iyengar, S. S., & Bhatt, H. S. (Eds.), *Efficient and Experimental Algorithms: Proceedings of the 12th International Conference on Advances in Computing and Communications (ADCOM 2004)* (pp. 173–178). New Delhi, Ahmedabad: Allied Publishers.
Rodríguez-Chía, A. M., Espejo, I., & Drezner, Z. (2010). On solving the planar k-centrum problem with Euclidean distances. *European Journal of Operational Research, 207*, 1169–1186.
Runge, C. (1895). Über die numerische auflösung von differentialgleichungen. *Mathematische Annalen, 46*, 167–178.
Russell, J. L. (1964). Kepler's laws of planetary motion: 1609–1666. *The British Journal for the History of Science, 2*, 1–24.
Sasaki, M., Suzuki, A., & Drezner, Z. (1997). On the selection of relay points in a logistics system. *Asia-Pacific Journal of Operational Research, 14*, 39–54.
Sasaki, M., Suzuki, A., & Drezner, Z. (1999). On the selection of hub airports for an airline hub-and-spoke system. *Computers & Operations Research, 26*, 1411–1422.
Schöbel, A., & Scholz, D. (2010). The big cube small cube solution method for multidimensional facility location problems. *Computers & Operations Research, 37*, 115–122.

Skorin-Kapov, D., Skorin-Kapov, J., & O'Kelly, M. (1996). Tight linear programming relaxations of uncapacitated p-hub median problems. *European Journal of Operational Research, 94*, 582–593.

Snyder, L. V. (2011). Covering problems. In Eiselt, H. A., & Marianov, V. (Eds.), *Foundations of location analysis* (pp. 109–135). Berlin: Springer.

Stackelberg, H. V. (1934). *Marktform und Gleichgewicht*. Vienne: Julius Springer.

Suzuki, A., & Drezner, Z. (1996). The p-center location problem in an area. *Location Science, 4*, 69–82.

Suzuki, A., & Drezner, Z. (1997). On the airline hub problem: The continuous model. *The Journal of the Operations Research Society of Japan, 40*, 62–74.

Suzuki, A., & Drezner, Z. (2009). The minimum equitable radius location problem with continuous demand. *European Journal of Operational Research, 195*, 17–30.

Suzuki, A., & Drezner, Z. (2013). Solving constrained two-facility location problems. *The Journal of the Operations Research Society of Japan, 56*, 157–165.

Suzuki, A., & Okabe, A. (1995). Using Voronoi diagrams. In Drezner, Z. (Ed.), *Facility location: A survey of applications and methods* (pp. 103–118). New York, NY: Springer.

Sylvester, J. (1857). A question in the geometry of situation. *Quarterly Journal of Mathematics, 1*, 79.

Sylvester, J. (1860). On Poncelet's approximate linear valuation of Surd forms. *Philosophical Magazine, 20* (Fourth series), 203–222.

Szabo, P. G., Markot, M., Csendes, T., & Specht, E. (2007). *New approaches to circle packing in a square: With program codes*. New York: Springer.

Taillard, É. D. (1995). Comparison of iterative searches for the quadratic assignment problem. *Location Science, 3*, 87–105.

Voronoï, G. (1908). Nouvelles applications des paramètres continus à la théorie des formes quadratiques. deuxième mémoire. recherches sur les parallélloèdres primitifs. *Journal für die reine und angewandte Mathematik, 134*, 198–287.

Weber, A. (1909). *Über den Standort der Industrien, 1. Teil: Reine Theorie des Standortes. English Translation: on the Location of Industries*. Chicago, IL: University of Chicago Press (Translation published in 1929).

Weiszfeld, E. (1936). Sur le point pour lequel la somme des distances de n points donnes est minimum. *Tohoku Mathematical Journal, 43*, 355–386.

Weiszfeld, E., & Plastria, F. (2009). On the point for which the sum of the distances to n given points is minimum. *Annals of Operations Research, 167*, 7–41. (English Translation of Weiszfeld (1936)).

Welch, S. B., Salhi, S., & Drezner, Z. (2006). The multifacility maximin planar location problem with facility interaction. *IMA Journal of Management Mathematics, 17*, 397–412.

Wesolowsky, G. O. (1993). The Weber problem: History and perspectives. *Location Science, 1*, 5–23.

Zangwill, W. I. (1969). *Nonlinear programming: A unified approach*. Englewood Cliffs, NJ: Prentice-Hall.

Chapter 2
Understanding the Weber Location Paradigm

Richard L. Church

2.1 Introduction

Location Science as a field started with the developments of von Thünen (1826), Launhardt (1872), Weber (1909), Christaller (1933), Hotelling (1990), Hoover (1937), and Lösch (1940). It further expanded with the developments of Isard (1949), Koopmans (1951a, b), Koopmans and Beckmann (1957), Moses (1958), Cooper (1963), Hakimi (1964, 1965), Balinski (1965), and Beckmann (1968). The problem of writing about Weber is that almost everyone in Location Science knows something about his "model." What could possibly be new that has not been included in previous assessments, especially given that his book has been in print for over 100 years? That being said, the objective of this paper is to demonstrate from the perspective of the field of location science that Weber has been pigeonholed, misunderstood, and under-appreciated.

Weber (1909) wrote that in each industry, there must be a "somewhere" as well as a "somehow" in terms of production, distribution, and consumption. He constructed a simple example that involved locating a factory that needed two raw materials that are sourced at specific locations along with a market that will be supplied with what is produced. This is depicted in Fig. 2.1, now known as a location triangle.

He assumed that transport costs for the two raw materials and the finished product could be estimated as the amount that is transported times the distance that the materials or product are carried (essentially weight times distance, e.g., ton-miles). His objective was to locate the facility at a place where total transport costs would be minimized. An underlying assumption is that transport distances were to be estimated using Euclidean distance measure. This summarizes the view of what

R. L. Church (✉)
Department of Geography, University of California at Santa Barbara, Santa Barbara, CA, USA
e-mail: Rick.church@ucsb.edu

H. A. Eiselt, V. Marianov (eds.), *Contributions to Location Analysis*, International Series in Operations Research & Management Science 281,
https://doi.org/10.1007/978-3-030-19111-5_2

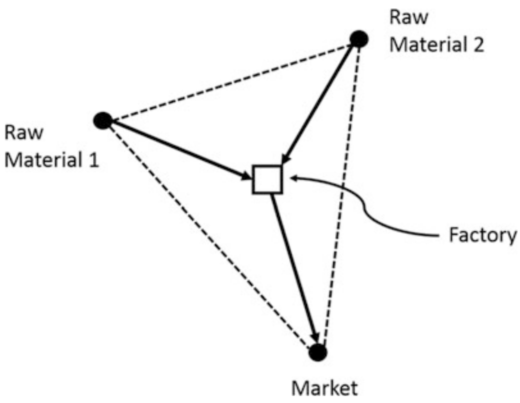


Fig. 2.1 Depiction of a classic location triangle with two raw materials and one market

Weber accomplished by most people in Location Science. There are reasons for this viewpoint and this will be explained below, however, before we describe a more complete Weber location paradigm, it is instructive to first discuss the historical roots of the "location triangle." Following that will be a discussion of Ackoff's definition of the phases of OR as a science. We use Ackoff's definition as a focal lens to understand one way that we can analyze Weber's book on industrial location. With this approach, we present a more complete set of Weber-defined location problems. We end with possible suggestions for research.

2.2 Historical Roots of the Planar Median Location Problem

Almost everyone credits Fermat as posing the following simple geometric problem in the 1600s: given three points on the plane, find the position of a fourth point which minimizes the sum of the distances from that fourth point to the three given points. It is quite possible Cavalieri or someone else originally suggested this problem (Wesolowsky 1993), but as Wesolowsky states: "the history of this problem is a bit murky." This mathematical puzzle has been solved a number of times with slightly different assumptions. Torricelli has been given the credit for first having solved this puzzle using several different geometric construction approaches. Simpson (1750) in his book on fluxions posed a similar problem, except that each of the three given points was assigned weights. Instead of finding a fourth point which minimized the sum of distances to the other three points, he proposed locating the fourth point so that the sum of the weighted distances to the other three points is minimized. Simpson also proposed a geometric construction method in solving this problem. Worth noting is that Krarup and Vajda (1997) provide an interesting discussion of the Fermat problem with respect to Torricelli and Simpson. It is also important to underscore the fact that these early geometric problems involved three given points and the problem of seeking the location of a fourth point. It is also important to note that Varignon (1687) proposed a mechanical frame of pulleys and weights that could be used to solve this problem. His apparatus is called a Varignon frame and a drawing of that can be found in the book by Weber (1909).

From an entirely different perspective, Launhardt (1872) proposed a transportation problem that involves connecting three locations in the following example. Suppose that there is a foundry that produces pig iron from coal and iron ore. If iron ore is sourced from point A, coal is sourced from point B and the pig iron is to be delivered to point C, where should the foundry be located in such a manner as to minimize the costs of transportation. Launhardt (1872) described this as both a transportation investment as well as a cargo hauling problem. Consider the following notation:

U is the annual interest of the capital costs, the yearly maintenance costs per kilometer, and all installation costs of the conveyance system (road or rail)
c symbolizes the transportation cost per ton per kilometer
d_i denotes the distance from point i to location at (x, y) where $i = A, B$, or C
w_i^S is the volume (tons) of annual traffic needed to be hauled from source where $i = A$ or B
w_C^M is the volume (tons) of annual traffic needed in supplying the market C

Using this notation, we can pose the following transportation design and location problem: find the point (x, y) which minimizes:

$$Z = \left(U + cw_A^S\right) d_A + \left(U + cw_B^S\right) d_B + \left(U + cw_C^M\right) d_C \qquad (2.1)$$

The three points A, B, and C form the so-called location triangle mentioned in the introduction. If we view the cost components as the weights associated with a given source material or market demand, this is exactly the same problem that was solved by Simpson (1750). Depending upon the relative value of the weights, the solution is ether a point in the interior of the triangle or at the boundary of the triangle and could even be at one of the original three points. Thus, it seems fitting to call the economic version of this location problem associated with the triangle of points, the *Launhardt* problem. Hidden in this literature is an analytic solution proposed by Launhardt that has since been used in road network design in forestry (Greulich 1999).

Just where does Weber fit into this development? He too developed what was depicted as a location triangle, but somewhat later. In 1909, Weber published a treatise called "Theory of the Location of Industries." In that work, Weber describes a problem like that of Launhardt, except Weber ignores the cost of the infrastructure (road or rail), but otherwise restates what Launhardt proposed years earlier. But, there is more to this history than describing the location triangle, especially with respect to Weber's work. Before we dig into the details of Weber's constructs, it is important to first cover a bit more of the history of the development of the problem of locating a point on a plane in order to minimize the lengths of connecting lines or the costs of transport.

Weiszfeld (1937) was the first to propose an algorithm to solve a problem that was expanded to locating a point on the plane while minimizing the weighted distances to n other points. The approach of Weiszfeld was later rediscovered independently by Miehle (1958), Kuhn and Kuenne (1962), Cooper (1963), and Vergin and Rogers (1967). A very nice presentation of this history, including the works of Weber, Launhardt, and Weiszfeld and the developments which followed Weiszfeld, can be found in Eckhardt (2008). Eckhardt (2008) describes in detail various mathematical and geometric approaches that have been developed to solve this problem.

Overall, this simple problem has been classified as the planar median location problem. However, Kuhn and Kuenne (1962) formally stated that they were solving a generalized form of the Weber problem, generalized in the sense that there were n points as compared to what had been viewed as a three-point problem. Kuhn and Kuenne's designation seems to have stuck, as most subsequent work has called this problem the Weber or generalized Weber problem. Wersan et al. (1962) suggested an application for the planar median location problem where distances are measured using the Manhattan metric (rectilinear or grid distance) in siting a municipal solid waste incinerator and demonstrated that it could be formulated as a simple linear programming problem. Vergin and Rogers (1967) proposed a simple algorithm that can be used to solve this problem. Since then many variations for the planar median problem have been proposed, including forms which contain generalized distance metrics (Brimberg and Love 1993), line barriers (Klamroth 2001), forbidden regions (Katz and Cooper 1981; Butt and Cavalier 1996), negative weights (Drezner and Wesolowsky 1991), and an inclined plane (Drezner and Wesolowsky 1989), just to name a few. There has also been innovative work on solution procedures (see, for example, Rosing 1992; Brimberg et al. 2000; Salhi and Gamal 2003; Drezner and Suzuki 2004). It seems fitting to mention here that Drezner is perhaps the most well-known of the researchers that have worked on various forms of the Weber and planar median problems since the late 1970s. Over the last 40 years, Drezner has published 18 papers that detailed special algorithms and proposed new constructs for planar median/Weber problems, and has garnered an average of 40 citations per publication.

Most people in the field of Location Science who are somewhat familiar with Weber seem to source material gleaned from an appendix of Weber's book written by Georg Pick or from some secondary source. Pick produced a geometrical construct that can be used to solve the three-point problem. It is this part of Weber's book that gets into specific details of analytics rather than a discussion of all of the elements and the issues involved in selecting a location. Outside of the purview of economic geographers, most of the descriptive details have been ignored, and Weber is viewed as having addressed only a three-point location triangle problem. After all, that is why Kuhn and Kuenne called their problem the generalized Weber problem because they had developed a solution approach that was not restricted to three points, but generalized in the sense that it dealt with n points. Cooper (1963) made this distinction as well when he proposed siting multiple facilities among *n*-demand points on the plane. He developed a heuristic to locate a fixed number of facilities in order to minimize total weighted distance while assigning each demand point to their closest located facility. Most would recognize this as the planar *p*-median problem. Whereas the classic one-facility median location problem is a pure location problem, the multi-facility location problem proposed by Cooper is a more complex location-allocation problem. Cooper (1963) in his presentation of this planar location problem makes several points when discussing the relationship between his work and that of Weber. First, he states that Weber restricted his work to three points, whereas his model contained a larger set, *n,* of demand points. Second, he stated: "only a single source is considered" in Weber. Cooper viewed facilities as sources (or suppliers), and he made the distinction that Weber located only one facility within the location triangle paradigm. The assessments of both Cooper and Kuhn and Kuenne about Weber are patently false and demonstrate a view that is sourced primarily from reading Pick's appendix and not the book itself.

Before moving on to a more complete presentation of the Weber paradigm, it is fitting to discuss a bit more about Launhardt. Without doubt, there is a growing understanding of what Launhardt accomplished. Many may ask why his important work was overlooked, and the economic form of the location triangle problem was not named in honor of Launhardt. First, Launhardt's work was published in German, and much of it was read or studied by civil engineers. Second, Weber appeared to have not known about Launhardt when he wrote his book. Third, Weber's book although written in German was translated into English in 1929 and was published by the University of Chicago press, making his work known to a much wider audience than Launhardt's. But, the importance of Launhardt's work has not gone completely un-noticed. Isard (1949), Beckmann (1955), Isard and Reiner (1962), Kuhn and Kuenne (1962) all recognized and discussed Launhardt's location triangle problem. In wasn't until Pinto (1977) wrote a lengthy article on Launhardt and his neglected book, did others begin to take notice. More recently, Laporte et al. (2015) in their introduction to a book on Location Science note Launhardt's development of the location triangle discussed above and discuss his approach in solving it. Finally, to gain a better perspective between Launhardt and Weber, one should consult the paper by Perreur (1998) in which there is a section titled: "Should Weber be forgotten?" In answering this question, Perreur (1998) digs into various elements of industrial location that were raised by Weber that clearly place his thinking on certain elements, such as labor and the tendencies to agglomerate specific activities, as advances over that of Launhardt. This paper addresses different aspects of Weber's industrial location work, which I will call here Weber's Industrial Location Paradigm, which is far more nuanced than relegating it to the so-called location triangle. More importantly, my approach here is decidedly different than Perreur (1998) and his comments and assessments can easily be added to what I offer here.

2.3 Location Science, Ackoff, and Weber

Location science is a discipline that has emerged principally in the fields of operations research (OR), economics, geography, and engineering. Many of the constructs are based upon a model and most normative constructs are cast as mathematical programming problems, some as linear programming problems, most as integer linear programming problems, and others involving bi-level programming, quadratic programming, among others. Ackoff (1956) in his describing the development of OR as a science noted that there were six phases of an OR project:

(1) "Formulating the problem," a verbal statement of what is needed;
(2) "Constructing a mathematical model to represent the system being studied";
(3) "Deriving a solution from the model";
(4) "Testing the model and its solution, ... checking against reality";
(5) "Establishing controls over the solution", determining the conditions under which the solution could be implemented; and
(6) "Putting the solution to work, implementation."

Most professionals in the field of OR follow this approach in general, although many may not see a problem being approached through all six phases. In fact, many of the papers in the literature in the past 50 years describe only the first three phases: problem exposition, model formulation, and model solution. Others concentrate on proposing a new solution approach for an existing problem. That is, nearly all the papers in our field involve contributing to one or perhaps more phases of Ackoff's paradigm. We can look back in the historical roots and see very good examples of this. For example, Fermat is credited with proposing the initial location problem of finding a fourth point which minimizes the combined distances to the other three points. Torricelli has been credited with being the first to structure and solve this problem geometrically. From my perspective, Weber's work is more complex as he described not just the simple location triangle problem but a series of interrelated problems. His descriptions are quite illuminating with respect to Ackoff's phase 1 of OR. That is, Weber's work viewed through the lens of Ackoff was a leader in the nascent field of location science, by having suggested a number of problems that are faced by industry that are central to the decision as to where to place a manufacturing plant. The remainder of this presentation is devoted to identifying a set of location problems that were described by Weber, in essence phase 1 of Ackoff's paradigm. We also move from Weber's verbal description to the next phase of analysis, that of model construction. We will leave for others further development, such as identifying appropriate solution techniques for these problems. Through this approach, we will sketch out the Weber paradigm and see that his work has been mostly "pigeonholed" as the Location Science community has not really looked beyond what is now viewed as Launhardt's location triangle.

2.4 A More Complete Form of the Weber Location Paradigm

As stated before, Weber wrote that in each industry there must be a "somewhere" as well as a "somehow" in terms of production, distribution, and consumption. Without going into a detailed discussion that is provided by Weber, the central issue is to determine how transportation costs influence the distribution of industries. Rather than discuss the issue with respect to many companies, each making their own decision, this discussion will assume that a company is making a decision on where to locate one or more facilities. As Weber states, industries "will be drawn to those locations which have the lowest costs of transportation" with regard to the transport of materials that are needed in the manufacturing of an item as well as the transport costs associated with shipping the product to market. Weber points out that shipping rates are a function of weight, distance, volume, the transportation facility (rail, road, etc.), and special properties. For example, if an item is bulky, more railcars may be necessary with a concomitant increase in transport rates. Weber suggested that many of the issues of transportation costs can be represented by the tonnage being transported times the distance of that transport, and multiplied by a rate per ton-mile. He recognized that there were nuances of transport costs that could not be handled exactly by calculating costs as a transport rate times the ton-miles of travel, but that this approach could cover many of the special circumstances one might encounter. So, he boiled his main question down to "how will places of minimum ton-miles actually distribute the production?"[1]

There are two types of raw materials needed in the production of something, localized and ubiquitous. Localized materials are geographically present at specific locations where ubiquitous materials are present everywhere. Ubiquitous materials require no transport except what is present in the final product, whereas localized raw materials (materials taken from local deposits) must be transported to the industrial plant. So, transportation costs boil down to transporting localized raw materials and the finished product for a problem involving a simple problem of production. The location triangle of Fig. 2.1 depicts the problem of having two localized raw materials and one market. Weber's discussion covers the location triangle at some depth, in terms of shape as well as characteristics that describe a particular orientation for the industry, market orientation or material orientation. For example, weight losing materials are those materials that have little of their initial weight being incorporated into the finished product. Weber states that weight losing materials "may pull production to their deposits."[2] The opposite is true perhaps for "pure" raw materials in which no weight is lost in production or when a ubiquitous material makes up a substantial proportion of the weight of the finished product. For example, soft drink bottling plants are likely to be located close to the market, as most of the weight and volume of the product is associated with the water that makes

[1]This quote can be found on page 48 of Weber's Theory of the Location of Industries, translated into English in 1929 by Carl J. Friedrich.

[2]See discussion on page 61, of the English translation of Weber (1929).

up the product, and not the syrup. Thus, the location of different types of factories will be based upon the relative values of material and product flows at the point of production and depending upon these values, the ultimate location may likely be oriented toward a market or some raw material.

More important is the fact that Weber acknowledges that there could be more than two raw materials[3] and even more than one market.[4] The location triangle is in some ways meant to be used as a construct to show the orientation that production might take in a simple problem. He states that more complicated problems can be addressed, e.g., more than two raw materials and one market, by Varignon's frame (see footnote 3). Consider the following notation which will be used throughout the rest of this paper in this or modified form:

i is an index used to refer to different localized raw materials, $i = 1, 2, \ldots, m$
j is an index used to refer to different locations of product consumption (markets), $j = 1, 2, \ldots, n$
wrm_i denotes the weight of raw material i needed per unit of finished product produced
wfp symbolizes the weight of finished product
a_i is the amount of finished product needed at market j where $A = \sum_j a_j$
(x_i, y_i) is the location of raw material i
$(x_j y_j)$ is the location of market j
(x, y) symbolizes the location of production plant (to be determined)
$t_i = ((x - x_i)^2 + (y - y_i)^2)^{½}$ is the Euclidean distance between the production facility and raw material i
$d_j = ((x - x_j)^2 + (y - y_j)^2)^{½}$ is the Euclidean distance between the production facility and market location j

Formally, we can define this pure location problem as:

Weber Model 1: The Classic Location Model with Multiple Source Materials and Multiple Markets

$$\text{Min } Z = A \sum_{i=1}^{m} wrm_i \; t_i + \sum_{j=1}^{n} wfp \; a_j d_j \tag{2.2}$$

The objective is to minimize all transportation (in ton-miles) of raw materials being shipped to the production plant (first term) and the distribution of product to various markets from the production plant (second term) by finding the best location (x, y) for the production plant. If we have costs in terms of per ton-mile or costs per unit product per mile of transport, we can introduce those cost factors into this problem definition so that costs of transport are minimized instead of ton-miles of

[3] See figure 9 on page 64 and related discussion of the English translation, Weber (1929).

[4] See figure 18 on page 115, also note on page 71 of Weber (1929) where Weber notes that a given raw material may be used by other production facilities serving other places of consumption. Clearly, his view included multiple markets and multiple sources of raw material, given one product type.

transport. Technically speaking, if we have a problem in which there are no localized raw materials to be shipped, the above form is essentially the same as that which was solved by Weiszfeld (1937). That form would be equivalent as well to what was solved by Kuhn and Kuenne (1962) and called the generalized Weber problem, where the number of demand points was some number *n*. The Weiszfeld algorithm and the approach by Kuhn and Kuenne can be used to solve Weber Model 1 as there does not need to be a distinction between raw materials and markets in this simple case, but the distinction is made here as other forms of Weber's paradigm rest on the distinction between the transportation of raw materials and the distribution of finished product. When the problem is represented by two raw materials and one market, it devolves to the Launhardt location triangle. But, note that what Kuhn and Kuenne call the generalized Weber problem is really not a generalization when formulating what Weber described within the context of Ackoff's phase 1 of OR.

Weber almost always attempts to describe a given problem situation within the simple context of two raw materials and one market (i.e., the locational figure) even though he states that there could be more such materials and more markets. In fact, he uses the term location(al) figure rather than triangle so that he hasn't placed limits on the number of resources and markets. He states: "while the locational figures will always be individual or specific for a particular plant, these weight figures are general, applying to all plants of the same production."[5] That is, he acknowledges that the locational figures can be drawn about each plant, a sort of plant specific resource and service area. But, he also acknowledges that there could be more plants, each producing the same product, each with their own resource and service area. The main focus appears to be on identifying the right place of each plant within its own service area, but not formally stated is the fact that in this landscape of plants there must be some process of location across all of the production. It is my belief that if Weber were working in today's environment of OR, the natural progression would have been to locate multiple plants simultaneously.

2.4.1 Weber Model 2: Single Plant Location with Alternate Sources of Raw Materials

Weber describes the case of having alternate sources of the same raw material and within the context of two location triangles reasons that there could be two places of production with equally low transport costs, where one potential production site is served by one of the local sites for that raw material and the other potential production site is served by the other local site for that raw material. With this, he reasoned that production sites would source their raw materials from the cheapest sources for that site and that these allocations would change depending on the locations of a given type of raw material as well as plant location. For example, wallboard in the USA is made of gypsum, a soft white mineral. It is mined in 19

[5] This quote can be found on page 55, Weber (1929).

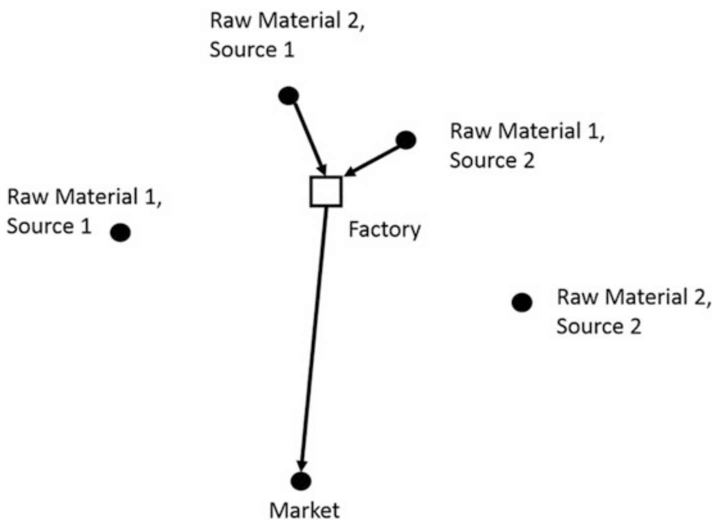


Fig. 2.2 Depiction of multiple sources of each raw material, where the source of a given raw material is based upon the location of the factory. This requires an allocation component to the associated model. Note the figure is a simple depiction as the problem could easily involve multiple markets

states. If you were locating a wallboard manufacturing plant, to serve a given market area, it is likely that you would find the closest/cheapest available supply of gypsum for this needed raw material. Figure 2.2 depicts a problem with two raw materials and one market, where each of the two raw materials can be sourced from several different locations.

A production plant has been placed at a specific location and its raw materials are assigned from its closest source for each raw material. To address this in a model, we need to introduce allocation variables that represent the assignment of raw material sources to the production plant. Consider then, the enhanced notation:

Θ_i is the number of localized sources for raw material i

ℓ denotes an index used to represent a given localized raw material, where $\ell = 1, 2, \ldots, \Theta_i$

$\left(x_i^\ell, y_i^\ell\right)$ is the location of the ℓ - th source location for raw material i

$$t_i^\ell = \left(\left(x - x_i^\ell\right)^2 + \left(y - y_i^\ell\right)^2\right)^{\frac{1}{2}}$$

$$s_i^\ell = \begin{cases} 1, \text{if source } \ell \text{ for raw material } i \text{ is used to supply the factory} \\ 0, \text{otherwise} \end{cases}$$

Using this notation, we can define a new form of the Weber problem, where raw material sourcing is a function of the plant location.

$$\text{Min } A = A\sum_{i=1}^{m}\sum_{\ell=1}^{\Theta_i} wrm_i\, t_i^{\ell}\, s_i^{\ell} + \sum_{j=1}^{n} wfp\, a_j\, d_j \tag{2.3}$$

subject to:

$$\sum_{\ell=1}^{\Theta_i} s_i^{\ell} = 1 \quad \text{for each } \; i = 1, 2, \ldots, m \tag{2.4}$$

$$s_i^{\ell} \in \{0, 1\} \quad \text{for each} \quad i = 1, 2, \ldots, m \quad \text{and} \quad l = 1, 2, \ldots, \Theta_i \tag{2.5}$$

The objective of the problem is to determine the location of the production plant (x, y) along with assigning/allocating raw material sources to the production plant while minimizing the total amount of ton-miles of transport in acquiring each of the needed raw materials as well as transporting the finished product to the markets. The first constraint requires that for each raw material, one of its sources is assigned to supply the production plant. The second constraint lists the binary restrictions on the raw material assignment variables. This problem is a discrete nonlinear programming problem. The objective function is not nicely convex like Model 1. Even though this problem was suggested as an issue by Weber in 1909 and discussed by Isard (1956), it appears to have been virtually ignored from an algorithmic perspective. It should also be mentioned that this problem is related to general production-location problems, although we are not attempting to substitute various raw materials, as compared to selecting the sources of specific raw materials (see Hurter and Martinich 2012).

2.4.2 *Weber Model 3: Multi-Plant Location Problem*

To develop a more general framework, one can combine the notion of alternate sources and multiple facilities into a single model construct. This is certainly within the confines of Weber's original work as he addressed the issue of multiple facilities, when stating that the "locational figures will always be individual or specific for a particular plant. These weight figures are general, applying to all plants of the same kind of production." Essentially, his notion was that each production plant would have sources of raw materials as well as a set of markets to serve, where each plant can be represented by what he called a locational figure (like that given in Fig. 2.1). Other plants of the same type would each have a locational figure, so he was in a way arguing that each plant should be optimally placed within its market and sourced raw materials. Just how each set of service areas are defined was not raised by Weber in his book, but then again the techniques of resource allocation within a mathematical and economic framework had not been formally proposed or formulated until the works of Koopmans (1949) and Kantorovich (1939), work for which they jointly received the Nobel Prize in Economics in 1975. There are two ways in which we might create a multi-plant location model with Weber's definition. The first approach would be based upon adding terms in the objective that involve plant construction and operations where the number of facilities that would be involved would be endogenously determined by minimizing the combined costs of transportation, production, and plant construction. The second approach would involve an imposed constraint on the number of plants that are to be located. This second approach was used by Cooper (1964) in defining a location-allocation model that involved minimizing weighted distances of facilities serving demand. We will take this second approach here, although if Weber's work had followed Koopmans and Kantorovich, there is an equally likely chance the former approach would have been taken over the latter. Consider then:

p is the number of factories that are to be located
k is an index for facilities, where $k = 1, 2, \ldots, p$
(x_k, y_k) symbolize the coordinates of the $k -$ th factory, where $k = 1, 2, \ldots, p$

$$t_{ik}^{\ell} = \left(\left(x_k - x_i^{\ell}\right)^2 + \left(y_k - y_i^{\ell}\right)^2\right)^{\frac{1}{2}}$$

$$d_{kj} = \left(\left(x_k - x_j\right)^2 + \left(y_k - y_j\right)^2\right)^{\frac{1}{2}}$$

$$s_{ik}^{\ell} = \begin{cases} 1, \text{if source } \ell \text{ for raw material } i \text{ is used to supply factory } k \\ 0, \text{otherwise} \end{cases}$$

$$r_{kj} = \begin{cases} 1, \text{if demand at } j \text{ is served by facility } k \\ 0, \text{otherwise} \end{cases}$$

Using this notation, we can now define a model, which involves the locating p factories, allocating raw materials to factories, and distributing product to markets while minimizing the transport of raw materials and minimizing the distribution of product to markets.

$$\text{Min } Z = \sum_{k=1}^{p} \sum_{i=1}^{m} \sum_{\ell=1}^{\Theta_i} wrm_i t_{ik}^{\ell} \left(\sum_{j=1}^{n} a_j r_{kj}\right) s_{ik}^{\ell} + \sum_{k=1}^{p} \sum_{j=1}^{n} wfp\ a_j\ d_{kj}\ r_{kj} \quad (2.6)$$

subject to:

$$\sum_{k=1}^{p} r_{kj} = 1 \quad \text{for each market } j \quad (2.7)$$

$$\sum_{\ell=1}^{\Theta_i} s_{ik}^{\ell} \quad = 1 \text{ for each raw material } i \text{ and each factory } k \tag{2.8}$$

$$r_{kj} \in \{0, 1\} \quad \text{for each factory } k \text{ and each market } j \tag{2.9}$$

$$s_{ik}^{\ell} \in \{0, 1\} \quad \text{for each source } \ell \text{ of raw material type } i \text{ and factory } k \tag{2.10}$$

The objective (2.6) seeks to minimize the total haulage of materials to factories (first term) and the distribution of product to markets (second term). Note that the amount of raw material of a given type that is transported to a specific factory is a function of the need for that raw material based upon the amount of product that is distributed from that factory to markets. This is reflected by the amount $\sum_{j=1}^{n} a_j r_{kj}$ in the first term of the objective, which represents that amount of product produced and shipped from factory k. Constraints (2.7) ensure that each market is served by a factory, and constraints (2.8) ensure that each factory has been assigned a source of each raw material. Constraints (2.9) and (2.10) list the needed binary restrictions on the decision variables. Note that this model is based upon not just the allocations of the raw materials and the distributions of product but also the locations of the factories, (x_k, y_k). Note, too, all terms are nonlinear due to the embedded distance functions, but they also all involve the use of integer variables, where some terms even involve quadratic terms of integer variables. Altogether, this problem presents significant challenges to solve to optimality. Some might call this a "wicked" problem.

If we eliminated the terms of the objective that involve raw material allocation and transportation along with constraints (2.8) and (2.10), the problem would involve what many would think of as a multi-facility Weber location problem, like that solved by Cooper (1964). But in an attempt to be as true to the original form as possible, such terms should be included. Consequently, those who have worked on the problem detailed by Cooper (1964) should now recognize that they worked on a special case of Weber's general problem. Weber from the outset stated that production includes both the transport of raw materials as well as the distribution of product. Virtually all locational figures in Weber's book involve both raw materials and distribution of product.[6]

[6]Note that depending upon the location of the plant, some of these transport terms can be zero.

2.4.3 Weber Model 4: Multi-Plant Location with Resource Constraints

In describing raw material sources, Weber stated that: "it can and will happen that the normal output of the most favorable" raw material source[7] "may not be sufficient to supply the demand of the place of consumption." "In that case less favorable.....material deposits will be brought into play."[8] My assessment here is that Weber was clearly stating that one must not only keep track of which local deposits of raw materials will be used and transported, but that all of the needs for that material must be met, even if it requires the use of raw material sources that involve higher transport costs. Although we can introduce constraints on raw materials by extending Model 2 that involves the location of one factory, I have chosen to introduce resource constraints into the multi-plant Model 3, as it seems logical that resource constraints play a larger role when there are several plants being positioned. Consider then the following additional or modified notation:

C_i^ℓ is the capacity of the ℓ - th source of raw material i

s_{ik}^ℓ signifies the fraction of demand for raw material i at factory k that is supplied from the ℓ - th location of that resource

r_{kj} denotes the fraction of demand at market j that is met by facility k

$$\text{Min } Z = \sum_{k=1}^{p} \sum_{i=1}^{m} \sum_{\ell=1}^{\Theta_i} wrm_i \, t_{ik}^\ell \left(\sum_{j=1}^{n} a_j r_{kj} \right) s_{ik}^\ell + \sum_{k=1}^{p} \sum_{j=1}^{n} wfp \, a_j \, d_{kj} \, r_{kj} \quad (2.11)$$

Subject to:

$$\sum_{k=1}^{p} r_{kj} = 1 \quad \text{for each } j \quad (2.12)$$

$$\sum_{\ell=1}^{\Theta_i} s_{ik}^\ell = 1 \quad \text{for each raw material } i \text{ and each factory } k \quad (2.13)$$

[7] I have chosen to use the word "source" here while Weber was referring to the locational figure that contained the best, or least transport distant source.

[8] Again, my editing has eliminated the use of the terminology of "less favorable locational figure" to emphasize the fact that alternate, more costly sources may come into play when a source cannot handle unlimited levels of demand.

$$\sum_{k=1}^{p}\sum_{j=1}^{n} wrm_i\; a_j\; r_{kj}\; s_{ik}^{\ell} \le C_i^{\ell} \quad \text{for each material } i \text{ and source } \ell = 1, 2, \ldots \Theta_i \tag{2.14}$$

$$1 \ge r_{kj} \ge 0 \quad \text{for each factory } k \text{ and each market } j \tag{2.15}$$

$$1 \ge s_{ik}^{\ell} \ge 0 \quad \text{for each source } \ell \text{ of raw material type } i \text{ and each factory } k \tag{2.16}$$

The objective function of Model 4 is essentially the same as that of Model 3, except here some raw materials may not be sourced from only one location. This is reflected in constraints (2.15) and (2.16) as the integer restrictions on the allocation/transportation variables have been lifted. Note that it is possible that some demands may not be entirely served by one factory, as resource constraints may dictate costs that would be favorable to using another factory to meet some of the demand for product at a given market. This would be especially true if there were upper limits on the capacity of a given factory (not added here). Note, the addition of constraint (2.14) which ensures that each raw material source supplies no more than what is available at that source. This model is a nonlinear programming problem.

2.5 Multiple Stages of Production

Perhaps one of the more intriguing questions raised by Weber was the notion that production of an item may not all take place at the same location. In Chap. 6, he wrote:

> Let us suppose that an industry is influenced only by the cost of transportation, and let us neglect all of the deviating influence of labor and agglomeration. What, given such assumptions, does it mean that the production process does need to be entirely performed at one location, but split into a number of parts which may be completed at different locations? The only cause which could lead to a resultant transfer of the parts to different locations would obviously be that some ton-miles would be saved in the process . . . .

He explained this further with an example:

> "Let us take a simple case, an enterprise with three raw materials" . . . and an enterprise "which is capable of being split, technologically into two stages. In the first stage two materials are combined into half-finished product; in the second stage the half finished product is combined with the third material into the final product."

When production of an item is somewhat complex, there can be many different stages and plants involved in the production of an item. A cell phone, for example, comprises many parts, many produced at different locations. Coffee goes through a much simpler but staged production process. After it is harvested, the "coffee cherries" are processed. There are several stages to this as it involves both drying and

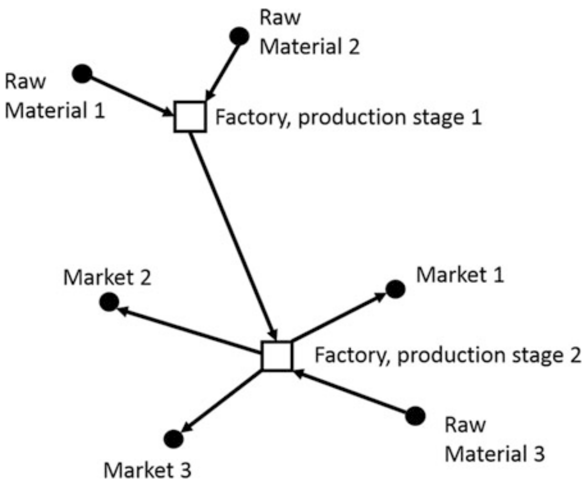


Fig. 2.3 Depiction of two stages of production, each taking place at a different location. Note for simplicity of presentation that multiple sources of a given raw material are not depicted as well as multiple plants at a given stage of production

removing the outer layers of the cherry which leaves the inner green bean. Through the initial stage of processing, what is left weighs only 20% of the original weight of the picked cherries. This initial stage of production takes place in the coffee-growing region. After it is processed, the green beans are shipped to a market region, where it is roasted and packaged, as roasted coffee does not have the same shelf life and different market areas have different preferences for roasting. Figure 2.3 depicts a simple two-stage production problem, where there are two raw materials involved in the first stage of production, one additional raw material in the second stage of production, and the distribution of the final product to three markets.

Rather than consider extending the multi-plant location Model 4 with the nuance of a staged production process, let us return to the initial Weber Model 1 and consider adding a second stage of production where there is one plant being located for the first stage and one plant being located for the second stage. To do this, consider the following additional/modified notation beyond what was used in Model 1:

(x^q, y^q) is the location of production plant for stage q, where $q = 1, 2$
Ω_q :– $\{i$: raw material i is required in stage q of production$\}$
$t_{iq} = ((x^q - x_i)^2 + (y^q - y_i)^2)^{½}$ is the Euclidean distance between raw material i and production stage q
$d_j = ((x^q - x_j)^2 + (y^q - y_j)^2)^{½}$ is the Euclidean distance between the production stage q facility and market location j
$e_{q,\,q+1} = ((x^{q+1} - x^q)^2 + (y^{q+1} - y^q)^2)^{½}$ denotes the distance between stage q facility and stage $q + 1$ facility
$e_{q,\,q+1} = ((x^{q+1} - x^q)^2 + (y^{q+1} - y^q)^2)^{½}$, and *wip* symbolizes the weight of the first-stage product.

With this notation, we can now formulate a simple two-stage production problem, where one factory is located in each of the two stages as:

$$\text{Min } Z = \sum_{q=1}^{2} \sum_{i \in \Omega_q} \sum_{j=1}^{n} wrm_i \; a_j \; t_{iq} + wip \left(\sum_{j=1}^{n} a_j \right) e_{1,2} + \sum_{j=1}^{n} a_j d_j \qquad (2.17)$$

The above formulation appears deceptively simple in that only two facilities are being located. A first-staged production facility makes an intermediate product from a set of raw materials that are listed in set Ω_1. The first-stage factory is located at (x^1, y^1). The second stage of production involves transporting the output from the first stage plus transporting an additional set of raw materials listed in the set Ω_2. The second-stage plant is located at (x^2, y^2). The variable $e_{1,2}$ represents the transport distance of the intermediate product from the stage 1 factory to the stage 2 factory. The first term of the objective sums up all of the weighted distances associated with transporting raw materials to the factories, the second term represents the weighted distance involved in transporting the intermediate product between the two stages of manufacturing, and the third term of the objective sums up all of the distribution costs (weighted distances) of shipping product to market.

This formulation is relatively simple and the notions of multiple facilities at each stage and alternate sources of raw materials of a given type can easily be added. To do this would require the introduction of resource assignment variables, distribution flow variables between stages, and finally the assignment of demand to various final-stage production sites. It is interesting to note that the one facility per stage problem (like the two-stage problem formulated above) with no alternate resource sites has fixed assignments between all points: raw material sites, staged production sites, and demand. Thus, this problem is related to the work of Miehle (1958) which is of great interest in the industrial engineering literature. It would appear that, the first description of this type of problem is, in fact, due to Weber. This simple model is a nonlinear optimization problem as well. Unfortunately, the techniques that have been developed to solve Weber Model 1 (e.g., Weiszfeld (1937)) do not directly apply to this problem.

2.6 Summary and Conclusions

Weber was constrained by the technology and mathematical reasoning of the time in describing his theory of the location of industries. He developed and used a "location triangle" to describe the location of an industry that required two localized raw materials and supplied one market. The ultimate location was the position which minimizes transport costs, the same structure as that described by Launhardt nearly 30 years earlier. Launhardt defined an approach to solve his location triangle problem, whereas Weber left this to a mathematician, Georg Pick, to provide details on how that simple problem could be solved in an appendix to his book. Weber concentrated on the details and properties that may give an "orientation" for a given production plant (e.g., oriented toward a market and oriented toward raw materials), the constraints that might involve labor availability, and the tendencies toward agglomeration of multiple industries. He was also interested in finding the least cost *feasible* site and suggested the drafting of a contour map of isodapanes (lines of equal transport cost), so that one could identify the feasible site on the lowest cost isodapane. This is an approach that was heavily used by Isard (1956) many years later in describing the location of industries.

A thorough review of Weber's book reveals detailed elements of locating a facility using the so-called location triangle whenever possible (actually Weber preferred to call it a locational figure), but often those descriptions detail a larger more complex problem. Through the lens of Ackoff, his descriptions can be thought of as the first phase of OR as a science. In this paper, we have attempted to take these descriptions and provide associated model formulations. These models represent the second phase in OR as a science. More importantly, these models represent the deep understanding and rich theory that Weber developed which has been overlooked for more than a century. Clearly, his work is more than a location triangle, but a location modeling paradigm.

Virtually all work in the last 50 years that have dealt with a "Weber" problem have assumed that there were only demands or markets, not material flow from raw material sources to production facilities, and finished product transport to markets. The notion that raw materials can be varied and deposits and sources may be limited gave rise to a rich modeling framework, even when only one facility is located. Even though the work of Cooper (1963) was groundbreaking in locating multiple facilities on the plane to serve a set of demands, this problem can be viewed as a special case of the multi-facility Weber Problem (Model 3 in this paper) that Weber described more than 50 years earlier. Further, Weber clearly realized that the production of a product may involve stages, where one plant produces parts and that another plant uses those parts and produces a finish product (and even perhaps requiring further and different resource materials as well as extended stages of production). Two stages or even multiple stages in production are common in today's industry. Weber was the first to describe this type of problem as a foundational location problem.

Finally, several of the models formulated here have not appeared in the literature or if they have in some form they have not been attributed to Weber. It is time to not only set the record straight, but to move onto Phase 3 of Ackoff's "OR as a Science" and begin the process of developing algorithms for his resource constrained and multi-staged location problems on a plane. It should also be pointed out that there are important elements in an industrial setting that Weber ignored; these include pricing and competition. Weber assumed some level of demand and ignored that demand is a function of price. Further, Weber ignored competition and even the notion of a market threshold. That is, from an economics perspective, there is a lot to be desired in Weber's theory. But from an industrial engineering, business management, and geographical perspective, one might ask where a factory should be built and what would it cost to supply a given amount of product to one or more markets? These are the very problems posed by Weber.

Acknowledgements I would like to thank Ross Gerrard for his thoughtful and constructive comments on an extremely early version of this manuscript. I would also like to thank Prof. Barry

Lentnek of the University of Buffalo for several insights that he provided in the early stages of this work a number of years ago. Finally, all errors and omissions are mine alone.

References

Ackoff, R. L. (1956). The development of operations research as a science. *Operations Research, 4*(3), 265–295.

Balinski, M. L. (1965). Integer programming: Methods, uses, computations. *Management Science, 12*(3), 253–313.

Beckmann, M. (1955). An activity analysis approach to location theory. *Kyklos, 8*(2), 125–143.

Beckmann, M. J. (1968). *Location theory*. New York, NY: Random House.

Brimberg, J., & Love, R. F. (1993). Global convergence of a generalized iterative procedure for the minisum location problem with lp distances. *Operations Research, 41*(6), 1153–1163.

Brimberg, J., Hansen, P., Mladenović, N., & Taillard, E. D. (2000). Improvements and comparison of heuristics for solving the uncapacitated multisource Weber problem. *Operations Research, 48*(3), 444–460.

Butt, S. E., & Cavalier, T. M. (1996). An efficient algorithm for facility location in the presence of forbidden regions. *European Journal of Operational Research, 90*(1), 56–70.

Christaller, W. (1933). *Die zentralen Orte in Süddeutschland: eine ökonomisch-geographische Untersuchung über die Gesetzmässigkeit der Verbreitung und Entwicklung der Siedlungen mit städtischen Funktionen*. Frankfurter Geographische Hefte 24. Jahrgang, Heft 1.

Cooper, L. (1963). Location-allocation problems. *Operations Research, 11*(3), 331–343.

Cooper, L. (1964). Heuristic methods for location-allocation problems. *SIAM Review, 6*(1), 37–53.

Drezner, Z., & Suzuki, A. (2004). The big triangle small triangle method for the solution of nonconvex facility location problems. *Operations Research, 52*(1), 128–135.

Drezner, Z., & Wesolowsky, G. O. (1989). The asymmetric distance location problem. *Transportation Science, 23*(3), 201–207.

Drezner, Z., & Wesolowsky, G. O. (1991). The Weber problem on the plane with some negative weights. *INFOR: Information Systems and Operational Research, 29*(2), 87–99.

Eckhardt, U. (2008). *Kürzeste Wege und optimale Standorte – von Industriestandorten, Bomben und Seifenblasen*. Retrieved 11 January, 2019, from https://www.math.uni-hamburg.de/home/eckhardt/Standort.pdf

Greulich, F. E. (1999). The barycentric coordinates solution to the optimal road junction problem. *Journal of Forest Engineering, 10*(1), 111–114.

Hakimi, S. L. (1964). Optimum locations of switching centers and the absolute centers and medians of a graph. *Operations Research, 12*(3), 450–459.

Hakimi, S. L. (1965). Optimum distribution of switching centers in a communication network and some related graph theoretic problems. *Operations Research, 13*(3), 462–475.

Hoover, E. M. (1937). *Location theory and the shoe leather industries* (Vol. 55). Cambridge: Harvard University Press.

Hotelling, H. (1990). Stability in competition. In *The collected economics articles of Harold Hotelling*. New York, NY: Springer.

Hurter, A. P., & Martinich, J. S. (2012). *Facility location and the theory of production*. Amsterdam: Springer Science & Business Media.

Isard, W. (1949). The general theory of location and space-economy. *The Quarterly Journal of Economics, 63*(4), 476–506.

Isard, W. (1956). *Location and space-economy*. New York: John Wiley and Sons.

Isard, W., & Reiner, T. A. (1962). Regional science and planning. *Papers in Regional Science, 8*(1), 1–35.

Kantorovich, L. V. (1939). Mathematical methods of organizing and planning production. *Management Science, 6*, 366–422.

Katz, I. N., & Cooper, L. (1981). Facility location in the presence of forbidden regions, I: Formulation and the case of Euclidean distance with one forbidden circle. *European Journal of Operational Research, 6*(2), 166–173.

Klamroth, K. (2001). Planar Weber location problems with line barriers. *Optimization, 49*(5-6), 517–527.

Koopmans, T. C. (1949). Optimum utilization of the transportation system. *Econometrica, 17*, 136–146.

Koopmans, T. C. (1951a). *Analysis of production as an efficient combination of activities. Analysis of Production and Allocation* (pp. 33–97).

Koopmans, T. C. (1951b). An analysis of production as an efficient combination of activities. In T. C. Koopmans (Ed.), *Activity analysis of production and allocation, proceeding of a conference* (pp. 33–97). London: John Wiley and Sons Inc..

Koopmans, T. C., & Beckmann, M. (1957). Assignment Problems and the location of economic activities. *Econometrica, 25*(1), 53–76.

Krarup, J., & Vajda, S. (1997). On Torricelli's geometrical solution to a problem of Fermat. *IMA Journal of Management Mathematics, 8*(3), 215–224.

Kuhn, H. W., & Kuenne, R. E. (1962). An efficient algorithm for the numerical solution to the generalized Weber problem in spatial economics. *Journal of Regional Science, 4*(2), 21–33.

Laporte, G., Nickel, S., & da Gama, F. S. (2015). Introduction to location science. In *Location science*. Cham: Springer.

Launhardt, W. (1872). *Kommercielle Tracirung der Verkehrswege*. Architekten- und Ingenieurverein.

Lösch, A. (1940). *The economics of location*. New Haven: Yale.

Miehle, W. (1958). Link-length minimization in networks. *Operations Research, 6*(2), 232–243.

Moses, L. N. (1958). Location and the theory of production. *The Quarterly Journal of Economics, 72*(2), 259–272.

Perreur, J. (1998). Industrial location theory in German thought–Launhardt and Weber. *Recherches Économiques de Louvain/Louvain Economic Review, 64*(1), 75–96.

Pinto, J. V. (1977). Launhardt and location theory: Rediscovery of a neglected book. *Journal of Regional Science, 17*(1), 17–29.

Rosing, K. E. (1992). An optimal method for solving the (generalized) multi-Weber problem. *European Journal of Operational Research, 58*(3), 414–426.

Salhi, S., & Gamal, M. D. H. (2003). A genetic algorithm based approach for the uncapacitated continuous location–allocation problem. *Annals of Operations Research, 123*(1-4), 203–222.

Simpson, T. (1750). *The doctrine and application of Fluxions*. London: J. Nourse.

Varignon, P. (1687). *Projet d'une nouvelle mechanique*. Paris: Chez C. Jombert.

Vergin, R. C., & Rogers, J. D. (1967). An algorithm and computational procedure for locating economic facilities. *Management Science, 13*(6), B–240.

von Thünen, J. H. (1826). *Der isolierte Staat in Beziehung auf Landwirtschaft und Nationaloekonomie*. Germany: Jena.

Weber, A. (1909). *Über den Standort der Industrien (Theory of the location of industries)*. Chicago: University of Chicago Press.

Weber, A. (1929). *Theory of the location of industries* (translated with an introduction and notes by C. J. Friedrich). Chicago: University of Chicago Press.

Weiszfeld, E. (1937). Sur le point par lequel le somme des distances de n points donnes est minimum. *Tohoku Mathematics Journal, 43*, 355–386.

Wersan, S. J. J. E. A., Quon, J., & Charnes, A. (1962). *Systems analysis of refuse collection and disposal practices*. Kansas: American Public Works Association Yearbook, American Public Works Association.

Wesolowsky, G. O. (1993). The Weber problem: History and perspectives. *Location Science, 1*(1), 5–24.

Chapter 3
A General Framework for Local Search Applied to the Continuous p-Median Problem

J. Brimberg and S. Salhi

3.1 Introduction

The basic aim of continuous location problems is to generate facility sites in a given continuous space, usually the Euclidean plane, in order to optimize some performance measure relative to a set of given points (customers). From a mathematical perspective, we may attribute the start of continuous location theory to Fermat who suggested the problem of locating a point in the plane that minimizes the sum of distances to three given points. The idea was generalized in an applied sense by Weber (1909) a few centuries later by extending the problem to any number "n" of given (or fixed) points representing markets and associating weights (or demands) to these points. The objective function, a weighted sum of distances from the facility to the given markets, now measured the cost of delivering goods to the markets as a function of the facility location. This function is known to be convex for any distance norm and hence amenable to solution by local descent methods. One such method developed for Euclidean distance, the well-known single-point iterative scheme by Weiszfeld (1936), has received much attention in the literature, including for example, seminal papers by Kuhn (1973) and Katz (1974), which studied the global and local convergence properties of this method. For further

J. Brimberg (✉)
Department of Mathematics and Computer Science, The Royal Military College of Canada, Kingston, ON, Canada
e-mail: jack.brimberg@rmc.ca

S. Salhi
Centre for Logistics and Heuristic Optimisation (CLHO), Kent Business School, Sibson Building, University of Kent, Canterbury, UK
e-mail: S.Salhi@kent.ac.uk

H. A. Eiselt, V. Marianov (eds.), *Contributions to Location Analysis*, International Series in Operations Research & Management Science 281,
https://doi.org/10.1007/978-3-030-19111-5_3

reading on the rich history of the continuous single-facility minisum (or 1-median) location problem, see Wesolowsky (1993) and Drezner et al. (2002).

Cooper (1963, 1964) extended Weber's problem for locating one new facility in the plane to locating any given number, say $p \geq 1$, new facilities. This important practical extension made the problem much more difficult to solve, but at the same time, much more interesting to researchers. This new *multisource Weber problem* (also referred to as the uncapacitated continuous location-allocation problem or continuous p-median problem) may be formulated as follows:

$$\min\ F(X) = \sum_{i=1}^{n} w_i \min_{1 \leq j \leq p} \left\{ d\left(X_j, A_i\right) \right\} \tag{3.1}$$

where $X_j = (x_j, y_j)$ denotes the unknown location of new facility j, $j = 1, \ldots, p$; $X = (X_1, X_2, \ \ldots \ , X_p)$ is the vector containing the p new facility locations; $A_i = (a_i, b_i)$ is the given location of fixed point (also called demand point, existing point, customer) i, $i = 1, \ldots, n$; $w_i \geq 0$ is the given weight associated with customer i, $i = 1, \ldots, n$; and $d(X_j, A_i)$ denotes the Euclidean distance between customer-facility pair (i, j), that is,

$$d\left(X_j, A_i\right) = \sqrt{\left(x_j - a_i\right)^2 + \left(y_j - b_i\right)^2} \quad \forall\, (i, j)\,. \tag{3.2}$$

We see that the objective function, $F(X)$, is a weighted sum of distances from the fixed points to their *closest facilities*. Thus, a basic assumption of the model is that a customer is always served by the facility that is closest to it. Also note that each term on the right-hand side of (3.1) is the product of a positive weight and the minimum of p distances, in other words, the minimum of p convex functions which itself is a highly nonconvex function.

An equivalent mathematical programming formulation of the multisource Weber problem (*MWP*) is given by (e.g., see Love et al. 1988):

$$\min G\,(X, W) = \sum_{i=1}^{n} \sum_{j=1}^{p} w_{ij} d\left(X_j, A_i\right) \tag{3.3}$$

$$\text{s.t.} \ \sum_{j=1}^{p} w_{ij} = w_i, \ \ i = 1, \ldots, n \tag{3.4}$$

$$w_{ij} \geq 0, i = 1, \ldots, n, \, j = 1, \ldots, p. \tag{3.5}$$

Here w_{ij} denotes the demand (or flow) from (or to) customer i that is allocated to facility j, $W = (w_{ij})$ is the vector of flow allocation variables, and the remaining notation is the same as defined for (3.1). A key point to observe is that the minimization objective will automatically allocate flows to the nearest facilities in an optimal solution. Also note that the objective function (3.3) is now expressed as a sum of nonconvex terms of the form $u * v$ and hence is nonconvex itself.

To illustrate the complicated nature of the objective function, consider the following simplest of problems of locating two facilities on a line to serve three customers with coordinates and weights given by: $a_1 = 0, a_2 = 1, a_3 = 2; w_1 = 0.5, w_2 = 1.5, w_3 = 1.0$. Letting x_j denote the position on the line of facility $j, j = 1, 2$, the objective function in (3.1) may be written as:

$$F(x_1, x_2) = 0.5 \min(|x_1|, |x_2|) + 1.5 \min(|x_1 - 1|, |x_2 - 1|) + 1.0 \min(|x_1 - 2|, |x_2 - 2|).$$

Suppose an initial (random) solution is selected with $x_1^{(0)} = 0.5, x_2^{(0)} = 1.2$. Holding facility 1 fixed at $x_1^{(0)} = 0.5$, and observing the profile of F as x_2 varies (see Fig. 3.1), we see that a local search would move facility 2 from $x_2^{(0)} = 1.2$ to $x_2^{(1)} = 1.0$. In a similar way, with facility 2 now fixed, facility 1would then move to $x_1^{(1)} = 0$, at which time a local minimum $(x_1,x_2) = (0, 1)$ would be reached with $F(0, 1) = 1$ (see Fig. 3.2). We may easily see that the global minimum occurs at $(x_1,x_2) = (1, 2)$ (or (2,1)) with $F(1, 2) = 0.5$, thus demonstrating that even for such a trivial problem, a local search may fail to find an optimal solution. Another interesting feature to observe is the saddle point in Fig. 3.2 at $(x_1,x_2) = (1, 1)$. We will return to saddle points later in the chapter.

The non-convexity of the objective function illustrated by the simple example above suggests that a global optimization technique is needed if we wish to solve the continuous p-median problem (*MWP*) exactly. There is more bad news—Megiddo and Supowit (1984) show that *MWP* is NP-hard. Thus, there is no surprise that many heuristics have been developed to solve it. Earlier heuristics include the

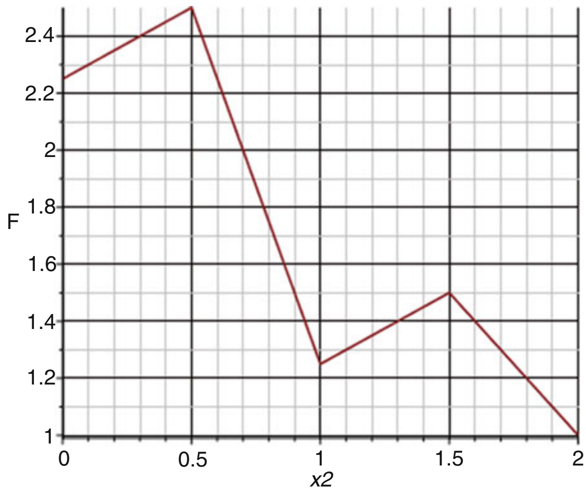


Fig. 3.1 Profile of F with x1 = 0.5

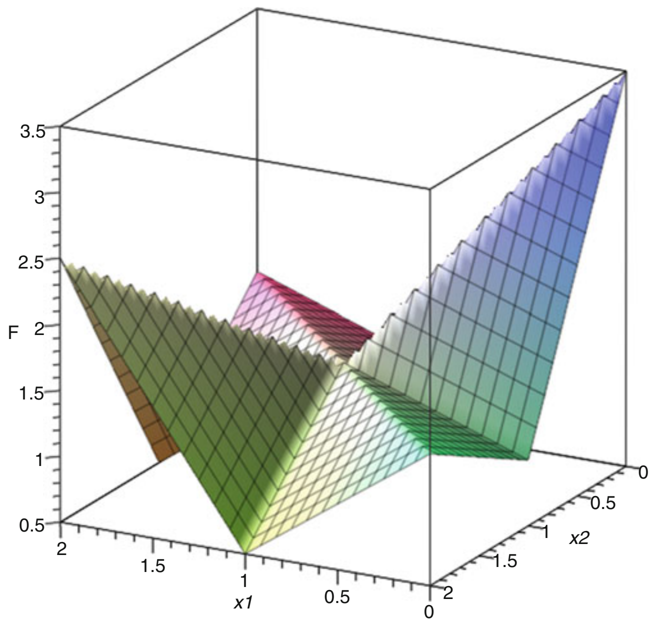


Fig. 3.2 Plot of F(x1, x2)

popular alternating locate-allocate algorithm by Cooper (1963, 1964), gradient-based methods such as Murtagh and Niwattisyawong (1982) and Chen (1983), and the projection method of Bongartz et al. (1994).These are all local search-based methods and hence terminate at a local optimum (local minimum point) in the solution space. It has been known for some time that the objective function of *MWP* may contain a large number of local minima of varying quality.

For example, in their well-known 50 customer problem, Eilon et al. (1971) were able to generate 61 local minima from 200 random restarts of Cooper's method when $p = 5$. It was proven much later (Krau 1997) that their best solution was indeed the global optimum. The fact that the worst solution deviated from the best by some 40% showed quite dramatically the danger of being "trapped" in a local optimum. The same problem is investigated by Brimberg et al. (2010) and Brimberg and Drezner (2013) where in each case 10,000 runs of a Cooper-style algorithm (*ALT*) are conducted from randomly generated initial solutions for different values of p. Some salient results from Brimberg and Drezner (2013) are given in Table 3.1. The first column of Table 3.1 gives the number p of facilities to locate; this is followed by the number of times (out of 10,000) that the known optimal solution (Krau 1997) was found, the average deviation of the found local solutions from the optimal objective value, and the maximum deviation (i.e., the worst solution). These results, which apply to relatively small instances by today's standards, demonstrate quite dramatically the tendency for the number of local minima to increase exponentially with problem size. A significant positive correlation between the qualities of the

Table 3.1 Summary results for *ALT* on $n = 50$ customer set (Brimberg and Drezner 2013)

p	# times optimum found (out of 10,000 runs)	# different local solutions	Ave % dev	Max % dev
5	674	261	4.89	51.31
10	54	4701	10.31	68.85
15	100	8727	17.95	60.38
20	0	9164	21.50	69.36
25	0	9035	27.37	81.94

initial and final solutions is also observed leading to the conclusion that one should select the best among several random solutions as a starting initial solution.

Cooper's alternating algorithm (*ALT* for short) uses a simple and elegant insight, namely, that the two phases of the problem—locating the facilities and allocating the customers to them—are easy to solve as separate sub-problems. That is, once the facility sites are fixed, each customer should be assigned to its nearest facility; and when the resulting partition of the customer set is fixed, the problem reduces to *p* convex single-facility location problems to find the corresponding best locations of the facilities. Thus we may view *MWP* in a strictly combinatorial sense where the objective is to find an optimal partition of the customer set, and where facilities are automatically located at corresponding 1-median points of their assigned customer subsets. The *ALT* heuristic became very popular with various modifications being introduced over the years (see the survey by Brimberg and Hodgson (2011) for a historical overview). The current version typically begins by randomly locating the *p* new facilities within the convex hull of the fixed points (or the smallest rectangle enclosing the points) to obtain an initial solution and allocation of customers to facilities. Also note that a random multi-start version of *ALT* is commonly used as a benchmark for comparing other heuristics (e.g., see Brimberg et al. 2000).

Every iteration of *ALT* involves the solution of up to *p* independent single-facility location problems, which is usually accomplished using the Weiszfeld procedure. Hence it is very important to make this procedure as efficient as possible. When the *p*-median problem occurs in the plane and Euclidean distances are being considered, which is the standard case, earlier studies have suggested that the pre-set step-size of the Weiszfeld procedure can be multiplied by a factor $\lambda \in [1, 2]$ to speed up convergence (e.g., see Ostresh 1978). Drezner (1992) recommends setting $\lambda = 1.8$ as a good compromise based on an empirical study, and also derives an effective formula from a local approximation of the objective function that calculates a new value of λ at each iteration. A different approach to speed up convergence is taken by Drezner (1995) where it is assumed that the differences between successive iterates form a geometric series. In most instances tested, the new procedure reduced run time by a factor of 2, and in extreme cases of slow convergence by over 100,000. Drezner (2015) develops a "fortified" Weiszfeld algorithm that uses a parabolic approximation of the objective function and tests a few demand points for optimality that lie near the sequence of iterates generated. This method is observed to work well when the number of customers assigned to a facility is small (< 50), which typically applies to test instances of the p-median problem, and is related to the probability of an optimal solution coinciding with a demand point that tends to be inversely proportional to this number (Drezner and Simchi-Levi 1992).

A new approach referred to as "reformulation local search" (*RLS*) is proposed by Brimberg et al. (2014). The main idea here is to exploit the relation between the continuous model and a discrete approximation of that model. A local search, for example using *ALT*, is carried out in the continuous space until a local optimum is reached. The local search then switches to a discrete version of the problem and tries to find an improved solution for the discrete p-median problem on an associated graph. The process of switching between the two formulations is repeated until no further improvement can be found in either space. The idea of solving a discrete p-median problem as an approximation of the continuous p-median problem, with n nodes on the graph corresponding to the set of fixed points, goes back as far as Cooper (1963). Hansen et al. (1998) solve this discrete version exactly and then apply one step of continuous adjustment (i.e., to locate each facility optimally with respect to its assigned customers from the discrete solution) to get the final solution. Excellent results are obtained even though the final solution is not guaranteed to be a local optimum in the continuous space. A nice feature of the *RLS* approach is its inherent flexibility. A choice of heuristics can be used in each phase (discrete and continuous) allowing larger instances to be handled without limitations on size imposed by exact algorithms. The discrete approximation is not restricted to the set of fixed points but could be constructed interactively by the user. More importantly, local solutions obtained in the continuous phase are added to the node set at each iteration so that in theory the discrete approximation is able to 'converge' to the continuous model.

Noting the importance of having a "good" starting solution, Brimberg and Drezner (2013) develop a composite heuristic consisting of a search on a grid representation of the planar p-median problem using simulated annealing, followed by an improved version of *ALT* (which we will discuss later) that takes the solution from the first phase as the starting point. Again we see the use of a discrete approximation of the original problem, but this time the goal is to find a good starting solution. As a result, substantial improvement over *ALT* is obtained. For example, referring to Table 3.1, and repeating 10,000 runs of the composite heuristic from random initial solutions, the number of different local optima obtained is reduced drastically to 1 (the global solution), 10, 16, 66, and 120, respectively, for $p = 5$, 10, 15, 20, and 25. The average deviation from the optimal solution's value is also drastically reduced to slightly over 1% in the worst case.

The first heuristic approach to impose a neighborhood structure on the planar p-median problem may be attributed to Love and Juel (1982). A given neighborhood is defined here as the set of solutions (or points) surrounding the current solution that are obtained by transferring a specified number of assignments of customers from their current facilities to new ones. The one- and two-exchange neighborhoods are only considered due to computational requirements. Facilities are always located optimally with respect to their assigned customers. Important to note is that the authors are able to show that a local optimum in the one-exchange neighborhood is not necessarily a local optimum in the two-exchange neighborhood. Hence, further improvement is possible by adding the larger neighborhood search. Of course, getting "trapped" in very deep troughs still poses the same problem as before. Mladenović and Brimberg (1995) apply the exchange neighborhood as follows in a fixed neighborhood (or iterated local) search. A certain number b of points is selected at random in the k-exchange neighborhood of the current solution (a local minimum point), where b and k are given parameters. Cooper's method is then applied to each of the selected points to obtain up to b new local minima. A descent version of their heuristic only allows a move to one of them if it is better than the current solution. A variable neighborhood approach where k is allowed to vary over a specified range, $1, \ldots, k_{max}$ (a parameter) is proposed by Brimberg and Mladenović (1996b), while a simple Tabu search using the one-exchange neighborhood is given by Brimberg and Mladenović (1996a). These two papers may be considered the first attempts of solving *MWP* with a metaheuristic.

The one-exchange neighborhood is used in a special way in a transfer follow-up step that is added to an *ALT* procedure by Brimberg and Drezner (2013). The transfer follow-up switches the assignment of "border" customers that are almost as near to their second-closest facility as the first, with the aim to move to a better adjacent local minimum point. More will be said about this improved alternating approach (*IALT*) in the next section.

A different neighborhood structure is applied by Brimberg et al. (2000) that is based on the relocation of facilities to unoccupied fixed points (i.e., customers that do not have facilities already coincident with them) instead of the reassignment (reallocation or exchange) of customers between facilities. Note that this is similar to the vertex exchange move originally developed by Teitz and Bart (1968) for the discrete p-median problem. A few variants of local search are proposed which examine selected points in the one-exchange relocation neighborhood followed by an *ALT* procedure from these points. Much better results are obtained compared to a standard *ALT*, and in a fraction of the time. Brimberg et al. (2000) also apply relocation-based neighborhoods within a variable neighborhood search framework. A point in the k-neighborhood of a current solution is obtained in the shaking operation by relocating k facilities randomly at unoccupied fixed points.

Several metaheuristic-based methods have been applied over the last years to solve *MWP*, including tabu, variable neighborhood, and genetic searches, as well as some hybrid schemes; see, for example, the survey by Brimberg et al. (2008). More recently, Drezner et al. (2015a) use a variable neighborhood search (*VNS*) with the following modified structure:

1. The improved *ALT* procedure (*IALT*) from Brimberg and Drezner (2013) is used in the local improvement step instead of a standard *ALT*.
2. $b \geq 1$ points in the k-neighborhood (allocation type exchange) are randomly selected one at a time in the shake operation. If a better solution is found by *IALT*, a move is made, k is reset to 1 and the counter for b is reset to 0. If none of the b local minima thus obtained is better than the current solution, k is increased by 1 (up to a limit k_{max}).

3. If m (another parameter) iterations of k from 1 to k_{max} fail to produce a better solution, the algorithm returns the current solution (which is also the best found) and stops.

Drezner et al. (2015a) also develop a new "concentric" search for solving *MWP*, which is a modification of an approach known as concentric tabu search devised by Drezner (2002, 2005) for the quadratic assignment problem. In their concentric algorithm, the current solution (also the current best as in *VNS*) is referred to as the center solution. Whereas in *VNS*, perturbations are always performed in the shake operation starting from the center solution, the concentric algorithm initiates the perturbation from the best solution in the smallest k-neighborhood of the center solution whose $flag = 0$. The *flag* of a k-neighborhood is always reset to 0 when an improved solution in that neighborhood is obtained. The algorithm stops when all such $flags = 1$. The two heuristics by Drezner et al. (2015a) performed consistently well on a wide range of test instances, with a small edge in favor of the concentric search. The study also showed that a front-end subroutine for finding "good" starting solutions improved the overall performance of the heuristics, and that a post-optimization routine that solves decomposed problems of three or four facilities obtained by Delaunay triangulation, and their assigned customers, using *RLS* (Brimberg et al. 2014) was able to improve some of the solutions.

Drezner et al. (2015b) propose a new variant of *VNS* termed distribution-based *VNS*, or *DVNS* for short, and apply it to the *MWP*. The basic idea here involves a variation of the shaking operation. Instead of systematically increasing the shake ($k = 1,..., k_{max}$) to obtain points successively further away from the current solution, the k-neighborhood in *DVNS* is selected randomly based on an empirically derived distribution that places more weight on those neighborhoods that are more likely to lead to an improved solution. A genetic algorithm is also developed and tested by Drezner et al. (2015b). A simple and effective merging process is used to generate offspring as follows: (1) two parents are selected randomly from the population; (2) an imaginary line is drawn at a random angle through the center of the cluster formed by the facilities; (3) the facilities on one side of the line that belong to one of the parents are selected and combined with those on the other side belonging to the other parent. The selection is adjusted to ensure that the new solution (the offspring) contains exactly p facilities, with about half of them coming from each parent. The two developed heuristics are also combined in a hybrid heuristic where the genetic algorithm is applied first, and the output solution is then used as the starting solution for the *DVNS* stage. All three heuristics performed very well on relatively large test instances ($n = 654$ and 1060) taken from the literature, with the hybrid heuristic performing the best of all. Also noteworthy is the fact that *DVNS* was comparable to standard (or Basic) *VNS* (*BVNS*) in terms of solution quality, but considerably faster. Both these heuristics achieved the best-known solutions about 90% of the time, and were able to find new best solutions for two test instances ($(n,p) = (654, 95)$ and (1060, 150)).

Drezner and Salhi (2017) incorporated two simple but effective neighborhood reduction schemes within a local search to speed up the search without significantly affecting the solution quality. The idea is to identify non-promising moves and hence avoid carrying out unnecessary computations. The first rule is based on the convex hull where a threshold is determined for each customer, whereas the second rule focusses on the borderline customers (i.e., those that lie between their first- and second-closest facilities). This implementation is tested on three problems with $n = 654$, 1060, and 3038 from the *TSP* library (Reinelt 1991) with very good results, such as problem instances with some 90% saving in computation time without loss in solution quality. This fast local search is then embedded into a metaheuristic based on *GA*/*VNS* where new best results are obtained. For more information on neighborhood reduction, see Salhi (2017).

The remainder of the chapter is organized as follows. Section 3.2 reviews a standard implementation of Cooper's (1963, 1964) famous alternating algorithm (*ALT*). It should be noted that there are many variants of Cooper's algorithm, and that this particular approach was incorporated later in a random multi-start implementation of *ALT* which remained state-of-the-art for a number of years. Some interesting extensions of *ALT* presented recently by Drezner et al. (2016) are also discussed. Section 3.3 proposes a general framework for alternating locate-allocate heuristics that encompasses these recent extensions and also allows us to construct new *ALT*-based heuristics. Some preliminary computational results are given in Sect. 3.4, followed in Sect. 3.5 by some discussion related to degenerate solutions and saddle points. Section 3.6 contains a brief conclusion to the chapter.

3.2 Cooper's Algorithm and Recent Extensions

Cooper (1963, 1964) is credited with the first formulation of the *MWP*. Recognizing the non-convexity of the objective function, Cooper proposed several heuristics to solve the problem. The best-known of these is an alternating locate-allocate heuristic now referred to as Cooper's algorithm. The general idea is as follows. Starting with p given facility locations, we construct a partition of the customer set into p mutually exclusive subsets by allocating each customer to its nearest facility. We then solve the resulting p independent 1-median problems. Customers are then reallocated to their closest facilities, and the process is repeated until a local minimum point is reached (i.e., the partition remains the same). The basic steps of the procedure which we call *ALT* (for *alternating*) may be summarized as follows.

Cooper's Algorithm (ALT) *Step 1* (*Initial solution*): Locate p facilities at random and assign each customer to its nearest facility (with any ties broken arbitrarily).

Step 2: (*Location phase*) Keeping the customer assignments fixed, solve p independent single-facility location problems.

Step 3: (*Allocation phase*) Re-assign customers to their nearest facilities. If there are no changes, STOP (the current solution is a local minimum); else return to step 2.

The Cooper algorithm typically reaches a local minimum after only a small number of iterations. Also, as observed above, the quality of the obtained solution can vary a lot. Brimberg and Drezner (2013) modify *ALT* by selecting one facility at random at a time in the location phase (step 2). After re-locating the selected facility optimally with respect to its assigned customer subset, the allocation phase (step 3) is carried out over the entire customer set. A *flag* is used for each facility to denote whether or not the facility is optimally located with respect to its currently assigned customers. Note that this process results in a more gradual descent than the standard *ALT*, and may lead to a different output solution when starting from the same initial solution. Brimberg and Drezner (2013) also add a transfer follow-up procedure once a local minimum is attained. This procedure re-assigns a customer to an adjacent (second-closest) facility one at a time, and then relocates the two affected facilities to their new median points. If the resulting solution is better, the heuristic (which they call *IALT* (for *improved alternating*)) resumes its modified *ALT* procedure to reach a new improved local solution, which is then followed by a new round of transfer follow-up. The customers selected for transfer are always those with the smallest difference in distance to their closest and second-closest facilities. If a better solution cannot be found within a specified maximum number of transfers, the heuristic stops. It is important to note that a local solution obtained by *IALT* is always a local solution for *ALT* but not vice versa. This may be one of the reasons why the average quality of local minima obtained by *IALT* is observed to be better than with *ALT*. Another contributing factor may be the slower convergence rate of the main loop due to the selection of one facility at a time for re-location.

Recently, Drezner et al. (2016) proposed several new local searches for solving the *MWP* in the plane. One of these constructs an initial solution using a greedy-add heuristic with a random component. The *IALT* algorithm is then applied on the initial solution. A multi-start version of this hybrid procedure (termed *START*) is observed to obtain significantly better solutions with much smaller variation in quality than *IALT* with random starting solutions. We discuss two more heuristics from Drezner et al. (2016) that can be easily modified to fit the general framework for *ALT*-type heuristics presented in the next section.

3.2.1 A Decomposition Algorithm (DECOMP)

Step 1: Find an initial solution using a suitable heuristic (or random selection) to locate p facilities. Let X_C denote this solution.

Step 2: Perform a Delaunay triangulation on the set of p facilities to obtain a list of triangles and associated quadrangles called *LIST*.

Step 3: If LIST is empty, stop and return X_C; *else*:

(a) Select a polygon at random from *LIST* (triangles first followed by quadrangles).
(b) Determine the set U of all demand points served by the facilities at the vertices of the selected polygon.
(c) Apply a suitable heuristic on the set U getting locations for $q = 3$ or 4 facilities.

(d) *If* the new solution for the q facilities is better than the original configuration serving U,

   i. Update the locations of the q facilities,
   ii. Apply *IALT* (or *ALT*) on the complete set of p facilities, update X_C, and
   iii. Return to step 2;

else remove the polygon from *LIST* and return to the top of step 3.

The use of a "suitable heuristic" in steps 1 and 3(c) allows a degree of flexibility in building different variants of *DECOMP*. For example, in one application of this algorithm, Drezner et al. (2016) use *START* in both these steps. Thus, the heuristic solves the sub-problem in step 3(c) from scratch; i.e., the current locations of the q facilities are not used for the initial solution.

If we select a random initial solution in step 1, use *ALT* on the sub-problem as the suitable heuristic in step 3(c), and delete step 3(d)(ii), i.e., return directly to step 2 after updating the locations of the q facilities in current solution X_C, we end up with a variant of *ALT* that includes decomposition which will fit nicely in the framework we discuss in the next section.

3.2.2 An Improvement Algorithm (IMP)

Another novel local search by Drezner et al. (2016) selects one facility (j) at a time to relocate as in *IALT*. Except now instead of fixing the customer allocation and solving the resulting 1-median problem, this new method known as *IMP* locates the selected facility j optimally while keeping the remaining $p-1$ facilities fixed at their current locations; that is, the following problem is solved:

$$\min \sum_{i=1}^{n} w_i \min \left\{ d\left(X_j, A_i\right), D_i \right\}, \tag{3.6}$$

where $D_i = \min_{k \neq j} \{d(X_k, A_i)\}$ is the distance from customer i at A_i to its closest facility other than facility j. This problem is equivalent to the limited distance model introduced by Drezner et al. (1991), and is solved here exactly using a branch-and-bound algorithm known as the Big Square Small Square method (e.g., see Plastria (1992)). The steps of *IMP* are outlined below.

Step 1: Obtain an initial solution.

Step 2: Repeat the following for each facility j selected one at a time in random order:

(a) Calculate D_i, $i = 1, \ldots, n$
(b) Relocate facility j by solving the limited distance problem (3.6).

Step 3: If the location of at least one facility has changed, repeat step 2; *else* return the current solution and *stop*.

Note that *IMP* in effect reallocates customers to facilities simultaneously as selected facility j is being moved to an optimal location with all other facilities fixed. In the general framework below, we can adapt this idea at least in an approximate sense by selecting the same facility (or facilities) in consecutive re-locations until the assigned customer subset stops changing. Also note that the computational results obtained with *IMP* by Drezner et al. (2016) are on average considerably better than those obtained by *ALT* and *IALT*.

3.3 A General Framework for Location-Allocation Heuristics

We have seen that a key feature of Cooper's algorithm and its variants is that whenever the process of moving a facility or subset of facilities from the current location(s) is taking place to improve the solution, the assigned subset of customers to that facility or subset is fixed and unchangeable. What we want to do is propose a general procedure that retains this fundamental property while unifying the variants or extensions of Cooper's algorithm, and that also allows us to construct new location-allocation heuristics.

In the general framework presented next, we consider a list of sub-problems denoted by $\{P_1, P_2, \cdots, P_K\}$, where K is a parameter, such that each sub-problem P_t is fully specified by a given subset of facilities and their current locations, and a subset of assigned customers for which these facilities are the closest (with ties broken arbitrarily), and each facility is included in at least one sub-problem.

3.3.1 General Location-Allocation Local Search (G-ALT)

Step 1: Construct an initial solution (e.g., select p random facility locations in the convex hull of the fixed points).

Step 2: Construct a list of sub-problems $P = \{P_1, P_2, \cdots, P_K\}$.

Step 3: Remove one of the sub-problems P_t from P according to a given rule, and solve it approximately with *ALT* using the current configuration as the starting solution.

Step 4: If the solution to P_t is improved, move the associated facilities to their new locations and return to step 2; *else* the current configuration remains unchanged and $P \leftarrow P \backslash \{P_t\}$.

Step 5: If P is not empty, return to step 3; *else stop*.

The basic Cooper algorithm (*ALT*) may now be viewed as the simplest application of *G-ALT* where a single sub-problem containing all p facilities is defined in step 2. For *IALT* (without transfer follow-up) each sub-problem P_t defines a 1-median problem, and the selection rule in step 3 randomly draws the next sub-problem. Meanwhile, the modified *DECOMP* discussed in the preceding section works with a list of sub-problems containing three or four adjacent facilities obtained by Delaunay triangulation with their allocated customers.

We see that for any variant of *G-ALT*, each constructed sub-problem translates to a location problem in a subspace defined by a selected subset of facilities and a fixed subset of customers allocated to them. In the fundamental result below, we assume that there are no ties in the output solution returned by *G-ALT* (i.e., no cases where a demand point A_i has a choice between two or more facilities with the same smallest distance to it and hence could be assigned to any one of them). Such solutions may be associated with saddle points similarly as in Fig. 3.2 at $(x_1, x_2) = (1, 1)$. That is, a re-assigning of A_i to an alternate tied facility could lead to further descent and a better solution. This unlikely situation can be handled by examining all such possible re-assignments before a final solution is declared. Thus we may assume that the terminal (or output) solution from *G-ALT* is always a local minimum point.

Property 3.1 Let X_* be a terminal solution for some variant of *G-ALT*. Then X_* is also a terminal solution for any other variant of *G-ALT*.

Proof Consider two versions, say, $G\text{-}ALT_1$ and $G\text{-}ALT_2$, and let X_* be a terminal solution of $G\text{-}ALT_1$. Now apply X_* as the initial solution to $G\text{-}ALT_2$. Since X_* is a local minimum point, it follows that each facility location must be a 1-median point for any partition of the customer set obeying the 'closest facility' rule. Hence applying *ALT* to any sub-problem P_t in $G\text{-}ALT_2$ will not improve the current configuration in P_t. We delete P_t from the list and continue in similar fashion until the list is empty leaving X_* as the final solution.

One might suspect as a result of Property 3.1 that all variants of *G-ALT* should behave in the same way. We will show briefly in the next section that this is not true; that is, different descent paths can affect the average quality of solutions obtained by different versions of *G-ALT*. This result is already known in other contexts. For example, Hansen and Mladenović (2006) demonstrate that first improvement and best improvement strategies may lead to different average solution qualities.

3.3.2 Constructing New Heuristics Based on G-ALT

Using the *G-ALT* framework we derive two new heuristics below. The first one is similar to the approach by Drezner et al. (2016) referred to as *IMP* that is based on the single-facility limited distance model. However, instead of solving globally for the optimal location of the selected facility keeping the remaining $p - 1$ facilities fixed, we solve approximately by repeated applications of *ALT* on the selected facility until no further improvement is possible. For this reason, we refer to the new procedure as *depth-first ALT*. The second proposed heuristic is a modified version of the decomposition algorithm (*DECOMP*) by Drezner et al. (2016). One of the unique features here is the different (and more flexible) way that the decomposed problems are constructed.

3.3.2.1 Depth-First Alternating Local Search (D-ALT)

Step 1: Initial locations for p facilities are generated (randomly or otherwise).

Step 2: Each demand point is assigned to its closest facility, thus partitioning the set of demand points into p disjoint subsets each attracted to a single facility.

Step 3: An assignment vector of length n is maintained along with an indicator vector of length p associated with the facilities that is initially set to all zeroes ($flag(j) = 0, j = 1, \ldots, p$).

Step 4: A facility j with $flag(j) = 0$ is randomly selected.

Step 5: Relocate facility j to the 1-median point X_j^* relative to its assigned subset of demand points. Set $flag(j) = 1$.

Step 6: All demand points are reallocated to their closest facilities. For each demand point that changed an assignment, the two facilities involved in the change (including possibly the facility that has just been relocated) have their *flags* reset to 0.

Step 7: If all $flags = 1$, *stop*; *elseif* $flag(j)$ of the facility j that was just selected $= 0$ again, re-select it and return to step 5 (following the *depth-first* strategy); else go to step 4.

3.3.2.2 Decomposition ALT (DEC(q))

Steps 1 to 3: Same as for *D-ALT*.

Step 4: A facility j at current location X_j, and with $flag(j) = 0$, is randomly selected.

Step 5: Perform *ALT* on the sub-problem defined by X_j and the $q - 1$ closest facilities to X_j (where $q < p$ is a parameter) using the subset of demand points currently allocated to these facilities as a given fixed set of demand points. (Also note that the $q - 1$closest facilities to X_j are selected irrespective of the value of their indicator variables.) Set $flag = 1$ for each of the q facilities that have just been relocated.

Step 6: All demand points are reallocated to their closest facilities. For each demand point that changed an assignment, the two facilities involved in the change (including possibly the facilities that have just been relocated) have their *flags* reset to 0.

Step 7: If all $flags = 1$, stop, *else* go to Step 4.

Many variants of $DEC(q)$ can be constructed and tested by varying the selected sub-problem size q. We may also select q randomly from a specified range of values each time step 5 is repeated, or gradually increase q to a limit p (or less), to add further variety for testing purposes. The depth-first approach of *D-ALT* may be incorporated as well by choosing X_j again if its *flag* is reset to 0 in step 6. Also note that *ALT* can be replaced with *IALT* in any such variant of $DEC(q)$.

3.4 Preliminary Computational Results

We plan to do extensive testing of *DEC*(q) in the future, but for now, present some preliminary results for *D-ALT* only. The experiments were coded in C++, and run on a PC Intel Core i7-6700/4GHz CPU vPro computer with single thread and 16 GB RAM. The results are summarized in Table 3.2 for a well-studied problem with $n = 1060$ points from Reinelt (1991). The first column specifies the number of facilities to be located, which varies in increments of 5 from $p = 5$ to 150, giving 30 problem instances in all. The next column gives the objective values of the best-known solutions obtained from the literature (see Drezner et al. 2016). Columns 3 and 4 give the % deviation from best-known value (column 2) for the best-found solution after 1000 runs of standard *ALT* and the new *D-ALT*, respectively. The average % deviation for the 1000 runs is given next (columns 5 and 6) followed by average CPU times (columns 7 and 8).

Initial solutions in step 1 are obtained by randomly selecting p demand points to locate the p facilities at. Also note that the same initial solutions are used for *ALT* and *D-ALT*. Based on the results in Table 3.2, we make the following observations:

1. Comparing best solutions, we see that standard *ALT* outperformed *D-ALT*. This may be due to the fact that *ALT* typically terminates after a small number of iterations, and this fast convergence may be advantageous for some (higher-quality) starting solutions.
2. Meanwhile *D-ALT* has a slight edge over *ALT* when we look at average performance. This is in spite of the higher observed degeneracy rate (facilities with no assigned customers) of *D-ALT* versus *ALT*, as shown in the next section.
3. *D-ALT* is roughly three times faster than *ALT*. This is a bit surprising as *D-ALT* requires a reallocation of all customers each time a selected facility is moved to a median point, whereas *ALT* performs this reallocation after solving p-independent 1-median problems. The faster CPU time may be explained by *D-ALT* having fewer calls to the Weiszfeld subroutine for solving 1-median problems because the *flags* of some facilities are never reset to 0 (or reset to 0 only a relatively small number of times).

Overall we see that even though these two variants of *G-ALT* have the same set of terminating solutions or local optima (Property 3.1), they follow different descent paths and hence may obtain different solutions from the same starting point.

3.5 Some Discussion

In the first stage of experiments, the initial facility locations were generated as random points within the smallest rectangle containing the set of demand points. This led to very poor solutions where several facilities had no customer assignments

Table 3.2 Summary results for large benchmark problem ($n = 1060$)

p	F_best	Dev_best (%)		Dev_ave (%)		CPU time (milsecs)	
		ALT	D-ALT	ALT	D-ALT	ALT	D-ALT
5	1851877.3	0	0	0.696	0.657	41.071	20.514
10	1249564.8	0	0.006	2.134	1.701	55.411	24.502
15	980131.7	0.007	0.013	3.354	3.144	57.435	22.669
20	828685.7	0.045	0.338	3.691	3.544	62.6	23.427
25	721988.2	0.215	0.632	4.713	4.575	68.324	23.325
30	638212.3	0.926	0.551	6.213	6.167	73.701	24.058
35	577496.7	0.991	2.029	7.092	6.535	81.875	24.579
40	529660.1	1.208	2.115	7.666	7.367	86.338	25.357
45	489483.8	1.368	1.662	8.338	8.165	90.53	26.198
50	453109.6	3.111	2.616	9.407	9.227	94.805	26.829
55	422638.7	3.516	3.038	10.223	10.145	100.133	28.507
60	397674.5	4.258	3.81	11.059	10.755	104.33	29.599
65	376630.3	3.701	4.984	11.016	11.203	108.739	31.089
70	357335.1	4.333	5.683	11.723	11.482	114.546	32.407
75	340123.5	4.987	5.501	11.978	12.221	116.844	33.809
80	325971.3	4.682	6.285	12.304	12.388	118.508	34.791
85	313446.6	5.301	6.181	12.336	12.221	123.838	35.894
90	302479.1	4.708	6.501	12.042	11.984	121.848	37.214
95	292282.6	4.685	6.271	11.907	11.96	124.649	37.957
100	282536.5	5.297	5.74	12.247	12.458	127.231	38.216
105	273463.3	5.599	7.258	12.201	12.255	132.431	39.004
110	264959.6	5.344	5.999	12.543	12.674	132.561	38.906
115	256735.7	6.919	6.429	12.762	12.768	137.8	39.484
120	249050.5	6.196	8.443	13.156	12.945	139.895	40.69
125	241880.4	6.78	7.542	13.299	13.55	141.778	41.495
130	235203.4	7.581	8.373	13.753	13.692	144.396	42.62
135	228999.2	7.341	8.157	13.982	13.965	145.741	43.741
140	223062	8.64	7.663	14.395	13.865	151.364	45.43
145	217462.8	8.737	9.211	14.586	14.563	154.197	46.973
150	212230.5	9.234	8.396	14.839	14.406	155.99	48.744
Average		4.190	4.714	10.189	10.086	110.297	33.601

because they were not the closest among the p facilities to any of the demand points. (See Table 3.3 for number of zeroes recorded under "Random I"). For this reason, we tried in a repeat stage of experiments reported in Table 3.2 (only for $n = 1060$ for brevity) to restrict the random selection of initial facility locations to the set of demand points. The results were much better (see "Random II").

Such degenerate solutions are clearly undesirable because some facilities are not being used and hence do not help at all in reducing the value of the objective function. However, if there are t such facilities with no customers assigned to them, we may have on the other hand a very good solution for the location problem with

Table 3.3 Summary results on degenerate solutions ($n = 1060$)

p	No. of zeroes (random I) ALT_ ave	ALT_ best	D-ALT_ ave	D-ALT_ best	No. of zeroes (random II) ALT_ ave	ALT_ best	D-ALT_ ave	D-ALT_ best
5	0.22	0	0.3	0	0	0	0	0
10	1.08	0	1.06	0	0	0	0	0
15	1.88	0	2.02	0	0	0	0	0
20	2.9	0	2.88	0	0	0	0	0
25	4.04	1	4.18	1	0	0	0	0
30	5.22	1	5.18	0	0	0	0	0
35	6.02	1	6.24	1	0	0	0	0
40	8.32	3	8.04	4	0	0	0	0
45	8.34	5	9	3	0	0	0.005	0
50	9.68	4	10.22	4	0	0	0.005	0
55	11.04	5	11.08	6	0	0	0.005	0
60	12.7	7	12.5	7	0.005	0	0.015	0
65	14.32	9	13.58	6	0.001	0	0.045	0
70	16	9	15.42	7	0.005	0	0.005	0
75	16.72	13	16.66	11	0.006	0	0.025	0
80	18.12	11	18.54	15	0.008	0	0.045	0
85	18.52	14	20.68	13	0.01	0	0.05	0
90	21.06	17	21.2	13	0.011	0	0.03	0
95	21.74	17	22.7	16	0.01	0	0.065	0
100	24.28	18	24.56	18	0.018	0	0.055	0
105	24.9	22	25.84	15	0.014	0	0.13	0
110	27.02	20	27.04	19	0.015	0	0.09	0
115	28.66	19	29.06	23	0.028	0	0.05	0
120	29.08	20	30.78	31	0.029	0	0.085	0
125	31.32	24	31.62	26	0.017	0	0.09	1
130	34.02	26	34.32	34	0.028	0	0.095	0
135	34.18	29	35.16	28	0.025	0	0.125	1
140	36.52	30	38.54	33	0.035	0	0.175	0
145	38.22	35	39.22	30	0.031	0	0.11	0
150	40.24	30	41.22	31	0.041	0	0.125	0
Ave	18.2120	13.0000	18.6280	13.1667	0.0112	0.0000	0.0475	0.0667

$p - t$ facilities. An effective insertion strategy for the t 'dormant' facilities, combined with the current configuration of the 'active' $p - t$ facilities, could therefore lead to a high-quality solution to the original problem. In other words, an undesirable situation may be turned into a profitable one! Although some recognition of degenerate solutions may be found in the literature (e.g., see Brimberg and Drezner (2013) and Brimberg and Mladenović (1999)), we believe this topic deserves much more attention.

This preliminary study also confirms previous observations (e.g., see Brimberg and Drezner 2013) on the importance of starting *ALT* with a 'good' initial solution. For example, notice in Table 3.3 that the best solution found in the first stage of experiments by *ALT* for $p = 150$ has 30 dormant facilities (or only 120 facilities in use). The objective value for this solution is 323560.8, which equates to a 29.9% deviation from the best-known solution for $p = 120$ (see Table 3.2). Meanwhile if we compare the results in Table 3.2 with different initial solutions, we see that the corresponding deviation for $p = 120$ is only 6.2%. Similar findings apply for the other instances tested, for both *ALT* and *D-ALT* and best and average results. We conclude that the way initial solutions are selected is an important component of these heuristics.

A related topic to degenerate solutions is the existence of saddle points (e.g., see Fig. 3.2), which to our best knowledge has not yet been recognized in the literature on *MWP*. Saddle points may be obtained intentionally with *ALT* (or *G-ALT*) by locating $p - t$ points, where t is now a parameter. For example, the case with $t = 1$ is equivalent to locating p facilities where two of them coincide at the same point. (Also note that this is equivalent to a degenerate solution where one of these two facilities is 'dormant'.) The idea then would be to explore the solution space in the vicinity of the saddle point to find local minimum points that are nearby. We can assume that the two coincident facilities are located at any one of the $p - 1$ current sites, and in this way generate several local minima from the same configuration. This too could be an exciting area for future research.

3.6 Conclusions

This chapter reviews some of the literature on the continuous p-median problem also known as the multisource Weber problem (*MWP*) or the continuous location-allocation problem, with a focus on some of the more recent contributions made by Professor Zvi Drezner in this area. We then extend a key element in most local searches applied to the *MWP* by providing a general framework for building variants of the classical alternating locate-allocate heuristic originally developed by Cooper (1963, 1964) and still popular to this date. This framework is then used to construct two heuristics which share some traits of two highlighted heuristics from Drezner et al. (2016).

Preliminary computational results confirm that although different heuristics constructed within the presented framework share the same set of local optima, they do not necessarily perform equally well. The computational tests also lead to some discussion on possible future research related to the exploitation of degenerate solutions and saddle points. Other directions for future research may include the development and testing of a wide range of 'alternating' heuristics constructed within the general framework presented here. This may also lead to a further understanding of the underlying structure of the problem, and to further insights on what makes such heuristics work well.

References

Bongartz, I., Calamai, P. H., & Conn, A. R. (1994). A projection method for l_p norm location-allocation problems. *Mathematical Programming, 66*, 283–312.

Brimberg, J., & Drezner, Z. (2013). A new heuristic for solving the p-median problem in the plane. *Computers & Operations Research, 40*, 427–437.

Brimberg, J., Drezner, Z., Mladenović, N., & Salhi, S. (2014). A new local search for continuous location problems. *European Journal of Operational Research, 232*, 256–265.

Brimberg, J., Hansen, P., & Mladenović, N. (2010). Attraction probabilities in variable neighborhood search. *4OR-Quarterly Journal of Operations Research, 8*, 181–194.

Brimberg, J., Hansen, P., Mladenović, N., & Salhi, S. (2008). A survey of solution methods for the continuous location-allocation problem. *International Journal of Operations Research, 5*, 1–12.

Brimberg, J., Hansen, P., Mladenović, N., & Taillard, É. (2000). Improvements and comparison of heuristics for solving the multisource Weber problem. *Operations Research, 48*, 444–460.

Brimberg, J., & Hodgson, M. J. (2011). Heuristics for location models. In H. A. Eiselt & V. Marianov (Eds.), *Foundations of location analysis: Industrial series in operations research & management science* (Vol. 155, pp. 335–355). New York, NY: Springer.

Brimberg, J., & Mladenović, N. (1996a). Solving the continuous location-allocation problem with Tabu search. *Studies in Locational Analysis, 8*, 23–32.

Brimberg, J., & Mladenović, N. (1996b). A variable neighborhood algorithm for solving the continuous location-allocation problem. *Studies in Locational Analysis, 10*, 1–12.

Brimberg, J., & Mladenović, N. (1999). Degeneracy in the multi-source Weber problem. *Mathematical Programming, 85*, 213–220.

Chen, R. (1983). Solution of minisum and minimax location-allocation problems with Euclidean distances. *Naval Research Logistics Quarterly, 30*, 449–459.

Cooper, L. (1963). Location-allocation problems. *Operations Research, 11*, 331–343.

Cooper, L. (1964). Heuristic methods for location-allocation problems. *SIAM Review, 6*, 37–53.

Drezner, Z. (1992). A note on the Weber location problem. *Annals of Operations Research, 40*, 153–161.

Drezner, Z. (1995). A note on accelerating the Weiszfeld procedure. *Location Science, 3*, 275–279.

Drezner, Z. (2002). A new heuristic for the quadratic assignment problem. *Journal of Applied Mathematics and Decision Sciences, 6*, 163–173.

Drezner, Z. (2005). The extended concentric Tabu for the quadratic assignment problem. *European Journal of Operational Research, 160*, 416–422.

Drezner, Z. (2015). The fortified Weiszfeld algorithm for solving the Weber problem. *IMA Journal of Management Mathematics, 26*, 1–9.

Drezner, Z., Brimberg, J., Mladenović, N., & Salhi, S. (2015a). Solving the planar p-median problem by variable neighborhood and concentric searches. *Journal of Global Optimization, 63*, 501–514.

Drezner, Z., Brimberg, J., Mladenović, N., & Salhi, S. (2015b). New heuristic algorithms for solving the planar p-median problem. *Computers & Operations Research, 62*, 296–304.

Drezner, Z., Brimberg, J., Mladenović, N., & Salhi, S. (2016). New local searches for solving the multi-source Weber problem. *Annals of Operations Research, 246*, 181–203.

Drezner, Z., Klamroth, K., Schöbel, A., & Wesolowsky, G. O. (2002). The Weber problem. In Z. Drezner & H. W. Hamacher (Eds.), *Facility location: Applications and theory* (pp. 1–36). Berlin: Springer.

Drezner, Z., Mehrez, A., & Wesolowsky, G. O. (1991). The facility location problem with limited distances. *Transportation Science, 25*, 183–187.

Drezner, Z., & Salhi, S. (2017). Incorporating neighbourhood reduction for the solution of the planar p-median problem. *Annals of Operations Research, 258*, 639–654.

Drezner, Z., & Simchi-Levi, D. (1992). Asymptotic behavior of the Weber location problem on the plane. *Annals of Operations Research, 40*, 163–172.

Eilon, S., Watson-Gandy, C. D. T., & Christofides, N. (1971). *Distribution management: Mathematical modelling and practical analysis*. London: Charles Griffin & Co., Ltd.

Hansen, P., & Mladenović, N. (2006). First vs best improvement: An empirical study. *Discrete Applied Mathematics, 154*, 802–817.

Hansen, P., Mladenović, N., & Taillard, E. (1998). Heuristic solution of the multisource Weber problem as a p-median problem. *Operations Research Letters, 22*, 55–62.

Katz, I. N. (1974). Local convergence in Fermat's problem. *Mathematical Programming, 6*, 89–104.

Krau, S. (1997). *Extensions du problème de Weber*. PhD Thèse, École Polytechnique de Montréal.

Kuhn, H. (1973). A note on Fermat's problem. *Mathematical Programming, 4*, 98–107.

Love, R. F., & Juel, H. (1982). Properties and solution methods for large location-allocation problems. *Journal of the Operational Research Society, 33*, 443–452.

Love, R. F., Morris, J. G., & Wesolowsky, G. O. (1988). *Facilities location: Models and methods*. New York: North-Holland.

Megiddo, M., & Supowit, K. J. (1984). On the complexity of some common geometric location problems. *SIAM Journal of Computing, 13*, 182–196.

Mladenović, N., & Brimberg, J. (1995). A descent-ascent technique for solving the multi-source Weber problem. *Yugoslav Journal of Operations Research, 5*, 211–219.

Murtagh, B. A., & Niwattisyawong, S. R. (1982). An efficient method for the multi-depot location-allocation problem. *Journal of the Operational Research Society, 33*, 629–634.

Ostresh, L. M., Jr. (1978). On the convergence of a class of iterative methods for solving the Weber location problem. *Operations Research, 26*, 597–609.

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

Reinelt, G. (1991). TSLIB a travelling salesman library. *ORSA Journal of Computing, 3*, 376–384.

Salhi, S. (2017). *Heuristic search: The emerging science of problem solving*. Cham, Switzerland.

Teitz, M. B., & Bart, P. (1968). Heuristic methods for estimating the generalized vertex median of a weighted graph. *Operations Research, 16*, 955–961.

Weber, A. (1909). *Über den Standort der Industrien*. Tübingen, Germany. (English translation: Friedrich CJ (translator) (1929), *Theory of the location of industries*. University of Chicago Press, Chicago)

Weiszfeld, E. (1936). Sur le point pour lequel la somme des distances de n points donnés est minimum. *Tohoku Mathematical Journal, 43*, 355–386.

Wesolowsky, G. O. (1993). The Weber problem: History and perspectives. *Location Science, 1*, 5–23.

Chapter 4
Big Triangle Small Triangle Method for the Weber Problem on the Sphere

Atsuo Suzuki

4.1 Introduction

We propose the Big Triangle Small Triangle (BTST) Method for solving the Weber problem on the sphere (WPS). It can also be applied to other single facility location problems on the sphere. The WPS is a variation of the Weber problem which is a classic and well-studied location problem. We assume that the demand points are distributed on the surface of a sphere, and our problem is to find the location of a facility so as to minimize the sum of the distances from demand points to the facility. The distance is measured by the great circle distance. The objective function of the WPS is not convex, and the Weiszfeld-type algorithm is not guaranteed to find the facility's optimal location. The BTST type algorithm divides the surface of the sphere into spherical triangles and applies a branch-and-bound method. We show that it finds the optimal solution within a short computational time.

As trade and services become global, there is an increased interest in location problems on a sphere. Most location problems, that are usually formulated on the plane, can also be formulated on the sphere. While planar distances are convex, great circle distances on the sphere are not. Therefore, even planar convex problems are not convex, in general, when formulated on the sphere. In order to find the global optimum on the sphere, global optimization techniques need to be utilized. One effective global optimization algorithm that was formulated on the plane is the "Big Triangle Small Triangle" (BTST) algorithm (Drezner and Suzuki 2004). We develop the spherical equivalent of the BTST algorithm.

The chapter is organized as follows: In Sect. 4.2 we review the BTST algorithm. In Sect. 4.3, we introduce the WPS, and provide a short literature review. In Sect. 4.4

A. Suzuki (✉)
Department of Systems and Mathematical Sciences, Nanzan University, Nagoya, Aichi, Japan
e-mail: atsuo@nanzan-u.ac.jp

H. A. Eiselt, V. Marianov (eds.), *Contributions to Location Analysis*, International Series in Operations Research & Management Science 281,
https://doi.org/10.1007/978-3-030-19111-5_4

we formulate the problem and list the properties we use for our algorithm. In Sect. 4.5 we describe the algorithm in detail. In Sect. 4.6 we derive the lower bound. In Sect. 4.7 we present the results of numerical experiments. We close the paper with the conclusion in Sect. 4.8. The derivation of a lower bound is detailed in the Appendix.

4.2 The Big Triangle Small Triangle Method

The Big Triangle Small Triangle method was proposed by Drezner and Suzuki (2004). The method solves optimally location problems with non-convex objective functions. Drezner and Suzuki (2004) solved the Weber problem with attraction and repulsion (WAR) and an obnoxious location problem. These problems have non-convex objective functions and unless we utilize the BTST method, it is difficult to obtain a guaranteed optimal solution.

The outline of the BTST method for a given relative accuracy ϵ is as follows:

1. Triangulation: The feasible region of the problem is divided into triangles. We utilize the Delaunay triangulation which is a dual graph of the Voronoi diagram generated by the demand points.
2. Initial upper bound *UB*: We evaluate the objective function at the centroid of each triangle and set the minimum of them as the initial upper bound.
3. Initial lower bound *LB* for each triangle: We evaluate a lower bound for each triangle.
4. Branch-and-bound

   (a) All triangles whose *LB* is greater than $UB(1-\epsilon)$ are discarded.
   (b) The triangle with the minimum *LB* (the "big triangle") is divided into four similar triangles ("small triangles") and the *UB* and *LB* in each triangle are evaluated. *UB* may be updated.
   (c) If the minimum of *LB* of all triangles is greater than $UB(1-\epsilon)$, then *UB* is the minimum of the objective function within a given relative accuracy ϵ. The solution is the centroid of the triangle which attains the *UB*.

The BTST method was motivated by the Big Square Small Square (BSSS) method proposed by Hansen et al. (1981, 1985). They solved variations of the Weber problem. The BTST method utilizes the triangulation of the feasible region into many triangles, while the BSSS method divides the feasible region, embedded in a big square, into small squares. The BSSS method is very simple and effective; however, the BTST method is also effective and is superior to the BSSS method in the following two points.

The first point is about the case when the feasible region is a polygon. In this case, the BSSS method divides the rectangle which includes the feasible region into rectangles which may have non-feasible areas. The BSSS method needs an additional procedure to avoid non-feasible solutions. On the other hand, the BTST method divides the polygon into triangles and all triangles and their interiors are feasible and the union of the triangles includes all feasible points.

The second point is that the triangles of the BTST method do not include demand points in their interior. A demand point in the interior of a triangle may complicate the derivation of an upper bound in a triangle. For example, in the Weber problem, the partial derivatives of the objective function are infinite at the demand points. In the BSSS method, some of the squares may include demand points and an upper bound may be ineffective. The triangles in the BTST method do not include demand points in their interior if we utilize the Delaunay triangulation described below.

The BTST method utilizes the Delaunay triangulation (Lee and Schachter 1980) which is a dual graph of the Voronoi diagram (Sugihara and Iri 1992; Okabe et al. 2000; Suzuki and Okabe 1995) to triangulate the feasible region. As described above, it is one of the reasons that BTST is better than BSSS. Voronoi diagrams and the Delaunay triangulation have been studied in the field of the computational geometry (Ohya et al. 1984; Sugihara and Iri 1992). The FORTRAN source program is available to the public on the WEB site of Kokichi Sugihara. Figure 4.1 is an example of the Voronoi diagram generated by randomly distributed 256 points on the plane.

Drezner and Suzuki studied various location problems using the Voronoi diagram. Examples include: the continuous p-center problem (Suzuki and Drezner 1996), the continuous hub model (Suzuki and Drezner 1997), the equitable radius circle problem (Suzuki and Drezner 2009), and the covering problem (Drezner and Suzuki 2010). They used the Delaunay triangulation (Lee and Schachter 1980) to triangulate the feasible region effectively in the application of the BTST algorithm.

Drezner utilized the BTST method for various location problems. They are listed in the introduction of Drezner (2007). After that Drezner has continued to apply the BTST method to various location problems, such as the multi-Weber problem

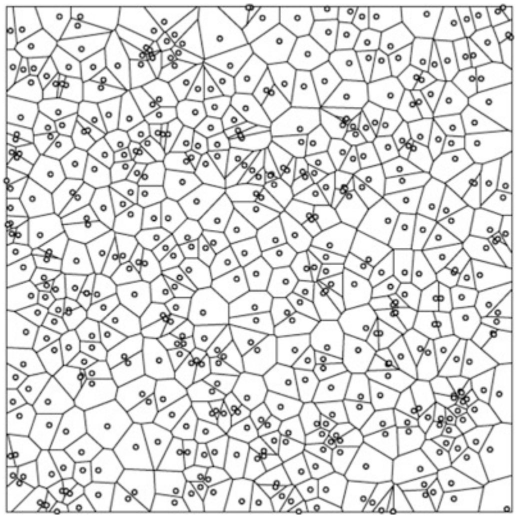


Fig. 4.1 Voronoi diagram for 256 generators by the program of Ohya et al. (1984)

(Suzuki and Drezner 2013), Huff based competitive location model (Drezner and Drezner 2004), and the network nuisance location problem (Drezner et al. 2009). It is almost impossible to obtain a guaranteed optimal solution for these problems without applying the BTST method.

4.3 The WPS Problem

The WPS problem is a variation of the Weber problem. The objective function of the classic Weber problem is convex, and by the algorithm first proposed by Weiszfeld (1936) and improved by Drezner (1996, 2015), the optimal solution is obtained. On the other hand, as the objective function of the WPS is not convex, the Weiszfeld-type algorithm cannot guarantee that the solution found is optimal. The WPS has been studied by many researchers, and many heuristics have been proposed for its solution. For example, Drezner and Wesolowsky (1978) proposed a heuristic based on the Weiszfeld algorithm. Other heuristic algorithms are surveyed by Hansen et al. (1995).

To the best of our knowledge, there are only two algorithms that obtain the optimal solution of the WPS. Drezner (1985) proposed an iterative algorithm based on the transformation of the problem into a series of minimax problems on the sphere. The algorithm adds extra points on the sphere one by one, and solves the minimax problem iteratively. The convergence of the algorithm to the optimal point is proved. However, numerical examples are not presented, and the convergence is expected to be slow.

Hansen et al. (1995) proposed another algorithm called the Big Region Small Region (BRSR) method based on the BSSS method. The algorithm of the BRSR divides the feasible region into spherical rectangles. It obtains a lower bound of the objective function in each spherical region. By the branch-and-bound process, they narrow the region where the solution exists. They implemented the algorithm and present computational results.

We apply the BTST type method described in the former section for solving the WPS. For the triangulation of the sphere, we utilize the algorithm of "spherical" Voronoi diagram developed by Sugihara (2002). Figure 4.2 is an example of the spherical Voronoi diagram drawn by the program available by Sugihara for randomly distributed 50 generators on the surface of the sphere.

The BTST method performs better than the BRSR method for the same reasons that the ordinary BTST method is superior to BSSS method. Furthermore, when the BRSR method divides the sphere into spherical rectangles, it uses latitude and longitude. As a result, it treats the regions near the north and south poles differently, while the BTST method does not require such special treatment.

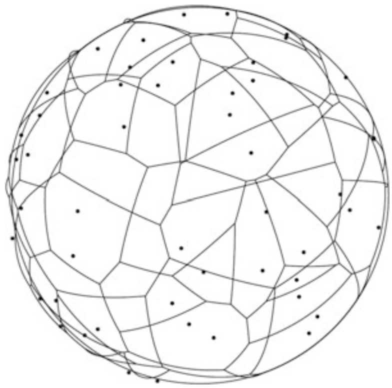


Fig. 4.2 Perspective view of a Voronoi diagram on the sphere for randomly distributed 50 generators by the program in Sugihara (2002)

4.4 Formulation of the Weber Problem on the Sphere

There are n demand points located on the sphere. Demand point i has an associated weight $w_i > 0$. The points on the sphere are represented by polar coordinates:

$$a_i = (\phi_i, \theta_i), -\frac{\pi}{2} \leq \phi_i \leq \frac{\pi}{2}, -\pi \leq \theta_i \leq \pi, i = 1, \ldots, n.$$

ϕ is the longitude and θ is the latitude if we consider the sphere as the globe.

The distance between two points on the sphere is measured by the great circle distance. The distance between $s(\phi, \theta)$ and $a_i(\phi_i, \theta_i)$ is

$$d(s, a_i) = \arccos\{\cos\phi \cos\phi_i \cos(\theta - \theta_i) + \sin\phi \sin\phi_i\}.$$

The great circle distance is the shortest distance from s to a_i when traveling on the surface of the globe. Note that $0 \leq d(s, a_i) \leq \pi$.

The objective function of the WPS is

$$F(s) = \sum_{i=1}^{n} w_i d(s, a_i). \tag{4.1}$$

We assume that $w_i > 0$. If $w_i < 0$ for some $1 \leq i \leq n$, we replace a_i by its antipode with weight $-w_i$. By Property 4.2 described below, we can transform the case where $w_i < 0$ for some values of i into problem (4.1). It means that if the WPS can be solved, the Weber problem with attraction and repulsion on the sphere can be solved as well.

As the WPS attracts significant interest, various properties have been studied. They are listed by Hansen et al. (1995). We show several definitions and properties which are utilized in our algorithm.

Definition 4.1 Given a center point and radius, a *spherical circle* is the set of the points whose shortest distance from the center is equal to the radius.

Definition 4.2 A set on the surface of the sphere is convex if all the points on the shortest arc connecting two points in the set are included in the set.

Definition 4.3 We define a function f on a convex set S on the surface of a sphere. f is convex if and only if for any two points a and b in S, and all $0 \leq \lambda \leq 1$,

$$f(\lambda a + (1-\lambda)b) \leq \lambda f(a) + (1-\lambda)f(b).$$

Property 4.1 The distance from a given point s to a point on the sphere is a convex function if the point is in the spherical circle whose center is s and radius is less than or equal to $\frac{\pi}{2}$.

Property 4.1 is proved in many papers such as Katz and Cooper (1980) and Drezner and Wesolowsky (1978). We utilize this property to obtain the lower bound in the BTST method.

Definition 4.4 The antipode of a point $a(\phi, \theta)$ on the sphere is $\bar{a}(-\phi, \theta - \pi)$ if $\theta \geq 0$ or $\bar{a}(-\phi, \theta + \pi)$ if $\theta \leq 0$.

The following property is easy to prove by the definition of the great circle distance and Definition 4.4.

Property 4.2 The sum of the distances from s to a and from s to $\bar{a}$ is equal to π.

Before describing the details of the algorithm, we re-write the objective function to facilitate the derivation of the lower bound for the BTST method. We define index sets of demand points. We consider a spherical triangle T whose vertices are T^k, $k = 1, 2, 3$. The polar coordinates of T^k are (ϕ^k, θ^k), $k = 1, 2, 3$. I is the index set of all demand points. We define the following:

$$I^k = \{i \in I | d(T^k, a_i) \leq \frac{\pi}{2}\}, k = 1, 2, 3$$

$$\bar{I^k} = I \setminus I^k.$$

I^k is a set of demand points whose distance from a vertex T^k of the triangle T is less than or equal to $\frac{\pi}{2}$.

We divide the objective function into three terms. The first term is the sum of the weighted distances from the demand points whose distance from all three vertices of the triangle is less than or equal to $\frac{\pi}{2}$. The second term is the sum of the weighted distances from the demand points whose distance from all three vertices of the triangle is greater than or equal to $\frac{\pi}{2}$. The third term is the sum for the rest of the demand points

$$
\begin{aligned}
F(s) &= \sum_{i \in I} w_i d(s, a_i) \\
&= \sum_{i \in \cap_{k=1}^{3} I^k} w_i d(s, a_i) + \sum_{i \in \cap_{k=1}^{3} \bar{I}^k} w_i d(s, a_i) + \sum_{i \in I \setminus (\cap_{k=1}^{3} I^k \cup \cap_{k=1}^{3} \bar{I}^k)} w_i d(s, a_i) \quad (4.2) \\
&\equiv F_1(s) + F_2(s) + F_3(s).
\end{aligned}
$$

The first term is convex for any s in the triangle, and the second term is concave. The third term is neither convex nor concave. We prove the convexity of the first term and the concavity of the second term.

First, we show that the first term of (4.2) is convex. Consider a spherical circle C_i whose center is a_i and radius $\frac{\pi}{2}$. By Property 4.1, $d(t, a_i)$ is a convex function in C_i. Note that $T^k \in C_i$. Let T' be a point on the edge T^1T^2. As $d(t, a_i)$ is a convex function in C_i, and C_i is convex, by Definition 4.3,

$$
d(T', a_i) \le \lambda d(T^1, a_i) + (1 - \lambda) d(T^2, a_i)) \le \frac{\pi}{2}, 0 \le \lambda \le 1.
$$

Let T'' be a point on the edge $T'T^3$. By the same reason above,

$$
d(T'', a_i) \le \lambda' d(T', a_i) + (1 - \lambda') d(T^3, a_i)) \le \frac{\pi}{2}, 0 \le \lambda' \le 1.
$$

As T'' moves in T when λ and λ' vary in their range,

$$
d(s, a_i) \le \frac{\pi}{2}, \forall s \in T
$$

Then, by Property 4.1, $d(s, a_i)$ is convex function of $\forall s \in T$. As $F_1(s)$ is the weighted sum of the convex functions and the weights are positive, $F_1(s)$ is a convex function.

Next, we show that the second term of (4.2) is concave. By Property 4.2, the term is rewritten as follows:

$$
\begin{aligned}
F_2(s) = \sum_{i \in \cap_{k=1}^{3} \bar{I}^k} w_i d(s, a_i) &= \sum_{i \in \cap_{k=1}^{3} \bar{I}^k} w_i \{\pi - d(s, \bar{a}_i)\} \\
&= \pi \sum_{i \in \cap_{k=1}^{3} \bar{I}^k} w_i - \sum_{i \in \cap_{k=1}^{3} \bar{I}^k} d(s, \bar{a}_i). \quad (4.3)
\end{aligned}
$$

As the distance from s to a_i is larger than $\frac{\pi}{2}$ for $a_i, i \in \cap_{k=1}^{3} \bar{I}^k$, the distance from s to the $\bar{a}_i$ is less than or equal to $\frac{\pi}{2}$. It means that the distance from s to the $\bar{a}_i$ is a convex function of s. As the second term of (4.3) is the sum of convex functions, it is convex. The first term of (4.3) is constant. As F_2 is a difference between a constant and a convex function, it is a concave function. This can also be proved by observing that all the distances from the antipode are less than or equal to $\frac{\pi}{2}$ and in general if $F(X)$ is convex, then $-F(X)$ is concave.

4.5 Outline of the BTST Method for the Weber Problem on the Sphere

The outline of the algorithm is similar to the planar BTST method. In the algorithm, we use spherical triangles rather than planar triangles. The calculation of the lower bound is more complicated than the calculation for the planar BTST method. We show the details in the next section.

1. Triangulation: The sphere is divided into spherical triangles. We utilize the Delaunay triangulation of the sphere which is a dual graph of the Voronoi diagram on the sphere.
2. Initial upper bound *UB*: We evaluate the objective function at the centroid of the spherical triangles, and set the minimum of them as the initial upper bound.
3. Initial lower bound *LB* for each spherical triangle: We evaluate a lower bound of the objective function for each spherical triangle.
4. Branch-and-bound
   (a) All spherical triangles whose *LB* is greater than $UB(1-\epsilon)$ are discarded.
   (b) The spherical triangle with the minimum *LB* is divided into four similar spherical triangles by connecting the middle points of the three edges of the triangle. The *UB* and *LB* in these triangles are evaluated. *UB* may be updated.
   (c) If the minimum of *LB* of all spherical triangles is greater than $UB(1-\epsilon)$, then *UB* is the solution.

As the initial upper bound, we calculate the weighted sum of the distances from the centroid of each spherical triangle to the demand points. The initial upper bound *UB* is the minimum of these upper bounds. Note that if there are n demand points, the number of the triangles in the initial triangulation is $2n-4$. The calculation of the initial upper bound is calculated in linear time.

By our experience, the effectiveness of the algorithm strongly depends on the quality of the lower bound like in any branch-and-bound method. In the next section we describe the details of the calculation of the lower bound used in the computational experiments.

In the branch-and-bound process, we use a heap with the key of the lower bound of the spherical triangles. We scan the heap to discard spherical triangles with *LB* greater than the $UB(1-\epsilon)$. When we pick the spherical triangle with the smallest *LB*, the spherical triangle is at the top of the heap, and we update the heap. Then we add four spherical triangles to the heap. As the spherical triangle with minimum *LB* is always on the top of the heap, comparing the minimum *LB* to $UB(1-\epsilon)$ is quite easy.

4.6 Calculation of the Lower Bound

To obtain a lower bound, we need to consider the first term, which is convex, and the third term of (4.2), which is neither convex nor concave. The second term is concave. Therefore, its minimum should be attained at one of the vertices of the spherical triangle T.

The first term of (4.2) is transformed as follows. A point s in T is represented as a linear combination of the three vertices of the spherical triangle T^1, T^2, and T^3

$$G_1(t_1, t_2, t_3) \equiv F_1(t_1T^1 + t_2T^2 + t_3T^3),$$

where

$$T^k = (\phi^k, \theta^k); \quad s = t_1T^1 + t_2T^2 + t_3T^3, t_1 + t_2 + t_3 = 1. \tag{4.4}$$

As F_1 is convex in T, G_1 is convex in the triangle defined by (4.4). For a lower bound of a convex function, we propose the tangent plane method in Drezner and Suzuki (2004). We calculate the tangent plane of the convex function and set the minimum of the tangent plane values on the three vertices of the triangle as the lower bound. In this case, the lower bound is obtained as follows:

$$\begin{aligned} G_1(t_1, t_2, t_3) = \sum_{i\in\cap_{k=1}^3 I^k} & w_i \arccos\{\cos(t_1\phi^1 + t_2\phi^2 + t_3\phi^3)\cos\phi_i \\ & \times \cos(t_1\theta^1 + t_2\theta^2 + t_3\theta^3 - \theta_i) \\ & + \sin(t_1\phi^1 + t_2\phi^2 + t_3\phi^3)\sin\phi_i\}. \end{aligned}$$

The tangent plane of $G_1(t_1, t_2, t_3)$ at $(\bar{t}_1, \bar{t}_2, \bar{t}_3)$ is

$$\begin{aligned} \bar{G}_1(t_1, t_2, t_3) = {} & G_1(\bar{t}_1, \bar{t}_2, \bar{t}_3) \\ & + \left.\frac{\partial G_1}{\partial t_1}\right|_{t_1=\bar{t}_1, t_2=\bar{t}_2, t_3=\bar{t}_3} (t_1 - \bar{t}_1) \\ & + \left.\frac{\partial G_1}{\partial t_2}\right|_{t_1=\bar{t}_1, t_2=\bar{t}_2, t_3=\bar{t}_3} (t_2 - \bar{t}_2) \\ & + \left.\frac{\partial G_1}{\partial t_3}\right|_{t_1=\bar{t}_1, t_2=\bar{t}_2, t_3=\bar{t}_3} (t_3 - \bar{t}_3). \end{aligned} \tag{4.5}$$

As the value of $F_1(s)$ is equal to that of $G_1(t_1, t_2, t_3)$, $s = t_1T^1 + t_2T^2 + t_3T^3$, and $G_1(t_1, t_2, t_3) \leq \bar{G}_1(t_1, t_2, t_3)$,

$$F_1(s) \geq \bar{G}_1(t_1, t_2, t_3), s = t_1T^1 + t_2T^2 + t_3T^3.$$

The derivation of $\partial G_1/\partial t_k$, $k = 1, 2, 3$ is given in the Appendix.

As mentioned above, $F_2(s)$ is concave. Therefore, $F_4(s) = F_2(s) + \bar{G}_1(t_1, t_2, t_3)$ is concave as a sum of a concave function and a linear function by Eq. (4.5). The minimum of $F_4(s)$ is attained at a vertex of T

$$F_4(s) = \sum_{i \in \cap_{k=1}^{3} \bar{I}^k} w_i d(s, a_i) + F_1(s)$$
$$\geq LB_1 = \min_{k=1,2,3} \left\{ \sum_{i \in \cap_{k=1}^{3} \bar{I}^k} w_i d(T^k, a_i) + \bar{G}^{(k)} \right\},$$

where $G^{(1)} = G_1(1, 0, 0)$; $G^{(2)} = G_1(0, 1, 0)$; $G^{(3)} = G_1(0.0.1)$.

$F_3(s)$ is neither convex nor concave. The distance from a demand point to any point in T is higher than the minimum distance from the demand point to the triangle T. Therefore,

$$F_3(s) = \sum_{i \in I \setminus (\cap_{k=1}^{3} I^k \cup \cap_{k=1}^{3} \bar{I}^k)} w_i d(s, a_i)$$
$$\geq LB_2 = \sum_{i \in I \setminus (\cap_{k=1}^{3} I^k \cup \cap_{k=1}^{3} \bar{I}^k)} w_i \min_{s \in T} d(s, a_i).$$

The details of the derivation of $\min_{s \in T} d(s, a_i)$ is given in the Appendix.

As the lower bound LB_2 is a naïve one, its quality may not be very good. However, the size of the triangles is relatively small as the algorithm progresses. The number of demand points in $I \setminus (\cap_{k=1}^{3} I^k \cup \cap_{k=1}^{3} \bar{I}^k)$ becomes small. Therefore, the quality of the lower bound improves as the algorithm progresses.

As the result,

$$F(s) \geq LB_1 + LB_2.$$

4.7 Computational Experiments

We implemented the procedure in FORTRAN and tested randomly generated demand points on the sphere. The computer used for the experiments has Intel Core i5-4200U CPU (1.6 GHz clock), the operating system is Windows 10 Professional, and the RAM is 4.00 GB. The FORTRAN Compiler we used is FORTRAN Power Station version 4.

We tested problems with 100, 200, 300, 400, and 500 demand points which are randomly distributed on the sphere. All the weights are equal to 1. We compare the effectiveness of our algorithm with the one in Hansen et al. (1995). As they used $\epsilon = 10^{-3}$ as the stopping criterion, we used the same ϵ in our tests. We also used $\epsilon = 10^{-6}$ to obtain more precise optimal function values. For every combination of the number of demand points and ϵ, we solved 10 sample problems. We measure the number of iterations and CPU time to obtain the optimum objective function values. In Table 4.1 we report the average and standard deviation of the number of iterations and the run times in seconds of the 10 runs. We also report the average value of the objective function and for $\epsilon = 10^{-3}$ and report its percentage above the average obtained using $\epsilon = 10^{-6}$. In Table 4.2 we give the details of the 10 results for the $n = 500$ instances.

The computer (Hansen et al. 1995) used for the computational experiments is SPARC 4/75-64 (28.5 MIPS, 64M RAM, 4.2 MFLOPS). The computer we used has 4.84 GFLOPS (2.42 GFLOPS per one core and the CPU has two cores). The ratio of the computer speeds is 1152 times. In Table 4.1, the average CPU times are 0.16, 0.41, 0.77, 1.09, and 1.26 s for 100, 200, 300, 400, and 500 demand points sample problems, respectively, while in Hansen et al. (1995), the CPU times for the same size of the sample problems are 517, 958, 2586, 3018, and 5233 s. The ratios between these are 3230, 2320, 3360, 2740, and 4000. As the speed of the computers increases 1152 times, the CPU time by our algorithm is at least twice as fast.

For more precise solutions, we use $\epsilon = 10^{-6}$ as the stopping rule of the algorithm. It appears that the CPU times to obtain the solutions increase moderately.

Table 4.1 Summary results for $\epsilon = 10^{-3}$ and $\epsilon = 10^{-6}$

	Iterations		CPU time		Objective value	
n	Mean	S.D.	Mean	S.D.	Mean	[a]
$\epsilon = 10^{-3}$						
100	212.6	76.7	0.17	0.05	145.4025	0.0019%
200	241.8	91.4	0.41	0.11	299.2156	0.0004%
300	292.5	158.4	0.78	0.28	454.1721	0.0005%
400	265.6	198.3	1.09	0.45	604.7062	0.0002%
500	190.5	75.4	1.26	0.22	757.1503	0.0003%
$\epsilon = 10^{-6}$						
100	14555.3	28154.0	22.41	54.42	145.3997	–
200	20435.0	23103.1	41.21	51.96	299.2144	–
300	12837.4	24230.9	34.90	75.29	454.1697	–
400	8235.3	12164.0	23.30	36.79	604.7047	–
500	11040.1	26156.0	46.32	115.86	757.1481	–

[a]% above optimum

Table 4.2 Results for $n = 500$

Case	$\epsilon = 10^{-3}$			$\epsilon = 10^{-6}$		
	Iteration	Time	Value	Iteration	Time	Value
1	245	1.44	770.8228	1524	5.05	770.8228
2	118	1.05	752.8868	2385	8.06	752.8784
3	157	1.14	757.7859	703	2.72	757.7853
4	254	1.44	765.9260	2631	8.30	765.9211
5	79	0.94	741.5615	310	1.61	741.5582
6	313	1.59	776.6203	1676	5.39	776.6203
7	259	1.47	752.1417	89327	393.50	752.1416
8	181	1.25	756.8979	1413	4.80	756.8967
9	212	1.33	761.8746	3460	10.98	761.8717
10	87	0.97	734.9853	6972	22.83	734.9851
Mean	190.5	1.26	757.1503	11040.1	46.32	757.1481
S.D.	75.4	0.22	12.01	26156.0	115.86	12.01

4.8 Conclusions

We proposed the spherical BTST algorithm for the Weber problem on a sphere (WPS). It obtained the exact solution in a short CPU time. For the calculation of the lower bound of the branch and bound process of the BTST, we divide the objective function into three terms. The first one is the sum of the weighted distances from the point in the triangle to the demand points whose distance is less than $\frac{\pi}{2}$ from all the points in the spherical triangle. We show the term is convex and use a variation of the tangent plane method to obtain the lower bound. The second one is the sum of the weighted distances from the point in the triangle to the demand points whose distance is larger than or equal to $\frac{\pi}{2}$. We show that the second term is concave. Therefore, the sum of the second term and the tangent plane constructed for the first term is also concave. The lower bound for the sum is attained at one of the vertices of the triangle. For the third one, we calculate the shortest distance from the demand points which is neither less than $\frac{\pi}{2}$ nor larger than or equal to $\frac{\pi}{2}$. We implemented the algorithm and found that it obtained the solution in a short CPU time.

The methodology proposed in this paper can be used to optimally solve other location problems on a sphere. Competitive facility location, forbidden regions (such as bodies of water or enemy countries) using the Weber or the minimax objective, finding the location that attracts the maximum weight within a given distance, equity location models such as minimizing the range, the variance, and other location models that were investigated in the plane.

Appendix

Partial Derivatives of G_1

The calculation of the partial derivatives of $G_1(t_1, t_2, t_3)$ is as follows:

$$g_{1i}(t_1, t_2, t_3) \equiv \cos(t_1\phi^1 + t_2\phi^2 + t_3\phi^3)\cos\phi_i \times \cos(t_1\theta^1 + t_2\theta^2 + t_3\theta^3 - \theta_i)$$
$$+ \sin(t_1\phi^1 + t_2\phi^2 + t_3\phi^3)\sin\phi_i.$$

Then

$$G_1(t_1, t_2, t_3) = \sum_{i \in \cap_{k=1}^3 I^k} w_i g_{1i}(t_1, t_2, t_3).$$

The partial derivatives of G_1 are

$$\frac{\partial G_1}{\partial t_k} = -\sum_{i \in \cap_{k=1}^3 I^k} w_i (1 - g_{1i}(t_1, t_2, t_3)^2)^{\frac{1}{2}} \frac{\partial}{\partial t_k} g_{1i}(t_1, t_2, t_3), k = 1, 2, 3.$$

The partial derivatives of g_{1k} are calculated as follows:

$$\frac{\partial}{\partial t_k} g_{1i}(t_1, t_2, t_3) = \cos\phi_i\{-\phi^k \sin(t_1\phi^1 + t_2\phi^2 + t_3\phi^3) \cdot \cos(t_1\theta^1 + t_2\theta^2 + t_3\theta^3 - \theta_i)$$
$$- \theta^k \cos(t_1\phi^1 + t_2\phi^2 + t_3\phi^3) \cdot \sin(t_1\theta^1 + t_2\theta^2 + t_3\theta^3 - \theta_i)\}$$
$$+ \sin\phi_i\{\phi^k \cos(t_1\phi^1 + t_2\phi^2 + t_3\phi^3)\}.$$

Shortest Distance to the Triangle T

For the lower bound of $F_3(s)$, we need to calculate the shortest distance from a demand point to the triangle T. The shortest distance from a demand point is attained on the edges or the vertices of T. Therefore, we need to obtain the shortest distance from the demand point $P(a_i)$ to an edge T^1T^2 of the triangle T. We calculate the distance from the demand point to the three edges and set the minimum of them as the shortest distance from the demand point to the triangle T.

For the calculation of the distance from the demand point $P(a_i)$ to an edge T^1T^2, we need to consider two cases. Consider the spherical angles $\alpha = \angle PT^1T^2$ and $\beta = \angle PT^2T^1$ (see Fig. 4.3).

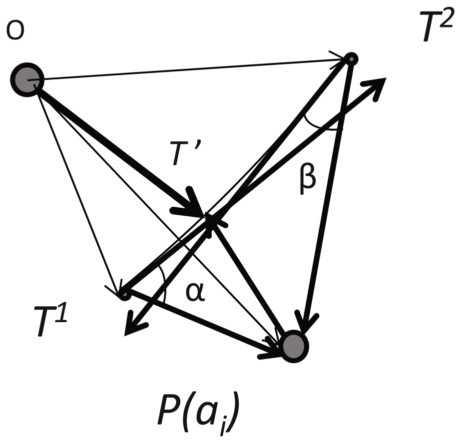


Fig. 4.3 The shortest distance from a demand point to the edge T^1T^2

The first case is that both α and β are less than $\pi/2$. In this case, the shortest distance from P to T^1T^2 is attained at a point T' on T^1T^2 see Fig. 4.3. The distance is calculated as

$$d(T', a_i) = \arccos(\overrightarrow{OT'} \cdot \overrightarrow{OP}),$$

where

$$\overrightarrow{OT'} = \frac{\overrightarrow{OP} - k(\overrightarrow{OT^1} \times \overrightarrow{OT^2})}{||\overrightarrow{OP} - k(\overrightarrow{OT^1} \times \overrightarrow{OT^2})||},$$

$$k = \frac{(\overrightarrow{OT^1} \times \overrightarrow{OT^2}) \cdot \overrightarrow{OP}}{||\overrightarrow{OT^1} \times \overrightarrow{OT^2}||}.$$

The second case is that either α or β is greater than or equal to $\frac{\pi}{2}$. If $\alpha \geq \frac{\pi}{2}$, the shortest distance is attained at T^1. If $\beta \geq \frac{\pi}{2}$, the shortest distance is attained at T^2.

References

Drezner, T., & Drezner, Z. (2004). Finding the optimal solution to the Huff competitive location model. *Computational Management Science, 1*, 193–208.

Drezner, T., Drezner, Z., & Scott, C. H. (2009). Location of a facility minimizing nuisance to or from a planar network. *Computers & Operations Research, 36*, 135–148.

Drezner, Z. (1985). A solution to the Weber location problem on the sphere. *Journal of the Operational Research Society, 36*, 333–334.

Drezner, Z. (1996). A note on accelerating the Weiszfeld procedure. *Location Science, 3*, 275–279.

Drezner, Z. (2007). A general global optimization approach for solving location problems in the plane. *Journal of Global Optimization, 37*, 305–319.
Drezner, Z. (2015). The fortified Weiszfeld algorithm for solving the Weber problem. *IMA Journal of Management Mathematics, 26*, 1–9.
Drezner, Z., & Suzuki, A. (2004). The big triangle small triangle method for the solution of non-convex facility location problems. *Operations Research, 52*, 128–135.
Drezner, Z., & Suzuki, A. (2010). Covering continuous demand in the plane. *Journal of the Operational Research Society, 61*, 878–881.
Drezner, Z., & Wesolowsky, G. O. (1978). Facility location on a sphere. *Journal of the Operational Research Society, 29*, 997–1004.
Hansen, P., Jaumard, B., & Krau, S. (1995). An algorithm for Weber's problem on the sphere. *Location Science, 3*, 217–237.
Hansen, P., Peeters, D., Richard, D., & Thisse, J.-F. (1985). The minisum and minimax location problems revisited. *Operations Research, 33*, 1251–1265.
Hansen, P., Peeters, D., & Thisse, J.-F. (1981). On the location of an obnoxious facility. *Sistemi Urbani, 3*, 299–317.
Katz, I. N., & Cooper, L. (1980). Optimal location on a sphere. *Computers & Mathematics with Applications, 6*, 175–196.
Lee, D. T., & Schachter, B. J. (1980). Two algorithms for constructing a Delaunay triangulation. *International Journal of Parallel Programming, 9*, 219–242.
Ohya, T., Iri, M., & Murota, K. (1984). Improvements of the incremental method of the Voronoi diagram with computational comparison of various algorithms. *Journal of the Operations Research Society of Japan, 27*, 306–337.
Okabe, A., Boots, B., Sugihara, K., & Chiu, S. N. (2000). *Spatial tessellations: Concepts and applications of Voronoi diagrams*. *Wiley series in probability and statistics*. New York: Wiley.
Sugihara, K. (2002). Laguerre Voronoi diagram on the sphere. *Journal for Geometry and Graphics, 6*, 69–81.
Sugihara, K., & Iri, M. (1992). Construction of the Voronoi diagram for "one million" generators in single-precision arithmetic. *Proceedings of the IEEE, 80*, 1471–1484.
Suzuki, A., & Drezner, Z. (1996). The *p*-center location problem in an area. *Location Science, 4*, 69–82.
Suzuki, A., & Drezner, Z. (1997). On the airline hub problem: The continuous model. *The Journal of the Operations Research Society of Japan, 40*, 62–74.
Suzuki, A., & Drezner, Z. (2009). The minimum equitable radius location problem with continuous demand. *European Journal of Operational Research, 195*, 17–30.
Suzuki, A., & Drezner, Z. (2013). Solving constrained two-facility location problems. *The Journal of the Operations Research Society of Japan, 56*, 157–165.
Suzuki, A., & Okabe, A. (1995). Using Voronoi diagrams. In Z. Drezner (Ed.), *Facility location: A survey of applications and methods* (pp. 103–118). New York: Springer.
Weiszfeld, E. (1936). Sur le point pour lequel la somme des distances de n points donnés est minimum. *Tohoku Mathematical Journal, 43*, 355–386.

Chapter 5
Integer Location Problems

Anita Schöbel

5.1 Introduction

The goal of this paper is to introduce *integer* location problems. These are continuous location problems in which we look for a new facility with integer coordinates. We motivate why research on integer location problems is useful and sketch an application within robust optimization. We then analyze the structure of optimal integer locations: We identify integer location problems for which a finite dominating set can be constructed and we identify cases in which the integer problem can be solved by rounding the solution of the corresponding continuous location problem. We finally propose a geometric branch-and-bound procedure for solving integer location problems.

Planar location problems, the most prominent being the Weber problem, have a long tradition in locational analysis and are by now well understood.

An instance of a *planar median location problem* (L) is given by a set $A_1, \ldots, A_m \in \mathbb{R}^2$ of m demand points in the plane. A weight $w_m \geq 0$ is assigned to each demand point A_i, $i = 1, \ldots, m$. The goal is to find a new facility $x \in \mathbb{R}^2$ which minimizes the weighted sum of distances

$$f(x) = \sum_{i=1}^{m} w_i d(x, A_i)$$

A. Schöbel (✉)
Department of Mathematics, Technical University Kaiserslautern, Kaiserslautern, Germany

Fraunhofer Institute for Industrial Mathematics ITWM, Kaiserslautern, Germany
e-mail: schoebel@mathematik.uni-kl.de

H. A. Eiselt, V. Marianov (eds.), *Contributions to Location Analysis*, International Series in Operations Research & Management Science 281,
https://doi.org/10.1007/978-3-030-19111-5_5

to the existing facilities, i.e.,

$$\text{(L)} \quad \min \left\{ f(x) : x \in \mathbb{R}^2 \right\}. \tag{5.1}$$

If the distance $d(\cdot, \cdot)$ is the Euclidean distance ℓ_2, the problem is called *Fermat-Torricelli* or *Weber problem*. Many variations of the Weber problem have been considered. We refer to numerous textbooks such as Love et al. (1988), Nickel and Puerto (2005), or to compendiums, e.g., Drezner (1995), Hamacher and Drezner (2002), Eiselt and Marianov (2011), Laporte et al. (2015), or to survey articles, e.g., Wesolowsky (1993), Drezner et al. (2002), ReVelle and Eiselt (2005) for details on the Weber problem including its history, algorithms, applications, and extensions to other location problems.

The planar median location problem has not only been studied with the Euclidean distance, but also with other distances, e.g., with the squared Euclidean distance ℓ_2^2 (White 1971), the Manhattan distance ℓ_1, with the Chebyshev distance ℓ_∞, or with ℓ_p-distances (Drezner and Wesolowsky 1980; Brimberg and Love 1993), block norms, or polyhedral gauges, see Thisse et al. (1984) and Ward and Wendell (1985).

As objective function, not only the sum of distances is used in location problems, but also *center location problems* in which the maximum distance

$$g(x) = \max_{i=1}^{m} d(x, A_i)$$

to the existing facilities is to be minimized (Drezner 2011). More general variations of the objective function of a location problem include ordered median location problems (Eiselt and Laporte 1995; Nickel and Puerto 2005), obnoxious facility location (Hansen et al. 1981; Drezner and Wesolowsky 1991), competitive facility location (Drezner 2009; Drezner and Eiselt 2002; Eiselt 2011; Drezner 2014; Eiselt et al. 2015), and various covering (García and Marín 2015; Snyder 2011) and equity models (Berman and Kaplan 1990; Drezner and Drezner 2007).

Also different types of restrictions have been studied, e.g., that the new point must not lie in a forbidden set (Hamacher and Nickel 1995), or there are barriers to travel (Klamroth 2002). Different approaches are necessary if one does not locate a new point x, but a line (Wesolowsky 1975), a circle (Drezner et al. 1996; Brimberg et al. 2009), a line segment, or any other dimensional structure. We refer to Díaz-Bánez et al. (2004) or Schöbel (2015) for surveys on the topic. The most prominent and very well researched generalization looks for a set of p new facilities to be located instead of just a single one. It is known as the p-median or the p-center problem. Both have many applications, and are treated in many publications, see, e.g., Brimberg et al. (2008), Mladenović et al. (2007) and the references therein.

Finally, the space may be varied from $\mathbb{R}^2$ to higher dimensions $\mathbb{R}^n$. Only few results are known here, straightforward generalizations are for the Manhattan and also for the squared Euclidean distance. Results and approaches for arbitrary norms in higher dimensions are known mainly for the location of hyperplanes (Martini and Schöbel 1998).

What has to the best of the author's knowledge not been studied so far are *integer* location problems restricting the location of the new facility to integer coordinates, i.e., requiring that $x \in \mathbb{Z}^n$. This changes the space for the new location from the continuous plane $\mathbb{R}^2$ to the grid points $\mathbb{Z}^2$. In this paper we mainly discuss *integer median location problems*:

Given a set $A_1, \ldots, A_m \in \mathbb{R}^2$ of m demand points in the plane with weights $w_m \geq 0, i = 1, \ldots, m$ and a distance measure $d : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$, find a new point $x \in \mathbb{Z}^2$ which minimizes the sum of weighted distances to the demand points, i.e.,

$$\text{(IL)} \quad \min \left\{ f(x) : x \in \mathbb{Z}^2 \right\}. \tag{5.2}$$

Generalizations of integer median location problems to other location problems (as described above) will be mentioned where appropriate.

Throughout the paper we use the following notation: Given an integer location problem

$$\text{(IL)} \quad \min \left\{ f(x) : x \in \mathbb{Z}^2 \right\},$$

its continuous relaxation

$$\text{(L)} \quad \min \left\{ f(x) : x \in \mathbb{R}^2 \right\}$$

is called its *corresponding* continuous location problem.

There are different reasons and applications why integer location problems are interesting. First, the new facilities might be restricted to be points on a grid, for example, due to accuracy reasons, or because the new facility should be built at a crossing (e.g. in Manhattan), or since a board can only be pinned at certain points. Such reasons may directly come from the application at hand. There is also another reason: optimal integer locations come in useful for solving robust integer optimization problems. This will be described in Sect. 5.2.

Note that integer location problems are also interesting from a theoretical point of view. It is well known that (general) integer optimization problems are harder to solve than continuous optimization problems. This is in general also the case for integer location problems, but as we will see in the following, there are integer location problems which can still be solved in the same time complexity as their corresponding continuous location problems. We believe that the structure which allows to solve integer location problems can also be exploited for solving more general integer programming problems.

The remainder of the paper is structured as follows. In Sect. 5.2 we start with a brief excursion to robust optimization giving another motivation why integer location problems are reasonable objects to be studied. We then discuss different approaches on how to tackle integer location problems: In Sect. 5.3 we first investigate the construction of a finite dominating (candidate) set for integer location problems whose corresponding continuous location problems are piecewise quasiconcave on convex cells and in Sect. 5.4 we use the sublevel sets of the corresponding continuous location problem and investigate in which cases their structure helps to find a solution to the integer location problem. We illustrate both approaches on the integer median location problem with rectangular distance ℓ_1, but also show for which other location problems these approaches may be applied. In Sect. 5.5 we finally propose a big-square-small-square method as a general algorithmic scheme for solving integer location problems. The paper is ended with some conclusions and suggestions for further research in Sect. 5.6.

5.2 Application in Robust Integer Optimization

In robust optimization, we consider optimization problems in which some (or all) parameters are uncertain or unknown. This means, we do not have an optimization problem with one fixed and given parameter set, but a different parameter set for each scenario $\xi \in \mathcal{U}$ that may occur. $\mathcal{U}$ is called the *uncertainty set* and contains all scenarios which should be taken into account. In most applications, $\mathcal{U}$ is an infinite set.

Taking the uncertainty into account, the optimization problem under consideration is specified as

$$P(\xi) \quad \min\{f_\xi(x) : x \in F_\xi\},$$

showing that both the objective function $f_\xi : \mathbb{R}^n \to \mathbb{R}$ and the feasible set $F_\xi \subseteq \mathbb{R}^n$ may depend on the (unknown) scenario $\xi \in \mathcal{U}$. The goal of robust optimization is to find a solution which is "good" for all scenarios that may occur. There exist many robustness concepts, each of them trying to define what "good" means in this context. In the conservative approach of *strict* or *minmax robustness* one looks for a solution which is feasible for all scenarios $\xi \in \mathcal{U}$, and best possible in the worst case. Such a solution can be found by solving the *robust counterpart* given as

$$\min\{\max_{\xi \in \mathcal{U}} f_\xi(x) : x \in F_\xi \text{ for all } \xi \in \mathcal{U}\},$$

see Ben-Tal et al. (2009). More recent concepts argue that the solution need not be good for all scenarios as long as it may be recovered (or repaired) quickly in the moment when the real scenario becomes known. This idea was independently proposed by Liebchen et al. (2009) and Erera et al. (2009). Unfortunately, finding such a recovery-robust solution is in most cases hard.

However, the deterministic problem $P(\xi)$ can usually be solved, i.e., an algorithm for solving the optimization problem for a *fixed* scenario $\xi \in \mathcal{U}$ is known. In Goerigk and Schöbel (2011, 2014) and Carrizosa et al. (2017), it is shown how such an algorithm can be used to determine a recovery-robust solution to the uncertain problem. The approaches developed in these papers basically work as follows.

Step 1. (Sampling)
In the first step, many of the problems

$$P(\xi) \quad \min\{f_\xi(x) : x \in F_\xi\}$$

are solved for fixed scenarios from a large, but finite set of sampled scenarios $\overline{\mathcal{U}} \subseteq \mathcal{U}$ (maybe even for the whole set $\mathcal{U}$ if it is finite and not too large). For each scenario $\xi \in \overline{\mathcal{U}}$ one obtains a feasible solution $x(\xi) \in \mathbb{R}^n$.

Step 2. (Finding a representative point for the sample)
In the second step, a location problem is solved. It takes the sampled solutions $x(\xi), \xi \in \overline{\mathcal{U}}$ as demand points, and looks for a new facility $x \in \mathbb{R}^n$ which is as close as possible to the given solutions $x(\xi)$. The distance $d(x, x(\xi))$ represents the costs to recover a solution x to a solution $x(\xi)$ and is defined such that it fits the application at hand. The demand points can even be weighted if a probability distribution for the scenario set $\mathcal{U}$ is known.

In the common situation that the problem $P(\xi)$ is a discrete or an integer location problem, it is required that the resulting solution x is integer, i.e., $x \in \mathbb{Z}^n$. In Step 2 of the procedure we are hence left with solving a location problem, e.g., the following median location problem (as in Goerigk and Schöbel 2011, 2014),

$$\min\{\sum_{\xi \in \overline{\mathcal{U}}} w_\xi d(x, x(\xi)) : x \in \mathbb{Z}^n\}.$$

Denoting the sampled scenarios $\overline{\mathcal{U}} = \{\xi_1, \ldots, \xi_m\}$ and defining demand points $A_i := x(\xi_i), i = 1, \ldots, m$ the problem to be solved in Step 2 turns out to be an integer location problem (IL) in $\mathbb{Z}^n$. In Carrizosa et al. (2017) not median, but center location problems are considered in Step 2.

5.3 Finding a Finite Dominating Set

For many location problems, finite dominating sets are known for a long time. A *finite dominating set* for a location problem (L) consists of a finite set of points (also called candidates) which contains an optimal solution to (L). Even if finite dominating sets usually have little algorithmic consequences they help understanding the structure of optimal locations and provide interesting properties.

Based on the *Hakimi property*, the probably best known finite dominating set is for median *network* location problems: In a network location problem, we have given a network $G = (V, E)$ where V is the set of demand points, and we are allowed to locate the new facilities anywhere along the edges. For the 1-median problem, Hakimi (1964, 1965) has shown that there always exists an optimal solution which is a node. This infers immediately that the set of nodes on the network is a finite dominating set for the 1-median problem, and hence, also for p-median problems on networks.

For planar location problems, a prominent finite dominating set is for the planar median location problem with rectangular distance ℓ_1: it is known that there exists an optimal solution which lies on the intersection of a vertical and a horizontal line, both passing through one of the existing facilities. This also holds for the planar p-median location problem: there exists a solution in which all p facilities lie on such intersection points. The property was also generalized to planar median problems with block norms or polyhedral gauges as distance measures (Durier and Michelot 1985): Here, the fundamental directions of the block norm or the gauge replace the vertical and horizontal lines of the ℓ_1 norm. There also exist finite dominating sets for location problems with restricted sets, e.g., in Hamacher and Nickel (1995) for restricted median problems with block norms and in Hamacher and Schöbel (1997) for center problems with Euclidean distance. Also for planar median line location problems, i.e., finding the location of a line minimizing the sum of distances to a set of given demand points, a finite dominating set is known. Namely, the set of all lines passing through two of the existing facilities are a finite dominating set whenever a norm is chosen as distance (Schöbel 1998). The property can be extended to hyperplanes and even to center objective functions (Martini and Schöbel 1998).

We now discuss finite dominating sets for (IL). The basic property for deriving a finite dominating set for continuous location problems (L) is quasiconcavity of the objective function on a *cell* structure of convex cells. In the case of planar median location problems with a block norms γ_B, these cells are defined by the grid of fundamental directions of γ_B, see Fig. 5.2 as an illustration for the rectangular distance (with horizontal and vertical lines as fundamental directions of the ℓ_1 block norm). For location problems (L) which admit such a cell structure, we can construct a finite dominating set for the integer case as follows:

Theorem 5.1 *Let (L) be a location problem which can be decomposed into a finite number of polyhedral cells $\mathscr{C}$ on which the objective function is quasiconcave. Then a finite dominating set for (IL) is given by*

$$\bigcup_{C \in \mathscr{C}} \text{ext}(\text{conv}(C \cap \mathbb{Z}^2)),$$

where $\text{conv}(A)$ *denotes the convex hull of a set* $A \subseteq \mathbb{R}^2$ *and* $\text{ext}(A)$ *denotes the extreme points of a convex set* $A \subseteq \mathbb{R}^2$.

Proof Let $S = C \cap \mathbb{Z}^2$ be the set of feasible points in a cell C, then by quasiconcavity of f we have that

$$\begin{aligned} \min\{f(x) : x \in S\} &= \min\{f(x) : x \in \text{conv}(S)\} \\ &= \min\{f(x) : x \in \text{ext}(\text{conv}(S))\} \end{aligned}$$

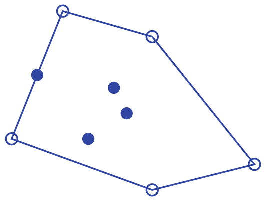


Fig. 5.1 The extreme points (not filled) of the convex hull of feasible points (filled and unfilled points)

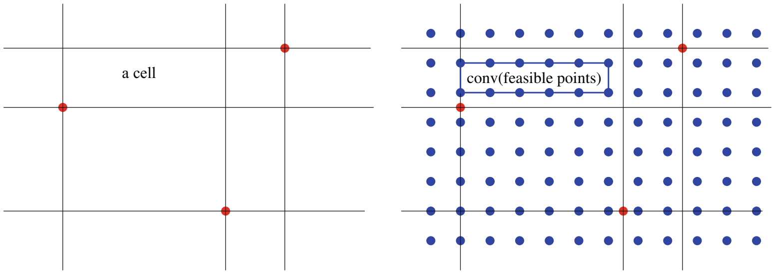


Fig. 5.2 *Left:* The 12 polyhedral cells for a planar median location problem with rectangular distance and three existing facilities (red balls). *Right:* The convex hull of the set of feasible points (blue balls) in one of the cells

since it is known that a quasiconcave function attains a minimum at an extreme point $\text{ext}(S)$ of a convex set, and that $\text{ext}(\text{conv}(S)) \subseteq S$. The situation is shown in Fig. 5.1. Moreover, due to Minkowski's theorem on representation of polyhedra (see, e.g., Nemhauser and Wolsey 1988) it is known that the extreme points of $\text{conv}(S)$ are a finite set.

A finite dominating set can now be derived whenever the set of extreme points of the convex hull of the feasible points within each cell can be determined. This works well for integer median location problems with polyhedral gauges and is shown exemplarily for the rectangular distance ℓ_1 and the Chebyshev-distance ℓ_∞ as illustrated below. Note that in this case the objective function is linear (and hence quasiconcave).

5.3.1 Integer Median Location Problems with Rectangular Distance

For the rectangular distance ℓ_1 and a location problem with m demand points, the $O(m^2)$ cells are known to be rectangles bounded by horizontal and vertical lines through the existing demand points, see the left side of Fig. 5.2.

The convex hull of the integer points in such a cell C is a smaller rectangle contained in C as shown exemplarily on the right side of Fig. 5.2. Its extreme points are its four vertices. These are the candidates see Fig. 5.3. They can be determined

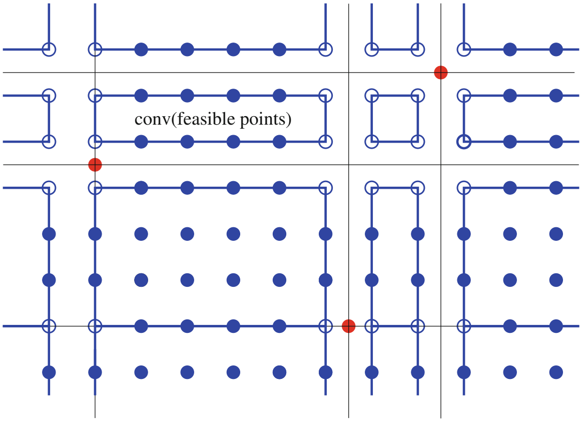


Fig. 5.3 The candidates for the integer median location problem with rectangular distance (the corresponding continuous location problem is shown in Fig. 5.2) are depicted as unfilled circles. The picture also shows the convex hulls of feasible points for each of the 12 cells

by taking the points in the finite dominating set of the continuous problem and rounding their coordinates up and down, i.e., for each point $x = (x_1, x_2) \in \mathbb{R}^2$ which is a candidate for the continuous location problem, we receive at most four candidates for the integer location problem, namely,

$$(\lfloor x_1 \rfloor, \lfloor x_2 \rfloor), (\lfloor x_1 \rfloor, \lceil x_2 \rceil), (\lceil x_1 \rceil, \lfloor x_2 \rfloor), (\lceil x_1 \rceil, \lceil x_2 \rceil).$$

Hence, the number of candidates we receive is of order $O(m^2)$.

Note that the finite dominating set constructed for the integer median location problem with rectangular distance also is a finite dominating set for the integer p-median problem with rectangular distance (since in an optimal solution to the integer p-median location problem each new facility is a solution of the integer median location problem with respect to a subset of the demand points).

5.3.2 Integer Median Location Problems with Chebyshev Distance

For the Chebyshev distance ℓ_∞ the $O(m^2)$ cells are known to be rectangles shifted by 45°. The convex hull of the set of feasible points $S = C \cap \mathbb{Z}^2$ within a cell C is illustrated in Fig. 5.4.

Figure 5.5 shows how the extreme points of conv(S) can be constructed: These are the points first met by the convex hull conv(S) if moving a horizontal or a vertical line towards the cell from the outside. We obtain at most 8 candidates per cell (unfilled circles) resulting in $O(m^2)$ candidates.

As for the rectangular distance, the finite dominating set constructed for the integer median location problem with Chebyshev distance is also a finite dominating set for the integer p-median location problem with Chebyshev distance.

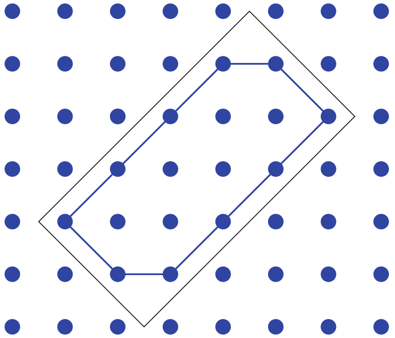


Fig. 5.4 The convex hull conv(S) of the set of feasible points $S = C \cap \mathbb{Z}^2$ in a cell C of the integer median location problem with Chebyshev distance

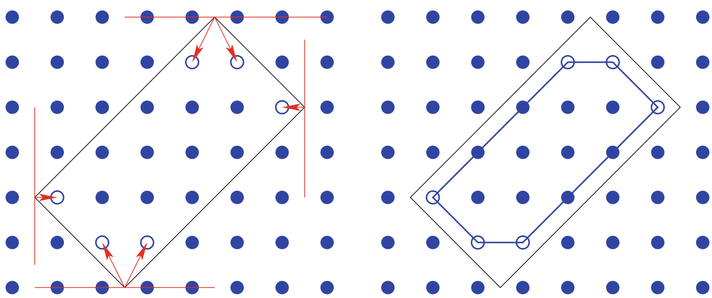


Fig. 5.5 The unfilled circles are the candidates of the depicted cell for the integer median location problem with Chebyshev distance. *Left:* construction of the candidates by moving vertical and horizontal lines towards the feasible points in the cell. *Right:* the candidates are the extreme points of the convex hull of the set of feasible points

5.3.3 Integer Median Location Problems with Polyhedral Gauges

For polyhedral gauges this approach can also be used where the fundamental directions of the gauge γ_B are needed to construct the extreme points of conv$(C \cap \mathbb{Z}^2)$ for the cells $C \in \mathscr{C}$. The number of cells and the number of extreme points within a cell both grow with the number G of fundamental directions of the gauge γ_B. Overall, the cardinality of the finite dominating set hence is $O(G^2m^2)$. The finite dominating set is also valid for the planar p-median location problem with the same polyhedral gauge γ_B.

5.3.4 Other Location Problems

We remark that Theorem 5.1 and hence the proposed approach for constructing a finite dominating set also works for higher dimensions as well as for other types of facilities to be located. For example, in the case of line location problems, we also receive a polyhedral cell structure in the called dual space (in which each line is mapped to a point) with a quasiconcave (non-linear) objective function (see Schöbel 1999) and Theorem 5.1 can also be applied.

In the case that the objective function is linear on each of the polyhedral cells it is also possible to use integer linear programming on each of the cells. For planar location problems only two variables are needed; for this special case several fast algorithms for integer linear programming exist (see, e.g., Feit 1984; Eisenbrand and Rote 2001) that may be used. If $r(m)$ is the runtime of such an algorithm for minimizing over a cell for a location problem with m demand point, the overall runtime for solving an integer location problem with polyhedral cells $\mathscr{C}$ adds up to $\mathrm{O}(r(m)|\mathscr{C}|)$; for an integer median location problem with a block norm with G fundamental directions we hence receive a complexity of $\mathrm{O}(r(m)Gm^2)$.

5.4 Using the Structure of the Sublevel Sets

For some integer optimization problems, an optimal solution can be found by rounding a solution of their continuous relaxations (up or down). If this is the case, the so-called *rounding property* holds, see Hübner and Schöbel (2014). We adapt the notation given there to integer location problems:

Notation 5.2 (Compare Hübner and Schöbel (2014)) *The integer location problem (IL) has the* rounding property *if for any optimal solution $x = (x_1, x_2)$ to its corresponding continuous location problem (L) there exists an optimal solution $x^* = (x_1^*, x_2^*)$ to (IL) such that*

$$x_1^* \in \{\lfloor x_1 \rfloor, \lceil x_1 \rceil\} \textit{ and } x_2^* \in \{\lfloor x_2 \rfloor, \lceil x_2 \rceil\}.$$

The sublevel sets

$$L_{\leq}^{f}(z) := \{x \in \mathbb{R}^2 : f(x) \leq z\}$$

of the location problem (L) can be used to check if an integer optimization problem admits the rounding property. In Hübner and Schöbel (2014), several conditions have been identified in which this is the case. The one which can be applied to location problems needs the definition of a box spanned by two points:

Let $a = (a_1, a_2), b = (b_1, b_2) \in \mathbb{R}^2$. Then

$$\begin{aligned} B(a, b) &:= [\min\{a_1, b_1\}, \max\{a_1, b_1\}] \times [\min\{a_2, b_2\}, \max\{a_2, b_2\}] \\ &= \{x : \ell_1(a, x) + \ell_2(x, b) = \ell_1(a, b)\}. \end{aligned}$$

Equivalently, $B(a, b)$ is the box with the four vertices $(a_1, a_2), (a_1, b_2), (b_1, a_2), (b_1, b_2)$.

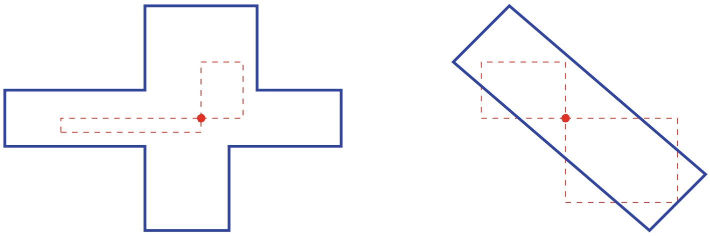


Fig. 5.6 *Left:* A set which is cross-shaped with respect to x_0 (red point). *Right:* A convex set which is not cross-shaped with respect to any point

Notation 5.3 (Hübner and Schöbel (2014)) *A set $M \subseteq \mathbb{R}^n$ is called* cross-shaped *w.r.t. $x_0 \in M$ if for any $y \in M$ the box $B(x_0, y) \subseteq M$.*

Euclidean balls, axis-parallel rectangles and axis-parallel ellipsoids are convex and cross-shaped sets. However, sets which have the shape of a cross (as shown in Fig. 5.6, left side) are cross-shaped, but not convex. On the other hand, convexity does not imply cross-shapedness, see Fig. 5.6, right side.

We use that box-shapedness of the sublevel sets guarantees the rounding property:

Theorem 5.4 (Hübner and Schöbel (2014)) *Let f be the objective function of a location problem (L), and let x be an optimal solution to (L). If the sublevel sets of f are cross-shaped w.r.t. the solution x, then (IL) has the rounding property.*

In the following we identify a few planar location problems which have cross-shaped level sets.

5.4.1 Integer Median Location Problems with Rectangular Distance

In this section we show that the sublevel sets of the planar median location problem with rectangular distance are cross-shaped.

Lemma 5.5 *Let x be an optimal solution to the planar median location problem with rectangular distance ℓ_1. Then for any $z \in \mathbb{R}$ the sublevel set $L^f_{\le}(z)$ is a cross-shaped set w.r.t. the solution x.*

Proof Denote the demand points $A_i = (a_{i1}, a_{i2})$ for $i = 1, \ldots, m$. If the sublevel set $L^f_{\le}(z)$ is empty, there is nothing to show. Otherwise let $y \in L^f_{\le}(z)$. We want to show that $B(x, y) \subseteq L^f_{\le}(z)$, i.e., that for any point $q \in B(x, y)$ we have that $f(q) \le z$.

To this end, take any point $q = (q_1, q_2) \in B(x, y)$. I.e., q_1 lies between x_1 and y_1 and q_2 lies between x_2 and w_2. We obtain that

$$
\begin{aligned}
f(q) &= \sum_{i=1}^{m} w_i \ell_1(q, A_i) \\
&= \sum_{i=1}^{m} w_i |q_1 - a_{i1}| + \sum_{i=1}^{m} w_i |q_2 - a_{i2}| \\
&=: f_1(q_1) + f_2(q_2).
\end{aligned}
$$

Note that f_1 is a convex function and $q_1 = \lambda_1 x_1 + (1 - \lambda_1) y_1$ for some $0 \leq \lambda_1 \leq 1$ is a convex combination of x_1 and y_1. Furthermore, $f_1(x_1) \leq f_1(y_1)$ since x is an optimal solution to (L). We hence have that

$$
\begin{aligned}
f_1(q_1) &= f_1(\lambda_1 x_1 + (1 - \lambda_1) y_1) \\
&\leq \lambda_1 f_1(x_1) + (1 - \lambda_1) f_1(y_1) \\
&\leq \lambda_1 f_1(y_1) + (1 - \lambda_1) f_1(y_1) \;=\; f_1(y_1).
\end{aligned}
$$

Analogously, $f_2(q_2) \leq f_2(y_2)$ and together we obtain that

$$
f(q) = f_1(q_1) + f_2(q_2) \leq f_1(y_1) + f_2(y_2) = f(y),
$$

i.e., $q \in L_{\leq}^{f}(z)$.

Together with Theorem 5.4 we now conclude:

Theorem 5.6 *The integer median location problem with rectangular distance* ℓ_1 *has the rounding property.*

This theorem can be used algorithmically as follows:

Let an integer median location problem (IL) with rectangular distance ℓ_1 be given and let (x_1, x_2) be an optimal solution to its corresponding planar median location problem. Then an optimal solution to (IL) is contained in

$$
(\lfloor x_1 \rfloor, \lfloor x_2 \rfloor), (\lfloor x_1 \rfloor, \lceil x_2 \rceil), (\lceil x_1 \rceil, \lfloor x_2 \rfloor), (\lceil x_1 \rceil, \lceil x_2 \rceil).
$$

I.e., we only have to compute the objective function value of the four possible points and choose the best of them. The integer location problem (IL) can hence be solved in the same time complexity as its continuous counterpart (L).

5.4.2 Integer Median Location Problems with Chebyshev Distance

Note that the rounding property does not hold for median location problems with the Chebyshev-norm as the example depicted in Fig. 5.7 demonstrates. The figure also shows that the sublevel sets of the problem need not to be cross-shaped.

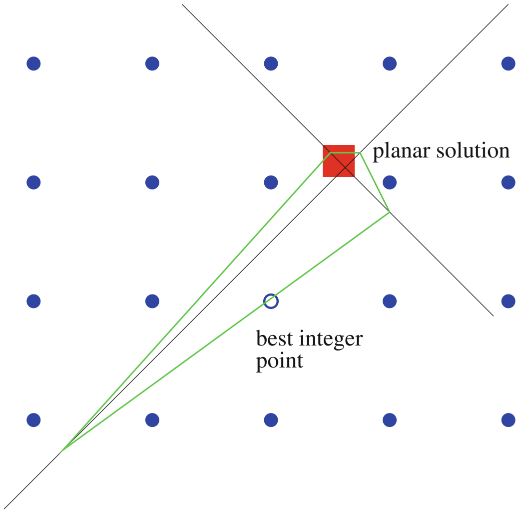


Fig. 5.7 The rounding property does not hold for integer median location problems with Chebyshev distance. The optimal solution to (L) is depicted as red point, the best integer solution as the unfilled circle. It can also be seen that the sublevel set (which intersects four different cells) is not cross-shaped

5.4.3 Integer Median Location Problems with Squared Euclidean Distance

The squared Euclidean distance is not piecewise linear, not even quasiconcave, hence it is not possible to derive a finite dominating set as in the first approach. However, integer location problems with the squared Euclidean distance turn out to be solvable easily in linear time by using the structure of the sublevel sets again. Note that the optimal solution to a location problem with squared Euclidean distance is unique and given as the center of gravity of its (weighted) demand points (White 1971).

Lemma 5.7 *Let x be the optimal solution to the planar median location problem with squared Euclidean distance ℓ_2^2. Then for any $z \in \mathbb{R}$ the sublevel set $L_{\leq}^{f}(z)$ is a cross-shaped set w.r.t. x.*

Proof Denote the demand points $A_i = (a_{i1}, a_{i2})$ for $i = 1, \ldots, m$. If the sublevel set $L_{\leq}^{f}(z)$ is empty, there is nothing to show. Otherwise let $y \in L_{\leq}^{f}(z)$. As in the proof of Lemma 5.5 we show that $B(x, y) \subseteq L_{\leq}^{f}(z)$.

Again, if the sublevel set is not empty, take any point $q = (q_1, q_2) \in B(x, y)$. Separability of the objective function in the squared Euclidean case gives that

$$\begin{aligned} f(q) &= \sum_{i=1}^{m} w_i \ell_2^2(q, A_i) \\ &= \sum_{i=1}^{m} w_i (q_1 - a_{i1})^2 + \sum_{i=1}^{m} w_i (q_2 - a_{i2})^2 \\ &=: \tilde{f}_1(q_1) + \tilde{f}_2(q_2), \end{aligned}$$

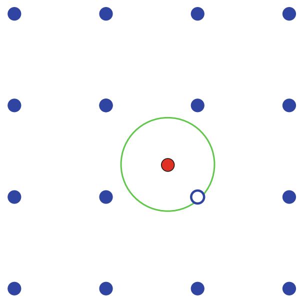


Fig. 5.8 The sublevel sets about the center of gravity (red point) for planar median location problems with squared Euclidean distance are Euclidean balls. The integer solution in the example is the unfilled circle

and $\tilde{f}_1, \tilde{f}_2$ are both convex functions which obtain their (unique) minima in x_1, and x_2, respectively. As in the proof of Lemma 5.5 we conclude

$$\tilde{f}_1(q_1) \leq \tilde{f}_1(y_1), \quad \text{and}$$
$$\tilde{f}_2(q_2) \leq \tilde{f}_2(y_2)$$

and obtain the required result:

$$f(q) = \tilde{f}_1(q_1) + \tilde{f}_2(q_2) \leq \tilde{f}_1(y_1) + \tilde{f}_2(y_2) = f(y).$$

Figure 5.8 shows an example of an integer median location problem with squared Euclidean distance. The figure shows its (integer) solution as well as a sublevel set of its corresponding continuous location problem about the center of gravity of the demand points.

Theorem 5.8 *The planar median location problem with squared Euclidean distance ℓ_2^2 has the rounding property.*

This theorem can even be strengthened as follows: In Hübner and Schöbel (2014) it is shown that box-shaped sublevel sets which are also coordinate axially symmetric admit the *strong* rounding property as defined below.

Notation 5.9 (Compare Hübner and Schöbel (2014)) *An integer location problem (IL) has the* strong rounding property *if for any optimal solution $x = (x_1, x_2)$ to (L) there exists an optimal solution $x^* = (x_1^*, x_2^*)$ to (IL) with*

$$x_1^* = round(x_1) \text{ and } x_2^* = round(x_2),$$

where $round(x)$ for some real number x means to round x to its closest integer using any fixed rule, e.g. the round half up rule in order to break ties.

Since the sublevel sets of the planar median location problem with squared Euclidean distance are Euclidean balls (see Hamacher 1995), they are coordinate axially symmetric. We hence conclude:

Corollary 5.10 *The planar median location problem with squared Euclidean distance ℓ_2^2 has the strong rounding property.*

Since the strong rounding property holds, the approach for solving integer location problems with squared Euclidean distance ℓ_2^2 is even easier as for the case of the rectangular norm ℓ_1: We compute the solution to the corresponding continuous problem, i.e., we determine its center of gravity $x \in \mathbb{R}^2$ by averaging the given demand points. Rounding both components of x to their closest integers gives the solution to the integer location problem.

5.4.4 Other Location Problems

As the example with the Chebyshev norm shows, the rounding property does not hold for integer location problems with general polyhedral gauges. It is also not satisfied for integer median location problems with the Euclidean distance.

On the other hand, the rounding property for the rectangular and for the squared Euclidean distance can easily be generalized to high-dimensional location problems, again using separability of the objective function into n convex functions as for the two-dimensional case. The rounding property can also be shown for center problems with the Euclidean distance ℓ_2 (also here the sublevel sets are cross-shaped), i.e., also in this case the corresponding integer location problem be solved by finding an optimal solution, e.g., by the algorithm of Elzinga-Hearn, and rounding its components up or down.

5.5 A Big-Square-Small-Square Approach

The big-square-small-square (BSSS) method is a geometric branch-and-bound procedure which has been successfully applied to location theory since the paper (Hansen et al. 1981). The method is interesting in itself, generalizations and the construction of good bounds is treated, e.g., in Plastria (1992), Drezner (2007), Blanquero and Carrizosa (2009), and Scholz and Schöbel (2010). A prominent modification is the big-triangle-small-triangle (BTST) method by Drezner and Suzuki (2004). BSSS has been adapted to more dimensional problems to a big-cube-small-cube method (BCSC) (Schöbel and Scholz 2010), generalized to mixed-integer optimization (Schöbel and Scholz 2014) and to multiobjective problems (Scholz 2010, 2011; Niebling and Eichfelder 2018). The method has also been applied to searching along segments (Berman et al. 2011) and recently, for searching on arcs, called big-arc-small-arc method (BASA) (Drezner et al. 2018).

The basic idea of BSSS is the following: Starting with a bounded box as feasible set, in each iteration, a box is chosen and decomposed into smaller boxes. On each box, a lower bound on the optimal objective value is computed. The bounds and the current best objective value are then used for pruning boxes which cannot contain an optimal solution. The remaining boxes are further decomposed. We now show how a BSSS algorithm can be refined in order to solve integer location problems (IL). To this end, assume that a BSSS algorithm for the continuous version (L) is given:

Input: (L) $\min\{f(x) : x \in \mathbb{R}^2\}$ and starting box B^0, accuracy $\epsilon > 0$.
Step 1. Initialization: List $:= \{B^0\}$, $\overline{z} := \infty$, $x^{inc} := \emptyset$.
Step 2. Stopping criterion: If List $= \emptyset$, Stop. Output: $x^* := x^{inc}$, $z^* := \overline{z}$.
Step 3. Selection:

3.1 Choose a box $\tilde{B} = B(\tilde{a}, \tilde{b}) \in$ List; List := List $\setminus \{\tilde{B}\}$.
3.2 Compute the center of the box $x^{\tilde{B}} := \frac{\tilde{a}+\tilde{b}}{2}$
3.3 Compute a lower bound LB on the box $\tilde{B}$.

Step 4. Pruning:

4.1. If $f(x^{\tilde{B}}) < \overline{z}$ set $\overline{z} := f(x^{\tilde{B}})$ and $x^{inc} := x^{\tilde{B}}$.
4.2. If $LB > \overline{z}$, goto Step 2.
4.3. If $f(x^{\tilde{B}}) < LB + \epsilon$, goto Step 2.

Step 5. Branching: Decompose $\tilde{B}$ in four smaller boxes B^1, B^2, B^3, B^4 with $\tilde{B} = \bigcup_{k=1}^{4} B^k$. Set List := List $\cup \{B^1, B^2, B^3, B^4\}$. Goto Step 2.

Note that the objective function value of (IL) is larger or equal to the objective function value of the corresponding location problem (L), i.e.,

$$z := \min\{f(x) : x \in \mathbb{R}^2\} \leq \min\{f(x) : x \in \mathbb{Z}^2\} =: z^*.$$

In particular, a lower bound for (L) is also a lower bound for (IL). This means, we can easily adapt BSSS to (IL) by using the same bounds as for the continuous counterpart. We can even strengthen the bounds and the boxes. The modifications we need are described below.

In Step 3.2 we need to compute an integer point instead of the center. This is done by

3.2 Compute the center $x^{\tilde{B}} := \frac{\tilde{a}+\tilde{b}}{2}$. Check if any of the four points

$$(\lfloor x_1^{\tilde{B}} \rfloor, \lfloor x_2^{\tilde{B}} \rfloor), (\lfloor x_1^{\tilde{B}} \rfloor, \lceil x_2^{\tilde{B}} \rceil), (\lceil x_1^{\tilde{B}} \rceil, \lfloor x_2^{\tilde{B}} \rfloor), (\lceil x_1^{\tilde{B}} \rceil, \lceil x_2^{\tilde{B}} \rceil)$$

lies in $\tilde{B}$. If yes, set $x^{\tilde{B}}$ as this point. Otherwise the box does not contain any integer point and can be discarded, i.e., goto Step 2.

In Step 3.3 we can strengthen the lower bound LB by using integer rounding (Nemhauser and Wolsey 1988). I.e., if the computation of the bound only involves integer operations, then we can round LB up to the next integer, i.e., we receive $\lceil LB \rceil$ as stronger bound.

Concerning the pruning in Step 4, note that in the modified Step 3.2 we already discarded boxes which do not contain any integer point. We can furthermore choose a small integer number ρ (e.g., $\rho = 4$) and prune boxes by optimality if they contain less than ρ integer points. To this end, we can add the following step at the beginning of Step 4:

4.0. If $|\tilde{B} \cap \mathbb{Z}^2| \leq \rho$ evaluate the objective function for each of the ρ integer points and choose the best of them. Let the corresponding optimal solution of the box be $x^{\tilde{B}}$.

Note that Step 4.0 is not performed if the box contains more than ρ integer points. In this case, $x^{\tilde{B}}$ is the rounded center of the box computed in Step 3.2.

Finally, when decomposing a large box $B = B(a, b)$ (with $a = (a_1, a_2)$, $b = (b_1, b_2)$ and $a_1 \leq b_1, a_2 \leq b_2$) into smaller boxes this can be done by taking the center q of the box, i.e., $q_i := \frac{a_i + b_i}{2}$ for $i = 1, 2$ and defining the four new boxes as

$$\begin{aligned}
B^1 &:= B((a_1, \lfloor q_1 \rfloor), (a_2, \lfloor q_2 \rfloor)),\\
B^2 &:= B((a_1, \lceil q_2 \rceil), (\lfloor q_1 \rfloor, b_2)),\\
B^3 &:= B((\lceil q_1 \rceil, a_2), (b_1, \lfloor q_2 \rfloor)),\\
B^4 &:= B((\lceil q_1 \rceil, \lceil q_2 \rceil), (b_1, b_2)),
\end{aligned}$$

see Fig. 5.9 as illustration. Note that for $q \notin \mathbb{Z}^2$ the sum of the areas of the four new boxes is strictly smaller than the area of the original box (which further strengthens BSSS for integer location problems compared to the continuous version).

The algorithm converges since the number of feasible points is finite and the new boxes generated in Step 5 are strictly smaller than the box which is decomposed such that eventually all points are enumerated.

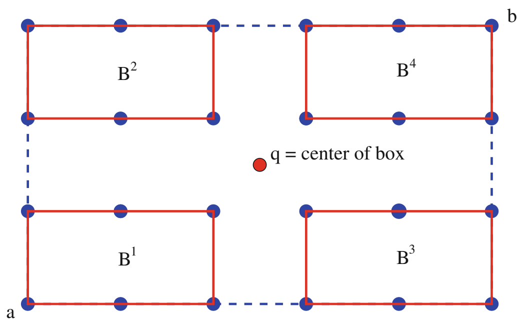


Fig. 5.9 decomposing a larger box (dashed line) into four smaller boxes (red lines) defined by the center of the box q (red point). Note that the sum of the areas of the smaller boxes is strictly smaller than the area of the original box

5.6 Conclusion and Further Research

In this paper we have introduced integer location problems. Apart from direct applications we have motivated the usage of integer location problems for robust optimization. We then have shown how a finite dominating set may be constructed and how sublevel sets may be used to establish a rounding property. We also propose a solution algorithm based on geometric branch-and-bound.

Research on integer location problems nevertheless is just at its beginning. In particular higher-dimensional problems are of interest, as well as location problems with restricted sets. Both extensions are important for finding recovery-robust solutions as sketched in Sect. 5.2, or for the ongoing topic of integrated optimization of interwoven systems as described in Klamroth et al. (2017).

We hope that this introductory paper triggers research towards integer location problems including integer median as well as integer center problems with different distance measures.

References

Ben-Tal, A., Ghaoui, L. E., & Nemirovski, A. (2009). *Robust optimization*. Princeton: Princeton University Press.

Berman, O., Drezner, Z., & Krass, D. (2011). Big segment small segment global optimization algorithm on networks. *Networks, 58*, 1–11.

Berman, O., & Kaplan, E. H. (1990). Equity maximizing facility location schemes. *Transportation Science, 24*, 137–144.

Blanquero, R., & Carrizosa, E. (2009). Continuous location problems and big triangle small triangle: Constructing better bounds. *Journal of Global Optimization, 45*(3), 389–402.

Brimberg, J., Hansen, P., Mladenović, N., & Salhi, S. (2008). A survey of solution methods for the continuous location allocation problem. *International Journal of Operations Research, 5*(1), 1–12.

Brimberg, J., Juel, H., & Schöbel, A. (2009). Locating a minisum circle in the plane. *Discrete Applied Mathematics, 157*, 901–912.

Brimberg, J., & Love, R. F. (1993). Global convergence of a generalized iterative procedure for the minisum location problem with l_p distances. *Operations Research, 41*, 1153–1163.

Carrizosa, E., Goerigk, M., & Schöbel, A. (2017). A biobjective approach to recovery robustness based on location planning. *European Journal of Operational Research, 261*, 421–435.

Díaz-Bánez, J. M., Mesa, J., & Schöbel, A. (2004). Continuous location of dimensional structures. *European Journal of Operational Research, 152*, 22–44.

Drezner, T. (2009). Competitive facility location. In *Encyclopedia of optimization* (2nd ed., pp. 396–401). Berlin: Springer.

Drezner, T. (2014). A review of competitive facility location in the plane. *Logistics Research, 7*, 1–12.

Drezner, T., & Drezner, Z. (2007). Equity models in planar location. *Computational Management Science, 4*, 1–16.

Drezner, T., Drezner, Z., & Schöbel, A. (2018). The Weber obnoxious facility location model: A big arc small arc approach. *Computers and Operations Research, 98*, 240–250.

Drezner, T., & Eiselt, H. A. (2002). Consumers in competitive location models. In Z. Drezner & H. W. Hamacher (Eds.), *Facility location: Applications and theory* (pp. 151–178). Berlin: Springer.

Drezner, Z. (Ed.). (1995). *Facility location: A survey of applications and methods*. New York: Springer.

Drezner, Z. (2007). A general global optimization approach for solving location problems in the plane. *Journal of Global Optimization, 37*(2), 305–319.

Drezner, Z. (2011). Continuous center problems. In H. A. Eiselt & V. Marianov (Eds.), *Foundations of location analysis* (pp. 63–78). Berlin: Springer.

Drezner, Z., Klamroth, K., Schöbel, A., & Wesolowsky, G. (2002). The Weber problem. In Z. Drezner & H. Hamacher (Eds.), *Facility location - applications and theory* (chap. 1, pp. 1–36). Berlin: Springer.

Drezner, Z., Steiner, S., & Wesolowsky, G. (1996). *On the circle closest to a set of points*. Technical report, California State University, Department of Management Science and Information Systems.

Drezner, Z., & Suzuki, A. (2004). The big triangle small triangle method for the solution of nonconvex facility location problems. *Operations Research, 52*, 128–135.

Drezner, Z., & Wesolowsky, G. O. (1980). Single facility ℓ_p distance minimax location. *SIAM Journal of Algebraic and Discrete Methods, 1*, 315–321.

Drezner, Z., & Wesolowsky, G. O. (1991). The Weber problem on the plane with some negative weights. *INFOR, 29*, 87–99.

Durier, R., & Michelot, C. (1985). Geometrical properties of the Fermat-Weber problem. *European Journal of Operational Research, 20*, 332–343.

Eiselt, H. A. (2011). Equilibria in competitive location models. In H. A. Eiselt & V. Marianov (Eds.), *Foundations of location analysis* (pp. 139–162). Berlin: Springer.

Eiselt, H. A., & Laporte, G. (1995). Objectives in location problems. In Z. Drezner (Ed.), *Facility location: A survey of applications and methods* (pp. 151–180). New York: Springer.

Eiselt, H. A., & Marianov, V. (Eds.). (2011). *Foundations of location analysis*. New York: Springer.

Eiselt, H. A., Marianov, V., & Drezner, T. (2015). Competitive location models. In G. Laporte, S. Nickel, & F. S. da Gama (Eds.), *Location science* (pp. 365–398). Berlin: Springer.

Eisenbrand, F., & Rote, G. (2001). Fast 2-variable integer programming. In *IPCO 2001: Integer programming and combinatorial optimization*. *Lecture notes in computer science* (vol. 2081, pp. 78–89).

Erera, A., Morales, J., & Savelsbergh, M. (2009). Robust optimization for empty repositioning problems. *Operations Research, 57*(2), 468–483.

Feit, S. (1984). A fast algorithm for the two-variable integer programming problem. *Journal of the Association for Computing Machinery, 31*(1), 99–113.

García, S., & Marín, A. (2015). Covering location problems. In G. Laporte, S. Nickel, & F. S. da Gama (Eds.), *Location science* (pp. 93–114). Berlin: Springer.

Goerigk, M., & Schöbel, A. (2011). A scenario-based approach for robust linear optimization. In *Proceedings of the 1st International ICST Conference on Practice and Theory of Algorithms in (Computer) Systems (TAPAS)*. *Lecture notes in computer science* (pp. 139–150). Berlin: Springer.

Goerigk, M., & Schöbel, A. (2014). Recovery-to-optimality: A new two-stage approach to robustness with an application to aperiodic timetabling. *Computers and Operations Research, 52*, 1–15.

Hakimi, S. (1964). Optimum locations of switching centers and the absolute centers and medians of a graph. *Operations Research, 12*(3), 450–459.

Hakimi, S. (1965). Optimum distribution of switching centers in a communication network and some related graph theoretic problems. *Operations Research, 13*(3), 462–475.

Hamacher, H. (1995). *Mathematische Lösungsverfahren für planare Standortprobleme*. Braunschweig: Vieweg.

Hamacher, H., & Drezner, Z. (Eds.). (2002). *Facility location. Applications and theory*. New York: Springer.

Hamacher, H., & Nickel, S. (1995). Restricted planar location problems and applications. *Naval Research Logistics, 42*(6), 967–992.

Hamacher, H., & Schöbel, A. (1997). A note on center problems with forbidden polyhedra. *Operations Research Letters, 20*, 165–169.
Hansen, P., Peeters, D., & Thisse, J.-F. (1981). On the location of an obnoxious facility. *Sistemi Urbani, 3*, 299–317.
Hübner, R., & Schöbel, A. (2014). When is rounding allowed in integer nonlinear optimization? *European Journal of Operational Research, 237*, 404–410.
Klamroth, K. (2002). *Single-facility location problems with barriers*. *Springer series on operations research*. New York: Springer.
Klamroth, K., Mostaghim, S., Naujoks, B., Poles, S., Purshouse, R., Rudolph, G., et al. (2017). Multiobjective optimization for interwoven systems. *Multi-Critieria Decision Analysis, 24*, 71–81.
Laporte, G., Nickel, S., & da Gama, F. S. (Eds.). (2015). *Location science*. Switzerland: Springer.
Liebchen, C., Lübbecke, M., Möhring, R. H., & Stiller, S. (2009). The concept of recoverable robustness, linear programming recovery, and railway applications. In R. K. Ahuja, R. Möhring, & C. Zaroliagis (Eds.), *Robust and online large-scale optimization*. *Lecture note on computer science* (vol. 5868). Berlin: Springer.
Love, R., Morris, J., & Wesolowsky, G. (1988). *Facilities location*. Amsterdam: North-Holland.
Martini, H., & Schöbel, A. (1998). Median hyperplanes in normed spaces — a survey. *Discrete Applied Mathematics, 89*, 181–195.
Mladenović, N., Brimberg, J., Hansen, P., & Moreno-Pérez, J. (2007). The p-median problem: A survey of metaheuristic approaches. *European Journal of Operational Research, 179*(3), 927–939.
Nemhauser, G. L., & Wolsey, L. A. (1988). *Integer and combinatorial optimization*. New York: Wiley.
Nickel, S., & Puerto, J. (2005). *Location theory: A unified approach*. Berlin: Springer.
Niebling, J., & Eichfelder, G. (2018). *A branch-and-bound based algorithm for nonconvex multiobjective optimization*. Technical report, Preprint-Series of the Institute for Mathematics, Technische Universität Ilmenau, Germany.
Plastria, F. (1992). GBSSS, the generalized big square small square method for planar single facility location. *European Journal of Operational Research, 62*, 163–174.
ReVelle, C., & Eiselt, H. (2005). Location analysis: a synthesis and survey. *European Journal of Operational Research, 165*, 1–19.
Schöbel, A. (1998). Locating least distant lines in the plane. *European Journal of Operational Research, 106*(1), 152–159.
Schöbel, A. (1999). *Locating lines and hyperplanes — Theory and algorithms*. *Applied optimization series* (vol. 25). Dordecht: Kluwer.
Schöbel, A. (2015). Location of dimensional facilities in a continuous space. In G. Laporte, S. Nickel, & F. S. da Gama (Eds.), *Location science* (chap. 7, pp. 63–103). Berlin: Springer.
Schöbel, A., & Scholz, D. (2010). The big cube small cube solution method for multidimensional facility location problems. *Computers and Operations Research, 37*, 115–122.
Schöbel, A., & Scholz, D. (2014). A solution algorithm for non-convex mixed integer optimization problems with only few continuous variables. *European Journal of Operational Research, 232*(2), 266–275.
Scholz, D. (2010). The multicriteria big cube small cube method. *TOP, 18*, 286–302.
Scholz, D. (2011). General further discarding tests in geometric branch-and-bound methods for non-convex multicriteria optimization problems. *Journal of Multi-Criteria Decision Analysis, 18*, 65–75.
Scholz, D., & Schöbel, A. (2010). The theoretical and empirical rate of convergence for geometric branch-and-bound methods. *Journal of Global Optimization, 48*(3), 473–495.
Snyder, L. (2011). Covering problems. In H. A. Eiselt & V. Marianov (Eds.), *Foundations of location analysis* (pp. 109–135). Berlin: Springer.
Thisse, J.-F., Ward, J. E., & Wendell, R. E. (1984). Some properties of location problems with block and round norms. *Operations Research, 32*(6), 1309–1327.

Ward, J., & Wendell, R. (1985). Using block norms for location modeling. *Operations Research, 33*, 1074–1090.

Wesolowsky, G. (1975). Location of the median line for weighted points. *Environment and Planning A, 7*, 163–170.

Wesolowsky, G. (1993). The Weber problem: history and perspectives. *Location Science, 1*(1):5–23.

White, J. (1971). A quadratic facility location problem. *American Institute of Industrial Engineers Transactions, 3*, 156–157.

Chapter 6
Continuous Location Problems

George O. Wesolowsky

6.1 Introduction

In this chapter I summarize many papers (out of 75) co-authored with Zvi mostly on continuous location models in the plane. Other topics that are described in other chapters include: production processes (Drezner et al. 1984; Drezner and Wesolowsky 1989e), optimal control (Drezner and Wesolowsky 1989f,c, 1991a,c, 1995b), and statistical methods (Drezner and Wesolowsky 1989a, 1990; Drezner et al. 1999).

Continuous location problems were some of the very earliest attempts at creating a body of knowledge relating to the most efficient ways of utilizing the location of facilities. These problems were shared among other disciplines such as mathematics, economics, and industrial engineering. Location problems also provided a firm base for more complex models and led to many problems of optimization which then could be applied in these other fields.

I first met Zvi when he became a Postdoctoral Fellow at McMaster University when he came to work with me. When he arrived, I set out to give him some ideas for location problems. I expected to have peace and quiet for weeks or months, but Zvi came back the next day with a sheaf of yellow note paper, on which he already solved all the problems I had proposed. I knew at that time that I had met someone special.

G. O. Wesolowsky (✉)
DeGroote School of Business, McMaster University, Hamilton, ON, Canada
e-mail: wesolows@mcmaster.ca

H. A. Eiselt, V. Marianov (eds.), *Contributions to Location Analysis*, International Series in Operations Research & Management Science 281,
https://doi.org/10.1007/978-3-030-19111-5_6

6.2 Minisum Models

In this section, models in which the objective is minimizing the sum of weighted distances between facilities and demand points are reviewed. This problem is commonly termed in the literature as the "Weber Problem" (Weber 1909).

Wesolowsky (1993) and Drezner et al. (2002a) provide a comprehensive review including historical background. The Weber problem has a long and convoluted history. Many players, from many fields of study, stepped on its stage. The problem seems simple, but is rich in possibilities. It has generated an enormous literature dating back to the seventeenth century. Kuhn (1967) provided historical background. It is usual to credit Pierre de Fermat (1601–1665) with proposing a basic form of the unweighted Weber problem based on three points. Other credits are given to Evangelista Torricelli (1608–1647), Battista Cavalieri (1598–1647), Thomas Simpson (1710–1761), Jacob Steiner (1796–1863), in addition to the more recent Alfred Weber (1868–1958).

As can be seen by perusing the references, some of the many names that have been used are: the Fermat problem, the generalized Fermat problem, the Fermat–Torricelli problem, the Steiner problem, the generalized Steiner problem, the Steiner–Weber problem, the Weber problem, the generalized Weber problem, the Fermat–Weber problem, the one median problem, the median center problem, the minisum problem, the minimum aggregate travel point problem, the bivariate median problem, and the spatial median problem.

The Weber problem is to find the best location X for a facility with the objective of:

$$\min_X \left\{ \sum_{i=1}^{n} w_i d_i(X) \right\} \tag{6.1}$$

where n is the number of demand points, w_i is the weight associated with demand point i, $d_i(X)$ is the distance between demand point i and the facility location X. The original application of the Weber problem was the location of a distribution center.

6.2.1 *Extensions to the Weber Problem*

Drezner and Wesolowsky (1978c, 1980c) solved the Weber problem when demand is originated in areas rather than demand points (Love 1972; Wesolowsky and Love 1971). This is a more realistic model when demand is generated in neighborhoods rather than at individual points.

Drezner and Wesolowsky (2000) investigated demand points clustered into groups with a weight associated with each group rather than area demand. The group-distance between the facility and a group of demand points is determined in three ways: the closest point in the group, the farthest one, and the average distance to all members in the group. Three objectives are considered: minisum (minimize the sum of weighted group-distances to the group), minimax, and maximin. There are nine possible models. Some of them are equivalent to known models. For example, the minisum objective with the average distance to the group is equivalent to the Weber problem. Two of the nine possible models (maximal group-distance using minisum objective, and average group-distance using minimax objective) are analyzed in the paper. The network version of this problem was analyzed in Berman et al. (2001). In Drezner and Drezner (2011) the equity objective (see Eiselt and Laporte 1995) is considered. The set of demand points is divided into two or more groups. For example, rich and poor neighborhoods and urban and rural neighborhoods. The objective is to provide "equitable" service to the groups. The objective function, to be minimized, is the sum of squares of differences between all pairs of service distances between demand points in different groups.

Drezner and Wesolowsky (1991d) considered the possibility that some of the weights can be negative. The facility may be obnoxious or "repulsive" to some demand points. It is also termed "the Weber problem with attraction and repulsion" (Maranas and Floudas 1993; Chen et al. 1992; Plastria 1991). Exact solution procedures are given for rectilinear and squared Euclidean distances. A heuristic is proposed for Euclidean distances. They proved that when the sum of the weights is positive, the optimal location is finite while if the sum is negative the optimal location is at infinity. When the sum of the weights is zero, the optimal location can be either finite or infinite. Examples for both cases are given. This problem was used as a test problem for global optimization procedures such as BTST (Drezner and Suzuki 2004) or BSSS (Hansen et al. 1981), which can solve the Euclidean distance problem optimally (see Sect. 6.7.2).

Drezner and Wesolowsky (1978a) investigated the Weber location problem on a sphere. When the area of interest is relatively small, the curvature of the surface can be ignored. However when the area of interest spans a large portion of the surface of the globe, the Euclidean distance may be significantly different from the distance on the surface of the earth. Using the shortest distances on a great circle, the problem may be non-convex and many local optima may exist. It was proven in Drezner and Wesolowsky (1978a) that the problem is convex when all demand points are located in a circle of radius $\frac{\pi}{4}$ of the radius of the globe. In Drezner (1981b) it was shown that when all demand points are located on a great circle, such as the equator, then the optimal location is on a demand point. In a later paper (Drezner 1985) an optimal procedure for solving the problem was proposed.

Drezner and Wesolowsky (1981) assumed that the weights associated with demand points in the Weber problem follow a distribution with given means, variances, and covariances. They found the probability that the optimal location is at a given location or a given region in the plane. For Manhattan distances (ℓ_1 norm) the probability that the optimal location is on a demand point is found. For general ℓ_p distances an approximate expression for the probability that the optimal solution is in a given region is developed. In a later paper, Drezner and Simchi-Levi (1992) showed that for Euclidean distances the probability that the optimal solution is on a demand point is approximately $\frac{1}{n}$ for the Weber problem with n demand points and random weights.

Drezner et al. (1991) introduced the Weber and minimax problems with limited distances. Up to a certain distance D, the distance is considered in the objective function and beyond distance D, the value of D is considered as the distance in the objective function. The distance $d_i(X)$ in (6.1) is replaced by $\min\{d_i(X), D\}$. Drezner et al. (1991) proposed a heuristic algorithm for the solution of the problem. It can be solved optimally by global optimization procedures detailed in Sect. 6.7.2, see, for example, Drezner et al. (2016a). Such a solution procedure is an important part of heuristic algorithms designed for the solution of the p-median problem (Daskin 1995). Drezner et al. (2016a) proposed to repeat the following until convergence. A facility is randomly selected and the optimal location of the selected facility while holding all other facilities in their locations is found. This is the Weber problem with limited distances. Let D_i be the minimum distance to the other facilities, then $d_i(X)$ is replaced by $\min\{d_i(X), D_i\}$. This approach is more effective than solving the Weber problem based only on the points closest to the selected facility by the Weiszfeld algorithm (Weiszfeld 1936; Drezner 2015) as suggested by Cooper (1963, 1964).

Most location models implicitly assume that travel time is proportional to the distance. Drezner et al. (2009b) considered vehicles (such as cars, airplanes, and trains) accelerating at the beginning of the trip and decelerating towards the end of the trip. This consideration is important in dispatching emergency services where time should be considered rather than distance. A heuristic approach employing the generalized Weiszfeld algorithm (Drezner 2009) and an optimal approach applying the big triangle small triangle global optimization method (Drezner and Suzuki 2004) are tested. These two approaches are very efficient and problems of 10,000 demand points are solved in about 0.015 s by the generalized Weiszfeld algorithm (Drezner 2009) and in about 1 min by the BTST technique (Drezner and Suzuki 2004). When the generalized Weiszfeld algorithm was repeated 1000 times, the optimal solution was found at least once for all test problems.

Drezner and Wesolowsky (1989b) considered the Weber problem and the minimax objective when the distance from point A to point B is not the same as the distance from B to A. This is common in rush hour traffic or for flights that in one direction have tail winds and in the opposite direction have head winds. The Weber problem with rectilinear distances is optimally solved and a heuristic procedure is proposed for Euclidean distances. Drezner and Drezner (2018) analyzed the asymmetric distance location problem where the distance (time) to get to the destination by air is affected by winds. Two models are proposed: the asymmetric Weber location problem and the round trip Weber location problem. The problems are analyzed and solved.

Berman et al. (2002b) considered the location of new facilities which serve only a certain proportion of the demand. The total weighted distances of the served demand is minimized. The problem is investigated in the plane for the location of one facility and on a network for the location of multiple facilities.

6.2.2 *Changing Costs*

Drezner and Wesolowsky (1991b) investigated the Weber problem (as well as the minimax objective) for locating one facility that serves a set of demand points over a time period. Demand at demand points changes over time in a "known" way. The facility can be relocated one or more times during the time horizon. The problem is to find the best time breaks for relocating the facility and the best location for the facility in each of the time windows. A follow-up paper is Farahani et al. (2009).

Berman et al. (2003d) considered the problem where there is a probability, depending on the distance from the facility, that the facility may not be able to provide satisfactory service to a customer. This probability is equal to 0 at distance zero, and is a monotonically increasing function of the distance. This problem is formulated and solved in a network environment. A given number of facilities need to be located such that the expected service level for all demand points combined will be maximized. Alternatively, one can state the problem as minimizing the expected demand that will not get satisfactory service.

Several papers extend the location-allocation problem by assuming that the prices charged to demand points depend on the demand and customers select the facility with the lowest total cost (charge plus transportation cost). Drezner and Wesolowsky (1996a) studied the case that demand points and facilities are located on a line segment and demand is continuous and follows a given distribution. Drezner and Wesolowsky (1999a) analyzed and solved the problem when demand is distributed in a convex region. The problem of two facilities to be located in a square with uniform demand is solved in detail. Drezner and Wesolowsky (1999b) analyzed and solved the problem in the plane. These three papers are summarized in Drezner and Wesolowsky (1996c). Extensions to these problems are proposed and solved in Averbakh et al. (1998, 2007).

6.3 Minimax and Maximin Models

In this section we review papers that investigate minimizing the maximum distance, maximizing the minimum distance, or a combination of both.

6.3.1 *Minimax Objective*

The basic minimax problem of finding the center of the smallest circle that encloses a set of points is credited to the English mathematician James Joseph Sylvester (1814–1897). Many papers discussed in Sect. 6.2 also investigated the minimax objective. Applications to the minimax objective include providing service so that the farthest customer will be as close as possible to the facility. For example, locating service stations for ambulances, fire trucks, and police depots. The following papers investigated solely the minimax objective.

The unweighted Euclidean (ℓ_2) problem was proposed and solved by Sylvester (1857, 1860). He showed that the problem is equivalent to finding the circle with the smallest radius that covers a set of points. Drezner and Wesolowsky (1980e) solved the weighted one facility minimax problem with ℓ_p distances for $p \geq 1$. The solution method is based on the property that the optimal point is the solution to a problem based on a subset of two or three demand points. Elzinga and Hearn (1972) proposed it for Euclidean distances. For a review see Drezner (2011). Drezner (1987) solved the unweighted p-center problem with Manhattan (ℓ_1) distances. The 1-center and 2-center problems are optimally solved in $o(n)$ time and the 3-center problem is optimally solved in $o(n \log n)$ time.

Drezner and Wesolowsky (1985a) considered the location of facilities or "movable" points on a planar area, on which there already exist fixed points. The minimax criterion for optimality is used and distances among points are assumed to be rectilinear. Two very efficient algorithms for the solution of the problem are presented. One is based on a univariate search, and the other on a steepest descent method.

Drezner et al. (2002b) found the circle whose circumference is as close as possible to a given set of points. Three objectives are considered: minimizing the sum of squares of distances, minimizing the maximum distance, and minimizing the sum of distances. These problems are equivalent to minimizing the variance, minimizing the range, and minimizing the mean absolute deviation, respectively. These problems are formulated and heuristically solved as mathematical programs. Follow-up papers for locating concentric circles are Drezner and Brimberg (2014) and Brimberg and Drezner (2015).

Berman et al. (2003f) considered the weighted minimax (1-center) location problem in the plane when the weights are not given but rather drawn from independent uniform distributions. The problem is formulated and analyzed. For certain parameters of the uniform distributions the objective function is proven to be convex and thus can be easily solved by standard software such as the Solver in Excel.

Berman et al. (2003b) introduced a new objective function for the minimax location problem. Every demand point generates demand for service with a given probability (during a given period of time) and the objective is to minimize the expected maximum distance. The planar problem is proven to be convex and thus standard solution techniques such as using the Solver in Excel can be applied for its solution. Properties for the problem on the network are proven and an efficient algorithm proposed for its solution.

6.3.2 Maximin (Obnoxious) Objective

Maximizing the minimum distance from a set of demand points was introduced by Church and Garfinkel (1978) as the obnoxious facility location problem. Their model was proposed for the network environment.

Drezner and Wesolowsky (1980b) proposed, analyzed, and tested the problem of locating a facility that maximizes the minimum weighted distance from a set of demand points. The facility must be located within a certain distance from all demand points. This condition restricts the set of potential locations to a finite set. The non-weighted version of this problem without maximum distance constraints was solved by Shamos and Hoey (1975) based on Voronoi diagrams (Suzuki and Okabe 1995; Okabe et al. 2000; Voronoï 1908).

Berman et al. (2003c) considered the location of an obnoxious facility, such as an airport, that serves only a certain proportion of the demand. Each demand point can be bought by the developer at a given price. An expropriation budget is given. Demand points closest to the facility are expropriated within the given budget. The objective is to maximize the distance to the closest point not expropriated. The problem is formulated and polynomial algorithms are proposed for its solution both on the plane and on a network.

Drezner and Wesolowsky (1985b) considered the location of facilities that are obnoxious in the sense that nearness of the facility to fixed points, which may represent population centers or other installations, is undesirable. Two model formulations are proposed. In the first formulation the maximum weighted distance in the system is minimized subject to constraints which require that the distances between the facilities and fixed points must exceed specified values. In the second formulation, the smallest weighted "facility-to-fixed-point" distance in the system must be maximized, given that every fixed point must be within "reach" of the closest facility. Certain useful duality relationships are established between these problems. A one-dimensional problem is solved using an algorithm that incorporates a version of the set covering problem.

Drezner and Wesolowsky (1983a) considered the basic obnoxious facility model. The minimum distance between demand points and the facility needs to be maximized. Rectangular (ℓ_1) distances are considered. Two approaches to solving the problem are proposed. In the first approach, the boundary and then the interior of the feasible region are searched for the optimum. The search is restricted to certain linear segments. The second algorithm essentially breaks down the problem into linear programming problems, one of which must yield the optimal solution.

Drezner and Wesolowsky (1989d) considered the location of a route or path through a set of given points in order to maximize the smallest weighted distance from the given points to the route. Applications may include the planning of pipelines carrying noxious material, and also certain problems in robotics. The first algorithm finds a non-linear path by iteratively solving network minimal-cut problems. A second algorithm solves the case where the route is restricted to be linear.

Drezner and Wesolowsky (1996b) considered the problem of locating a point that is as far as possible from arcs and nodes of a network. Each arc or node may have a different multiplicative factor (weight) for its distance. A graphical solution approach, as well as a computational algorithm, is presented. In a follow-up paper, Drezner et al. (2009a) also investigated the location of a facility anywhere inside a planar network but a different objective is proposed. The objective is to locate a facility where the total nuisance to links and nodes of the network is minimized.

Berman et al. (1996) find a location of a new facility on a network so that the total number (weight) of nodes within a pre-specified distance R is minimized. This problem is applicable when locating an obnoxious facility such as garbage dumps, nuclear reactors, prisons, and military installations. The paper includes an analysis of the problem, identification of special cases where the problem is easily solved, an algorithm to solve the problem in general, and a sensitivity analysis of R. The planar version with Euclidean distances is solved in Drezner (1981a).

Berman et al. (2000) considered the minimization of the impact of hazards located on or near a network. Two situations are considered: (1) a hazard is located on a network and affects off-network sites and (2) an off-network hazard which can affect traffic on the network. Eight models aimed at optimizing different objectives are developed and solved, including finding a route between two nodes on a network which minimizes the hazard along it and finding a location on a network where the hazard is minimized.

6.3.3 Incorporating Both Minimax and Maximin Objectives

Drezner and Wesolowsky (1983b) considered the following problem. There are n demand points on a sphere. Each demand point has a weight which is a positive constant. A facility must be located so that the maximum of the weighted distances (distances are the shortest arcs on the surface of the sphere) is minimized; this is called the minimax problem. Alternatively, in the maximin problem, the minimum weighted distance is maximized. A setup cost associated with each demand point may be added for generality. It is shown that any maximin problem can be re-parametrized into a minimax problem. A method for finding local minimax points is described and conditions under which these are global are derived. Finally, an efficient algorithm for finding the global minimax point is constructed.

Berman et al. (2003a) considered the weighted minimax and maximin location problems on the network when the weights are drawn from a uniform distribution. In the minimax (maximin) problem with stochastic demand the probability that the maximum (minimum) weighted distance between the facility and demand points exceeding (falling short of) a given value T is minimized. Properties of the solution points for both problems are proven and solution algorithms are presented.

Drezner et al. (1986) considered a new objective function for the placement of a public facility with reference to variations in accessibility: the minimization of the range between the maximal and the minimal distances to users (one of the equity objectives listed in Eiselt and Laporte 1995). Some properties of the solution are given. Algorithms for the Euclidean and rectilinear distance cases are presented. In the follow-up papers (Drezner and Drezner 2007; Drezner 2007) the global optimization algorithm BTST (Drezner and Suzuki 2004) was applied to optimally solve this problem, as well as other objectives.

6.4 Cover Models

Cover location models are one of the main branches of location analysis. In the original models (Church and ReVelle 1974; ReVelle et al. 1976) as well as in many follow-up models, a demand point is covered by a facility within a certain distance. In maximum covering models (Church and ReVelle 1974), facilities need to be located in an area to provide as much cover as possible. Set covering problems (ReVelle et al. 1976) aim to cover all demand points with the minimum number of facilities. Such models are used for cover provided by emergency facilities such as ambulances, police cars, or fire trucks. They are also used to model cover by transmission towers such as cell-phone towers, TV or radio transmission towers, and radar coverage among others. For a review of cover models see Plastria (2002), García and Marín (2015), and Snyder (2011).

Drezner et al. (2004) investigated the gradual covering problem. Within a certain distance r from the facility the demand point is fully covered, and beyond another specified distance $R \geq r$ the demand point is not covered. Between these two given distances the coverage is linear in the distance from the facility. If $R = r$, the gradual cover reduces to the original model where the drop in cover is abrupt. This formulation can be converted to the Weber problem by imposing a special structure on its cost function. The cost is zero (negligible) up to a certain minimum distance, and it is a constant beyond a certain maximum distance. Between these two extreme distances the cost is linear in the distance. The problem is analyzed and a branch and bound procedure is proposed for its solution. The gradual cover concept, sometimes referred to as partial cover, is investigated in many follow-up papers, including Berman et al. (2003e), Drezner et al. (2010), Drezner and Drezner (2014, 2019), Karatas (2017), Bagherinejad et al. (2018), Drezner et al. (2019a), Eiselt and Marianov (2009), and Berman et al. (2009e).

Berman et al. (2009d) investigated the maximal covering problem on a network when some of the weights can be negative. Demand points with a negative weight are demand points we do not wish to cover. For example, if all weights are negative, we wish to cover as little weight as possible. Integer programming formulations are proposed and tested with ILOG-CPLEX. Heuristic algorithms, an ascent algorithm, and simulated annealing (Kirkpatrick et al. 1983) are proposed and tested. The simulated annealing approach provided the best results.

Drezner and Wesolowsky (2005) considered the problem of covering the most probability of a multivariate normal distribution by selection of a given number of hypercubes of a given size. The problem is formulated and meta heuristic procedures proposed for its solution.

Berman et al. (2009b) proposed a covering problem where the covering radius of a facility is controlled by the decision-maker. The cost of achieving a certain covering distance is assumed to be a monotonically increasing function of the distance (i.e., it costs more to establish a facility with a greater covering radius). The problem is to cover all demand points at a minimum cost by finding the optimal number, locations, and coverage radii for the facilities. Both, the planar and discrete versions of the model are considered. Heuristic approaches are suggested for solving large problems in the plane. Mathematical programming formulations are proposed for the discrete problem, and a solution approach is suggested and tested.

Drezner and Wesolowsky (2014) analyzed the problem of locating facilities in a feasible area covering some parts of network links within a given radius. The feasible area can be the interior (convex hull of the nodes) of a planar network or any union of convex polygons. Both minimization and maximization of coverage are considered. The single facility location problem is solved by the global optimization approach BTST (Drezner and Suzuki 2004). The multiple facility maximization problem is solved by a specially designed heuristic algorithm. The idea of the heuristic algorithm may prove to work well on other planar multiple facility location problems. Computational experience with problems of up to 40,000 links demonstrates the effectiveness of the single facility and multiple facilities algorithms. The largest single facility minimal cover problem is solved in about 1 min and the largest single facility maximal cover problem is solved in less than 4 min.

Drezner et al. (2016b) proposed a stochastic model for the location of emergency facilities. The model is formulated and analyzed. The location of one facility in the plane is optimally solved. Optimal algorithms are proposed for the location of multiple facilities on a network. Computational experiments illustrate the effectiveness of these solution procedures.

Berman et al. (2009a) consider the situation where p facilities need to be located by a leader, on the nodes of a network, to provide maximum coverage of demand generated at nodes of the network. At some point in the future it is expected that one of the links of the network will become unusable either due to a terrorist attack or a natural disaster (referred to as the follower). The leader's objective is to retain as much cover as possible following the worst disruption. In case of a terrorist attack the selected link is intentional and in case of a natural disaster the leader wants to protect against the worst possible scenario. The follower's objective is to remove the link that causes the most damage. The leader's objective is to cover the most demand following such a damage to a link. The problem is formulated and analyzed from the leader's perspective. An efficient approach to solving the follower's problem is constructed. The leader's problem is solved heuristically by an ascent algorithm, simulated annealing (Kirkpatrick et al. 1983), and tabu search (Glover and Laguna 1997), using the efficient algorithm for the solution of the follower's problem.

Drezner and Wesolowsky (1994) considered a problem applicable to both obnoxious facility location and problems in production where one has to select an area of material with the least weight of defects. In the first version a boundary circle containing weighted points is given. The objective is to find a location of an interior covering circle of a given radius that encloses the smallest weight of points. In the second version we need to find a rectangle inside a rectangle. Two objectives are considered, minimizing the sum of weights or minimizing the maximum weight. Algorithms are constructed for solving both problems. An earlier version of the paper is Drezner and Wesolowsky (1993).

6.5 Hub Related Objectives

Berman et al. (2007) introduced the transfer point location problem. Demand for emergency service is generated at a set of demand points which need the services of a central facility (such as a hospital). Patients are transferred to a helicopter pad (transfer point) at normal speed, and from there they are transferred to the facility at increased speed. The general model involves the location of p helicopter pads and one facility. The special case where the location of the facility is known and the best location of one transfer point that serves a set of demand points is solved. Both minisum and minimax versions of the models are investigated. Berman et al. (2005) investigated the location of a facility and several transfer points. Heuristic approaches were proposed for the solution of this problem. Berman et al. (2008) applied the results of Berman et al. (2005) to solve the problem when the location of the facility is known. Both minisum and minimax versions of the models are investigated both in the plane and on the network.

Drezner and Wesolowsky (2001) investigated the problem of locating a new facility servicing a set of demand points. A given set of collection depots is also given. When service is required by a demand point, the server travels from the facility to the demand point, then from the demand point to one of the collection depots (which provides the shortest route back to the facility), and back to the facility. Applications include a septic tank cleaning service, garbage collection, or tree pruning. The service truck travels to the customer, collects the load, selects the collection depot that provides the shortest route from the demand point back to the facility to wait for the next call. When a depot must be included on the way to the customer, the model is the same. This is the case of a service where the vehicle collects some materials on the way to the customer and returns to the facility empty. The problem is analyzed and properties of the solution point are formulated and proved. The network version of this problem was investigated in Berman et al. (2002a). Drezner et al. (2019b) investigated the multiple facilities collection depots problem in the plane.

6.6 Other Objectives

Berman et al. (2009c) considered the problem of finding locations for p facilities such that the weights attracted to each facility will be as close as possible to one another. The problem is modeled as minimizing the maximum among all the total weights attracted to the various facilities. Solution procedures for the problem on a network and for the special cases of the problem on a tree or on a path are proposed.

Drezner and Wesolowsky (1980a) find the expected value of perfect information (EVPI) in a simple facility location problem where the weights, which summarize cost and volume parameters, are random draws from a multivariate probability distribution. A model with rectangular distances and one with squared Euclidean (center of gravity) are used. The analysis is developed for a multivariate normal distribution of weights but simulation is used to show that the expressions derived are reasonably accurate for other distributions.

Drezner and Wesolowsky (1980d) introduced the following problem. There are n points on the plane that are to be observed from some point on a circle of given radius that encloses all of the points. We wish to find the observation point that has the best possible view of the n points in the sense that if we draw lines of sight from the observation point to the given points, the smallest angle between the lines is maximized. Applications include the planning of photographs or displays. This is a "maximin problem" in which the function to be maximized has many local optima.

Drezner and Wesolowsky (1998) considered the rectilinear minisum and minimax location problems from a different point of view in that the orientation of the axes, which define the distances, is now also to be optimized. This corresponds to the situation where the grid of roads or aisles which connects the demand points to the facility can be designed at the same time as the location of the facility is chosen.

Drezner and Wesolowsky (1995a) considered two basic location problems: the Weber problem, and the minimax problem on a regular grid of alternating one-way routes or streets. Both the facility to be located and the demand points are restricted to any point on the network. The one-way restriction is often used for efficiency in traffic flow, but complicates the distances in the system.

Drezner and Wesolowsky (1997) considered the situation where an event may occur anywhere in a planar area or on a linear region such as a route. One or more detectors are to be located within this region with the objective of maximizing the smallest probability of the detection of an event anywhere in the region. In other words, the minimum protection in the region is to be maximized. The probability that an event is detected by a detector is a decreasing function of the distance.

Drezner et al. (1985) investigated the location of a facility among n points where the points are serviced by "tours" taken from the facility. Tours include m points at a time and each group of m points may become active (may need a tour) with some known probability. Distances are assumed to be rectilinear. For $m \leq 3$, it is proved that the objective function is separable in each dimension and an exact solution method is given that involves finding the median of numbers appropriately generated from the problem data. It is shown that the objective function becomes multi-modal when some tours pass through four or more points. A bounded heuristic procedure is suggested for this latter case.

Drezner and Wesolowsky (1997) considered the following problem. An event may occur anywhere in a planar area or on a linear region such as a route. One or more detectors are to be located within this region with the objective of maximizing the smallest probability of the detection of an event anywhere in the region. In other words, the minimum protection in the region is to be maximized. The probability that an event is detected by a detector is a decreasing function of the distance. Two solution procedures are proposed for the problem on a line segment: a mathematical programming model and a specially designed algorithm. The problem in an area is solved by a univariate search, a Demjanov-type algorithm (see Sect. 6.7.3), a mathematical programming model, and simulated annealing.

Drezner and Wesolowsky (2003) introduced new network design problems. A network of potential links is given. Each link can be either constructed or not at a given cost. Also, each constructed link can be constructed either as a one-way or two-way link. The objective is to minimize the total construction and transportation costs. Two different transportation costs are considered: (1) traffic is generated between any pair of nodes and the transportation cost is the total cost for the users and (2) demand for service is generated at each node and a facility is to be located on a node to satisfy the demand. The transportation cost in this case is the total cost for a round trip from the facility to each node and back. Two options in regard to the links between nodes are considered. They can either be two-way only, or mixed, with both two-way and one-way (in either direction) allowed. When these options are combined with the two objective functions, four basic problems are created. These problems are solved by a descent algorithm, simulated annealing (Kirkpatrick et al. 1983), tabu search (Glover and Laguna 1997), and a genetic algorithm (Goldberg 2006). A follow-up paper is Drezner and Salhi (2002).

Drezner et al. (1998) showed that the random utility model (Drezner and Drezner 1996; Leonardi and Tadei 1984) can be approximated by a logit model. The proportion of the buying power at a demand point that is attracted to the new facility can be approximated by a logit function of the distance to it. This approximation demonstrates that using the logit function of the distance for estimating the market share is theoretically found in the random utility model. A simplified random utility model is defined and approximated by a logit function. An iterative Weiszfeld-type algorithm is designed to find the best location for a new facility using the logit model.

6.7 Solution Approaches

6.7.1 The Trajectory Approach

The idea of solving location problems by a trajectory approach is described in three papers (Drezner and Wesolowsky 1978d,b, 1982). The idea is to use a known solution point for a similar model and calculate the trajectory of the solution by changing a parameter and solving a set of differential equations numerically. For example, the solution to the Weber problem with squared Euclidean distances is the center of gravity which can be easily calculated. Consider the Weber problem with the distances $d_i(X)$ replaced by $d_i(X)^{\lambda}$. For $\lambda = 2$ the solution point is known (center of gravity), while the desired solution point is the solution for $\lambda = 1$. There is a trajectory of solution points for $1 \leq \lambda \leq 2$. A set of two differential equations of the locations (x, y) by the parameter λ can be constructed and numerically solved by, for example, Runge (1895, 1901), Kutta (1901), resulting in the desired solution point for $\lambda = 1$.

6.7.2 Global Optimization Techniques

Many non-convex single facility non-convex problems can be optimally solved by global optimization algorithms such as "big square small square" (BSSS, Hansen et al. 1981) and "big triangle small triangle" (BTST, Drezner and Suzuki 2004).

The BSSS algorithm starts with a list consisting of a "big square" enclosing the feasible region. The upper bound in the big square, $\overline{UB}$, is the best upper bound found so far. The big square is divided into four "small squares" and lower and upper bounds established in each small square. $\overline{UB}$ is possibly updated. The big square is removed from the list. If the lower bound LB in a small square satisfies $LB \geq \overline{UB}(1-\epsilon)$ for a given relative accuracy ϵ, the small square is eliminated from further consideration. Otherwise, the small square is added to the list of squares. The process continues by selecting a square in the list with the smallest LB as a "big square" and split it into four "small squares" until the list of squares is empty.

The BTST algorithm starts with a Delaunay triangulation (Lee and Schachter 1980) of the convex hull of the demand points and the vertices of the feasible region as the initial list of triangles. An upper bound and a lower bound for each triangle in the list are calculated and $\overline{UB}$ is the best one. Many of the triangles in the list for which $LB \geq \overline{UB}(1 - \epsilon)$ are eliminated. The remainder of the process is very similar to the BSSS algorithm. A selected triangle in the list is divided into four small triangles by connecting the centers of the sides of the big triangle.

If the solution is restricted to the convex hull of the demand points, or any convex polygon, the triangulation by BTST takes care of it automatically while BSSS requires an extra check whether the solution point is feasible or not. The lower and upper bounds required for applying the BSSS algorithm may also be affected by the feasibility issue.

Extensions to these global optimization algorithms include big region small region (BRSR, Hansen et al. 1995) for location on a sphere, big cube small cube (BCSC, Schöbel and Scholz 2010) for three or more dimensional problems, big segment small segment (BSSS, Berman et al. 2011) for location on a network, and big arc small arc (BASA, Drezner et al. 2018) for location on circumferences of circles.

6.7.3 The Demjanov Optimization Technique

This method was proposed in Demjanov (1968) and applied in Drezner and Wesolowsky (1997, 1985a). It is designed to solve the minimization of $f(X) = \max_{1 \le i \le n} \{f_i(X)\}$ where X is a vector of p variables. It can be used to solve heuristically, for example, the location of p facilities among n demand points. For such an application the vector X is a vector of locations, which is a vector of $2p$ variables.

The Demjanov algorithm starts with a random location for the p variables and improves the solution by moving in the direction of steepest descent. The objective function is optimized on the ray of steepest descent using a one-dimensional optimization procedure such as the golden section search (Zangwill 1969).

Calculation of the gradient is done as follows. A tolerance δ is selected defining the set $I(X)$:

$$I(X) = \{ i \mid f_i(X) \ge f(X) - \delta \}$$

The set $I(X)$ is the set of "binding" variables. If the function $f_i(X)$ is reduced for all binding variables when X is changed infinitesimally, then the objective function is reduced. Consider changing variable j by Δx_j for $j = 1, \ldots, p$. The steepest descent direction of $f(X)$ is obtained by solving the following problem:

$$\min \sum_{j=1}^{p} \Delta x_j^2$$

$$\text{subject to: } \sum_{j=1}^{p} \frac{\partial f_i(X)}{\partial x_j} \Delta x_j \le -1 \quad \text{for } i \in I(X)$$

References

Averbakh, I., Berman, O., Drezner, Z., & Wesolowsky, G. O. (1998). The plant location problem with demand-dependent setup cost and centralized allocation. *European Journal of Operational Research, 111*, 543–554.

Averbakh, I., Berman, O., Drezner, Z., & Wesolowsky, G. O. (2007). The uncapacitated facility location problem with demand-dependent setup and service costs and flexible allocation. *European Journal of Operational Research, 179*, 956–967.

Bagherinejad, J., Bashiri, M., & Nikzad, H. (2018). General form of a cooperative gradual maximal covering location problem. *Journal of Industrial Engineering International, 14*, 241–253.

Berman, O., Drezner, T., Drezner, Z., & Wesolowsky, G. O. (2009a). A defensive maximal covering problem on a network. *International Transactions on Operational Research, 16*, 69–86.

Berman, O., Drezner, Z., & Krass, D. (2011). Big segment small segment global optimization algorithm on networks. *Networks, 58*, 1–11.

Berman, O., Drezner, Z., Krass, D., & Wesolowsky, G. O. (2009b). The variable radius covering problem. *European Journal of Operational Research, 196*, 516–525.

Berman, O., Drezner, Z., Tamir, A., & Wesolowsky, G. O. (2009c). Optimal location with equitable loads. *Annals of Operations Research, 167*, 307–325.

Berman, O., Drezner, Z., Wang, J., & Wesolowsky, G. O. (2003a). The minimax and maximin location problems with uniform distributed weights. *IIE Transactions, 35*, 1017–1025.

Berman, O., Drezner, Z., & Wesolowsky, G. O. (1996). Minimum covering criterion for obnoxious facility location on a network. *Networks, 18*, 1–5.

Berman, O., Drezner, Z., & Wesolowsky, G. O. (2000). Routing and location on a network with hazardous threats. *Journal of the Operational Research Society, 51*, 1093–1099.

Berman, O., Drezner, Z., & Wesolowsky, G. O. (2001). Location models with groups of demand points on a network. *IIE Transactions, 33*, 637–648.

Berman, O., Drezner, Z., & Wesolowsky, G. O. (2002a). The collection depots location problem on networks. *Naval Research Logistics, 49*, 15–24.

Berman, O., Drezner, Z., & Wesolowsky, G. O. (2002b). Satisfying partial demand in facilities location. *IIE Transactions, 24*, 971–978.

Berman, O., Drezner, Z., & Wesolowsky, G. O. (2003b). The expected maximum distance objective in facility location. *Journal of Regional Science, 43*, 735–748.

Berman, O., Drezner, Z., & Wesolowsky, G. O. (2003c). The expropriation location problem. *Journal of the Operational Research Society, 54*, 769–776.

Berman, O., Drezner, Z., & Wesolowsky, G. O. (2003d). Locating service facilities whose reliability is distance dependent. *Computers & Operations Research, 30*, 1683–1695.

Berman, O., Drezner, Z., & Wesolowsky, G. O. (2005). The facility and transfer points location problem. *International Transactions in Operational Research, 12*, 387–402.

Berman, O., Drezner, Z., & Wesolowsky, G. O. (2007). The transfer point location problem. *European Journal of Operational Research, 179*, 978–989.

Berman, O., Drezner, Z., & Wesolowsky, G. O. (2008). The multiple location of transfer points. *Journal of the Operational Research Society, 59*, 805–811.

Berman, O., Drezner, Z., & Wesolowsky, G. O. (2009d). The maximal covering problem with some negative weights. *Geographical Analysis, 41*, 30–42.

Berman, O., Kalcsics, J., Krass, D., & Nickel, S. (2009e). The ordered gradual covering location problem on a network. *Discrete Applied Mathematics, 157*, 3689–3707.

Berman, O., Krass, D., & Drezner, Z. (2003e). The gradual covering decay location problem on a network. *European Journal of Operational Research, 151*, 474–480.

Berman, O., Wang, J., Drezner, Z., & Wesolowsky, G. O. (2003f). A probabilistic minimax location problem on the plane. *Annals of Operations Research, 122*, 59–70.

Brimberg, J., & Drezner, Z. (2015). A location-allocation problem with concentric circles. *IIE Transactions, 47*, 1397–1406.

Chen, P., Hansen, P., Jaumard, B., & Tuy, H. (1992). Weber's problem with attraction and repulsion. *Journal of Regional Science, 32*, 467–486.

Church, R. L., & Garfinkel, R. S. (1978). Locating an obnoxious facility on a network. *Transportation Science, 12*, 107–118.

Church, R. L., & ReVelle, C. S. (1974). The maximal covering location problem. *Papers of the Regional Science Association, 32*, 101–118.

Cooper, L. (1963). Location-allocation problems. *Operations Research, 11*, 331–343.

Cooper, L. (1964). Heuristic methods for location-allocation problems. *SIAM Review, 6*, 37–53.

Daskin, M. S. (1995). *Network and discrete location: Models, algorithms, and applications*. New York: Wiley.

Demjanov, V. F. (1968). Algorithms for some minimax problems. *Journal of Computer and System Sciences, 2*, 342–380.

Drezner, T., & Drezner, Z. (1996). Competitive facilities: Market share and location with random utility. *Journal of Regional Science, 36*, 1–15.

Drezner, T., & Drezner, Z. (2007). Equity models in planar location. *Computational Management Science, 4*, 1–16.
Drezner, T., & Drezner, Z. (2011). A note on equity across groups in facility location. *Naval Research Logistics, 58*, 705–711.
Drezner, T., & Drezner, Z. (2014). The maximin gradual cover location problem. *OR Spectrum, 36*, 903–921.
Drezner, T., & Drezner, Z. (2018). Asymmetric distance location model (in review).
Drezner, T., & Drezner, Z. (2019). Extensions to the directional approach to gradual cover (in review).
Drezner, T., Drezner, Z., & Scott, C. H. (2009a). Location of a facility minimizing nuisance to or from a planar network. *Computers & Operations Research, 36*, 135–148.
Drezner, T., Drezner, Z., & Schöbel, A. (2018). The Weber obnoxious facility location model: A Big Arc Small Arc approach. *Computers and Operations Research, 98*, 240–250.
Drezner, T., Drezner, Z., & Goldstein, Z. (2010). A stochastic gradual cover location problem. *Naval Research Logistics, 57*, 367–372.
Drezner, T., Drezner, Z., & Kalczynski, P. (2019a). A directional approach to gradual cover. *TOP, 27*, 70–93.
Drezner, T., Drezner, Z., & Kalczynski, P. (2019b). The planar multifacility collection depots location problem. *Computers and Operations Research, 102*, 121–129.
Drezner, Z. (1981a). On a modified one-center model. *Management Science, 27*, 848–851.
Drezner, Z. (1981b). On location dominance on spherical surfaces. *Operations Research, 29*, 1218–1219.
Drezner, Z. (1985). A solution to the Weber location problem on the sphere. *Journal of the Operational Research Society, 36*, 333–334.
Drezner, Z. (1987). On the rectangular p-center problem. *Naval Research Logistics Quarterly, 34*, 229–234.
Drezner, Z. (2007). A general global optimization approach for solving location problems in the plane. *Journal of Global Optimization, 37*, 305–319.
Drezner, Z. (2009). On the convergence of the generalized Weiszfeld algorithm. *Annals of Operations Research, 167*, 327–336.
Drezner, Z. (2011). Continuous center problems. In H. A. Eiselt & V. Marianov (Eds.), *Foundations of location analysis* (pp. 63–78). Berlin: Springer.
Drezner, Z. (2015). The fortified Weiszfeld algorithm for solving the Weber problem. *IMA Journal of Management Mathematics, 26*, 1–9.
Drezner, Z., & Brimberg, J. (2014). Fitting concentric circles to measurements. *Mathematical Methods of Operations Research, 79*, 119–133.
Drezner, Z., Brimberg, J., Mladenovic, N., & Salhi, S. (2016a). New local searches for solving the multi-source Weber problem. *Annals of Operations Research, 246*, 181–203.
Drezner, Z., Drezner, T., & Wesolowsky, G. O. (2009b). Location with acceleration-deceleration distance. *European Journal of Operational Research, 198*, 157–164.
Drezner, Z., Klamroth, K., Schöbel, A., & Wesolowsky, G. O. (2002a). The Weber problem. In Z. Drezner & H. W. Hamacher (Eds.), *Facility location: Applications and theory* (pp. 1–36). Berlin: Springer.
Drezner, Z., Marianov, V., & Wesolowsky, G. O. (2016b). Maximizing the minimum cover probability by emergency facilities. *Annals of Operations Research, 246*, 349–362.
Drezner, Z., Mehrez, A., & Wesolowsky, G. O. (1991). The facility location problem with limited distances. *Transportation Science, 25*, 183–187.
Drezner, Z., & Salhi, S. (2002). Using hybrid metaheuristics for the one-way and two-way network design problem. *Naval Research Logistics, 49*, 449–463.
Drezner, Z., & Simchi-Levi, D. (1992). Asymptotic behavior of the Weber location problem on the plane. *Annals of Operations Research, 40*, 163–172.
Drezner, Z., Steiner, G., & Wesolowsky, G. O. (1985). One-facility location with rectilinear tour distances. *Naval Research Logistics Quarterly, 32*, 391–405.

Drezner, Z., Steiner, S., & Wesolowsky, G. O. (2002b). On the circle closest to a set of points. *Computers & Operations Research, 29*, 637–650.

Drezner, Z., & Suzuki, A. (2004). The big triangle small triangle method for the solution of non-convex facility location problems. *Operations Research, 52*, 128–135.

Drezner, Z., Szendrovits, A. Z., & Wesolowsky, G. O. (1984). Multi-stage production with variable lot sizes and transportation of partial lots. *European Journal of Operational Research, 17*, 227–237.

Drezner, Z., Thisse, J.-F., & Wesolowsky, G. O. (1986). The minimax-min location problem. *Journal of Regional Science, 26*, 87–101.

Drezner, Z., & Wesolowsky, G. O. (1978a). Facility location on a sphere. *Journal of the Operational Research Society, 29*, 997–1004.

Drezner, Z., & Wesolowsky, G. O. (1978b). A new method for the multifacility minimax location problem. *Journal of the Operational Research Society, 29*, 1095–1101.

Drezner, Z., & Wesolowsky, G. O. (1978c). A note on optimal facility location with respect to several regions. *Journal of Regional Science, 18*, 303.

Drezner, Z., & Wesolowsky, G. O. (1978d). A trajectory method for the optimization of the multifacility location problem with lp distances. *Management Science, 24*, 1507–1514.

Drezner, Z., & Wesolowsky, G. O. (1980a). The expected value of perfect information in facility location. *Operations Research, 28*, 395–402.

Drezner, Z., & Wesolowsky, G. O. (1980b). A maximin location problem with maximum distance constraints. *AIIE Transactions, 12*, 249–252.

Drezner, Z., & Wesolowsky, G. O. (1980c). Optimal location of a facility relative to area demands. *Naval Research Logistics Quarterly, 27*, 199–206.

Drezner, Z., & Wesolowsky, G. O. (1980d). The optimal sight angle problem. *AIIE Transactions, 12*, 332–338.

Drezner, Z., & Wesolowsky, G. O. (1980e). Single facility lp distance minimax location. *SIAM Journal of Algebraic and Discrete Methods, 1*, 315–321.

Drezner, Z., & Wesolowsky, G. O. (1981). Optimum location probabilities in the lp distance Weber problem. *Transportation Science, 15*, 85–97.

Drezner, Z., & Wesolowsky, G. O. (1982). A trajectory approach to the round trip location problem. *Transportation Science, 16*, 56–66.

Drezner, Z., & Wesolowsky, G. O. (1983a). The location of an obnoxious facility with rectangular distances. *Journal of Regional Science, 23*, 241–248.

Drezner, Z., & Wesolowsky, G. O. (1983b). Minimax and maximin facility location problems on a sphere. *Naval Research Logistics Quarterly, 30*, 305–312.

Drezner, Z., & Wesolowsky, G. O. (1985a). Layout of facilities with some fixed points. *Computers & Operations Research, 12*, 603–610.

Drezner, Z., & Wesolowsky, G. O. (1985b). Location of multiple obnoxious facilities. *Transportation Science, 19*, 193–202.

Drezner, Z., & Wesolowsky, G. O. (1989a). An approximation method for bivariate and multivariate normal equiprobability contours. *Communications in Statistics: Theory and Methods, 18*, 2331–2344.

Drezner, Z., & Wesolowsky, G. O. (1989b). The asymmetric distance location problem. *Transportation Science, 23*, 201–207.

Drezner, Z., & Wesolowsky, G. O. (1989c). Control limits for a drifting process with quadratic loss. *International Journal of Production Research, 27*, 13–20.

Drezner, Z. & Wesolowsky, G. O. (1989d). Location of an obnoxious route. *Journal of the Operational Research Society, 40*, 1011–1018.

Drezner, Z., & Wesolowsky, G. O. (1989e). Multi-buyer discount pricing. *European Journal of Operational Research, 40*, 38–42.

Drezner, Z., & Wesolowsky, G. O. (1989f). Optimal control of a linear trend process with quadratic loss. *IIE Transactions, 21*, 66–72.

Drezner, Z., & Wesolowsky, G. O. (1990). On the computation of the bivariate normal integral. *Journal of Statistical Computation and Simulation, 35*, 101–107.

Drezner, Z., & Wesolowsky, G. O. (1991a). Design of multiple criteria sampling plans and charts. *International Journal of Production Research, 29*, 155–163.

Drezner, Z., & Wesolowsky, G. O. (1991b). Facility location when demand is time dependent. *Naval Research Logistics, 38*, 763–777.

Drezner, Z., & Wesolowsky, G. O. (1991c). Optimizing control limits under random process shifts and a quadratic penalty function. *Communications in Statistics: Stochastic Methods, 7*, 363–377.

Drezner, Z., & Wesolowsky, G. O. (1991d). The Weber problem on the plane with some negative weights. *Information Systems and Operational Research, 29*, 87–99.

Drezner, Z., & Wesolowsky, G. O. (1993). Finding the circle or rectangle containing the minimum weight of points. *Studies in Locational Analysis, 4*, 105–109.

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

Drezner, Z., & Wesolowsky, G. O. (1995a). Location on a one-way rectilinear grid. *Journal of the Operational Research Society, 46*, 735–746.

Drezner, Z., & Wesolowsky, G. O. (1995b). Multivariate screening procedures for quality cost minimization. *IIE Transactions, 27*, 300–304.

Drezner, Z., & Wesolowsky, G. O. (1996a). Location-allocation on a line with demand-dependent costs. *European Journal of Operational Research, 90*, 444–450.

Drezner, Z., & Wesolowsky, G. O. (1996b). Obnoxious facility location in the interior of a planar network. *Journal of Regional Science, 35*, 675–688.

Drezner, Z., & Wesolowsky, G. O. (1996c). Review of location-allocation models with demand-dependent costs. *Studies in Locational Analysis, 10*, 13–24.

Drezner, Z., & Wesolowsky, G. O. (1997). On the best location of signal detectors. *IIE Transactions, 29*, 1007–1015.

Drezner, Z., & Wesolowsky, G. O. (1998). Optimal axis orientation for rectilinear minisum and minimax location. *IIE Transactions, 30*, 981–986.

Drezner, Z., & Wesolowsky, G. O. (1999a). Allocation of demand when cost is demand-dependent. *Computers & Operations Research, 26*, 1–15.

Drezner, Z., & Wesolowsky, G. O. (1999b). Allocation of discrete demand with changing costs. *Computers & Operations Research, 26*, 1335–1349.

Drezner, Z., & Wesolowsky, G. O. (2000). Location problems with groups of demand points. *Information Systems and Operational Research, 38*, 359–372.

Drezner, Z., & Wesolowsky, G. O. (2001). On the collection depots location problem. *European Journal of Operational Research, 130*, 510–518.

Drezner, Z., & Wesolowsky, G. O. (2003). Network design: Selection and design of links and facility location. *Transportation Research Part A, 37*, 241–256.

Drezner, Z., & Wesolowsky, G. O. (2005). Maximizing cover probability by using multivariate normal distributions. *OR Spectrum, 27*, 95–106.

Drezner, Z., & Wesolowsky, G. O. (2014). Covering part of a planar network. *Networks and Spatial Economics, 14*, 629–646.

Drezner, Z., Wesolowsky, G. O., & Drezner, T. (1998). On the Logit approach to competitive facility location. *Journal of Regional Science, 38*, 313–327.

Drezner, Z., Wesolowsky, G. O., & Drezner, T. (2004). The gradual covering problem. *Naval Research Logistics, 51*, 841–855.

Drezner, Z., Wesolowsky, G. O., & Wiesner, W. (1999). A computational procedure for setting cutoff scores for multiple tests. *The Journal of Business and Management, 6*, 86–98.

Eiselt, H. A., & Laporte, G. (1995). Objectives in location problems. In Z. Drezner (Ed.), *Facility location: A survey of applications and methods* (pp. 151–180). New York, NY: Springer.

Eiselt, H. A., & Marianov, V. (2009). Gradual location set covering with service quality. *Socio-Economic Planning Sciences, 43*, 121–130.

Elzinga, J., & Hearn, D. (1972). Geometrical solutions for some minimax location problems. *Transportation Science, 6*, 379–394.

Farahani, R., Drezner, Z., & Asgari, N. (2009). Single facility location and relocation problem with time dependent weights and discrete planning horizon. *Annals of Operations Research, 167*, 353–368.

García, S., & Marín, A. (2015). Covering location problems. In G. Laporte, S. Nickel, & F. S. da Gama (Eds.), *Location science* (pp. 93–114). Heidelberg: Springer.

Glover, F., & Laguna, M. (1997). *Tabu search*. Boston: Kluwer Academic Publishers.

Goldberg, D. E. (2006). *Genetic algorithms*. Delhi: Pearson Education.

Hansen, P., Jaumard, B., & Krau, S. (1995). An algorithm for Weber's problem on the sphere. *Location Science, 3*, 217–237.

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

Karatas, M. (2017). A multi-objective facility location problem in the presence of variable gradual coverage performance and cooperative cover. *European Journal of Operational Research, 262*, 1040–1051.

Kirkpatrick, S., Gelat, C. D., & Vecchi, M. P. (1983). Optimization by simulated annealing. *Science, 220*, 671–680.

Kuhn, H. W. (1967). On a pair of dual nonlinear programs. In J. Abadie (Ed.), *Nonlinear programming* (pp. 38–45). Amsterdam: North-Holland.

Kutta, W. (1901). Beitrag zur näherungweisen integration totaler differentialgleichungen. *Zeitschrift für Angewandte Mathematik und Physik, 46*, 435–453.

Lee, D.-T., & Schachter, B. J. (1980). Two algorithms for constructing a Delaunay triangulation. *International Journal of Computer & Information Sciences, 9*, 219–242.

Leonardi, G., & Tadei, R. (1984). Random utility demand models and service location. *Regional Science and Urban Economics, 14*, 399–431.

Love, R. F. (1972). A computational procedure for optimally locating a facility with respect to several rectangular regions. *Journal of Regional Science, 12*, 233–242.

Maranas, C. D., & Floudas, C. A. (1993). A global optimization method for Weber's problem with attraction and repulsion. In W. W. Hager, D. W. Hearn, & P. M. Pardalos (Eds.), *Large scale optimization: State of the art* (pp. 259–293). Dordrecht: Kluwer.

Okabe, A., Boots, B., Sugihara, K., & Chiu, S. N. (2000). *Spatial tessellations: Concepts and applications of Voronoi diagrams*. Wiley series in probability and statistics. Hoboken: Wiley.

Plastria, F. (1991). The effects of majority in Fermat-Weber problems with attraction and repulsion. *Yugoslav Journal of Operations Research, 1*, 141–146.

Plastria, F. (2002). Continuous covering location problems. In Z. Drezner & H. W. Hamacher (Eds.), *Facility location: Applications and theory* (pp. 39–83). Berlin: Springer.

ReVelle, C., Toregas, C., & Falkson, L. (1976). Applications of the location set covering problem. *Geographical Analysis, 8*, 65–76.

Runge, C. (1895). Über die numerische auflösung von differentialgleichungen. *Mathematische Annalen, 46*, 167–178.

Runge, C. (1901). Über empirische funktionen und die interpolation zwischen äquidistanten ordinaten. *Zeitschrift für Mathematik und Physik, 46*, 224–243.

Schöbel, A., & Scholz, D. (2010). The big cube small cube solution method for multidimensional facility location problems. *Computers & Operations Research, 37*, 115–122.

Shamos, M., & Hoey, D. (1975). Closest-point problems. In *Proceedings 16th Annual Symposium on the Foundations of Computer Science* (pp. 151–162).

Snyder, L. V. (2011). Covering problems. In H. A. Eiselt & V. Marianov (Eds.), *Foundations of location analysis* (pp. 109–135). Berlin: Springer.

Suzuki, A., & Okabe, A. (1995). Using Voronoi diagrams. In Z. Drezner (Ed.), *Facility location: A survey of applications and methods* (pp. 103–118). New York: Springer.

Sylvester, J. (1857). A question in the geometry of situation. *Quarterly Journal of Mathematics, 1*, 79.

Sylvester, J. (1860). On Poncelet's approximate linear valuation of Surd forms. *Philosophical Magazine, 20*(Fourth series), 203–222.

Voronoï, G. (1908). Nouvelles applications des paramètres continus à la théorie des formes quadratiques. deuxième mémoire. recherches sur les parallélloèdres primitifs. *Journal für die reine und angewandte Mathematik, 134*, 198–287.
Weber, A. (1909). *Über den Standort der Industrien, 1. Teil: Reine Theorie des Standortes. English Translation: On the Location of Industries*. Chicago, IL: University of Chicago Press. Translation published in 1929.
Weiszfeld, E. (1936). Sur le point pour lequel la somme des distances de n points donnes est minimum. *Tohoku Mathematical Journal, 43*, 355–386.
Wesolowsky, G. O. (1993). The Weber problem: History and perspectives. *Location Science, 1*, 5–23.
Wesolowsky, G. O., & Love, R. F. (1971). Location of facilities with rectangular distances among point and area destinations. *Naval Research Logistics Quarterly, 18*, 83–90.
Zangwill, W. I. (1969). *Nonlinear programming: A unified approach*. Englewood Cliffs, NJ: Prentice-Hall.

Chapter 7
Voting for Locating Facilities: The Wisdom of Voters

Mozart B. C. Menezes

7.1 Introduction

In this chapter we discuss the quality of decisions made by stakeholders with voting rights under some assumptions on their intention of maximising their own personal utility rather than the common good. In especial, we focus on decisions made by a particular social choice mechanism, Condorcet method, and how good are the voted decisions when compared to those of a central decision maker. Thus, this chapter brings to the front row the opposing views of centralised versus decentralised decision making. It makes a review of the state of the art in this subject, and highlights very recent results, in especial, the work presented in Menezes and Huang (2015), Drezner and Menezes (2014), and Menezes et al. (2016).

We consider herein that each voter and each candidate has a particular position and the closest candidate to the voter's position gets her vote. The context we develop our discussions and arguments below will be that of (desirable) facility location. As an inspiring example, consider that voters live in neighbourhoods and they all need to vote for the location of a new hospital. Candidate locations for the hospitals that are closer to the voters' neighbourhoods will have preference (the vote) as opposed to further away candidate locations. The implied assumption is that hospitals are desirable facilities to have nearby, possibly increasing the real estate value in its proximity. Thus, each voter has an implicit (or explicit) ranked list of all candidate locations, from the most preferred one (the closest) to the least preferred one (the farthest). The distance is a measure of (dis)utility in this case. A voting mechanism would then, based on those ranked lists, choose the winner through a democratic process. A central decision maker, on the other hand, would choose a

M. B. C. Menezes (✉)
NEOMA Business School – Rouen/Paris/Reims, Mont-Saint-Aignan, France

H. A. Eiselt, V. Marianov (eds.), *Contributions to Location Analysis*, International Series in Operations Research & Management Science 281,
https://doi.org/10.1007/978-3-030-19111-5_7

candidate location that minimises the sum of the weighted distances from the new hospital to all neighbourhoods. There are cases where the objective function that maximises the social welfare is not the weighted sum but something else. In this chapter we consider only problems where the "ideal" objective is the weighted sum.

The comparison between the chosen locations, by vote or by a central actor, is vital to this chapter. With some special care, one can go beyond traditional facility location problems. For example, a voter could have beliefs in several dimensions such as economic policy, social policy, human rights, and individual rights. The distance from the candidate to the voters' beliefs would give the disutility obtained by the candidate. The underlying assumptions used herein are: (1) for any candidate and any voter, there is always a way to quantitatively measure the utility the voter gets from the candidate, that is, there is a function mapping the candidate to a positive scalar for that voter; (2) voters know exactly their position and the position of each candidate in the domain, and, no less important, they know how to calculate the individual utility they get from each candidate; (3) a voter ranks candidates according to the utility the candidate brings to her, thus, acting in a selfish manner; (4) a central decision maker knows the position of each voter and the position of each candidate, and equally important, the utility that each voter would get from candidate; and finally, in the work herein, (5) all voters and the central decision maker use a same utility function when evaluating candidates.

We present results regarding theoretical worst-case performance bounds for choosing a solution through vote as opposed to via a central decision maker. We also focus on algorithms for finding voted solutions computationally when all voters' intentions are disclosed that allow us to understand beyond worst cases to average cases when instances are of realistic sizes. That is, we contrast the theoretical results and those originating from realistic settings. Finally, we discuss the implications of those results and point directions for future research. In this process we hope to give a good, although not comprehensive, review of the state of the art in this matter.

The wisdom of the crowd has been under the spotlight for a while (Galton 1907; Surowiecki 2004; Rich 2010). When asked to guess some exact numerical figures, under the right incentives, the average of the crowd's answers beats (i.e., gets closer to the correct answer) the best of the experts' answers most of time (see Surowiecki 2004, for numerous real experiments). This is a very important result and, besides making a good conversation in a party and be thought provoking, these are empirical observations with no guaranteed regularity on the outcome. Moreover, for those of us interested in decision making, saying that crowds are good in guessing is not the same thing as saying that they are good at decision making. For example, individual users of shared resources acting independently, and according to their own self-interest, may end up depleting or spoiling those same resources through their collective, and "greedy", action. That behaviour, which is contrary to the common good of all users, is known as the tragedy of the commons (Hardin 1968) with examples that abound.

When shifting from guessing to voting, three problems arise. First is the issue of self-interest, which is absent from the former but not from the latter in most cases. When a decision is made by vote, it is reasonable to assume that voters are interested in the outcome of the election and have strong incentives toward making decisions that maximise each voter's individual utility.

Second, voting processes have their own issues; since Arrow's (2012) work in the 1950s it is well known that voting methods have severe limitations leading to elections results not free from paradoxes such as winners not representing voters' top preferences, or even no outright winners (see Knoblauch 2001; Alós-Ferrer and Granić 2012; Balinski and Laraki 2014; Plassmann and Tideman 2014). Plassmann and Tideman (2014) show that these paradoxes may occur with different methods, and can either increase or decrease with the number of voters.

Hence, election planners need to invest in the hard task of identifying voting mechanisms that maximise voters' utility and simultaneously reduce chances of paradoxical situations. For starters, there is a broad portfolio of election methods and scoring systems to choose from (see, for example Balinski and Laraki 2014; Plassmann and Tideman 2014). Moreover, the quality of a method measured by how its results compare to an optimal, centralised solution by an "omniscient and benevolent dictator" (Mueller et al. 1972, p. 66) may vary depending on contingency issues such as number of candidates and voters (Arrow 2012; Plassmann and Tideman 2014; Drezner and Menezes 2014; Pritchard and Slinko 2006).

Third, previous studies have brought negative results on the quality of the voted solution when compared to that of the benevolent dictator. For example, when locating a facility in a network that minimises the average distance from the facility to each voter (the Median problem) using a network setting, the authors in Hansen and Thisse (1981) derive upper bounds on the ratio between the Median objective function at the voted solution and the optimal solution. They proved that when the voting method used is the Condorcet method—a major voting method as discussed in Taylor and Pacelli (2008)—that ratio is bounded by 3. That is, when trying to minimise the distances from all voters to the facility, the distance induced by selfish behaviour can be three times higher than the one obtained by the benevolent dictator. Not an encouraging information about the quality of voted decisions. That gap between the objective functions brought by the democratic solution and that of the benevolent dictator could be referred to as the price of democracy, as the concept of "benevolent dictator" is theoretically nice by providing a best-case scenario (i.e., a bound) but unfortunately such a creature does not seem to exist.

Given the three problems mentioned above, in special the latter, one is tempted to stop short of further investigating this instigating issue and concede that a worst-case bound of 3 is a sufficient reason to move on from this topic. Nonetheless, if one removes some of the strong assumptions that make easy to construct contrived instances, then that worst-case bound drops an order of magnitude. Furthermore, by relaxing those somewhat restrictive assumptions, voters deciding through voting processes are, statistically speaking, excellent decision makers, with solutions of quality nearly indistinguishable from centralised systems, including expensive and sophisticated computer systems. These results, albeit interesting by themselves from the theoretical point of view, ask for further analysis and, perhaps, advocacy for using more frequently social choice mechanisms as decision-making processes. It should not be a bad idea to bypass dictators, benevolent, or not.

This chapter is organised as follows: first we review and explain selected voting mechanisms, on Sect. 7.3 we discuss the Weber and the Condorcet problems. Section 7.4 focuses on the relative error worst-case bound, and Sect. 7.5 presents a discussion on numerical methods. Section 7.6 discuss the use of different norms as proxy for utility functions, and we conclude the chapter in Sect. 7.7.

7.1.1 On Voting Methods

In order to start advocating for direct and universal suffrage, we first need to understand what the voting methods in discussion are and how they function: the rules of the game.

Voting methods are set rules indicating how voters' choices, or preferences, are drawn out and ordered to selecting winners (Sinopoli 2000). In a multidimensional spatial setting, being a network or continuous space, a voter's preference for a candidate is assumed to be directly proportional to their proximity, where the different dimensions may represent different utility criteria (Merrill-III 1985; Gouret et al. 2011; Henry and Mourifié 2013). The distances are weighted according to the importance given by voters to each dimension. Ideally, the voting method of choice would elicit a clear winner that maximised the products of voters' preferences and weights along each utility dimension (Bhadury et al. 1998).

In the Condorcet method, candidates are compared pairwise (Taylor and Pacelli 2008). Each voter provides a single ballot with an ordered list of preferred candidates creating rankings. A candidate W beats candidate L when there is a larger number of voters having candidate W higher ranked than L is in their ranked lists or ballots. The (Condorcet) winner is a candidate who beats all other candidates. If there are ties, then the winner is the candidate who beats all others after ties are removed or, equivalently, the candidate who does not lose to any other candidate. When after removing ties more than one candidate meets the criterion, then the result is decided randomly. Note that when distances are taken from real numbers then ties are unlikely to happen. There is a possibility to not have a winner, which is known as the Condorcet paradox (Taylor and Pacelli 2008; Balinski and Laraki 2014; Plassmann and Tideman 2014).

The Condorcet method has most of the desirable features that one expects from a social choice mechanism. These features are:

- Monotonicity: if a candidate choice W is a winner choice then by moving the choice to a higher ranking in the ranked list of one or more voters and keeping everything else the same, then W should be a winner if the voting process is repeated.
- Pareto: if W is higher than L on every voter's ballot, then L cannot be a winner.
- Independence of irrelevant alternatives: Assume that in a vote process W is declared a winner and L is not. If some voters change ranks of one or more candidates keeping the relative position between W and L in each list (i.e.,

whether W is higher- or lower-ranked than L), then candidate L would not be a winner in a repetition of the voting process.

- Condorcet winner criterion: if W is a candidate that the majority of voters prefer over other candidates, then W should be the winner.
- Always a winner: after computing all ballots the voting mechanism will always appoint a winner.

Arrow (2012) has shown that if a social choice procedure has no dictators imposing her will, then there is no system with three or more candidates that guarantees that the ranked preferences of individuals will map into a winner candidate while meeting the desirable conditions of monotonicity, Pareto efficiency, and independence of irrelevant alternatives.

The Condorcet method has many advantages but does not guarantee a winner—see Taylor and Pacelli (2008) for more on the subject. For example, using the notation "$X \succ Y$" means that X is preferred to or higher ranked than Y, if 8 voters have the following ballot ($A \succ B \succ C$), 9 have preferences ($C \succ A \succ B$), and 10 ($B \succ C \succ A$), then we can find that in a pairwise comparison: A would get 17 votes while B would get 10 votes so A beats B; B beats C in a direct confrontation with 18 votes; and C beats A with 19 votes against 8 vote on A's favour. In this case, there is no candidate that beats all other candidates in a pairwise comparison, which is known as the Condorcet paradox.

Recall that the Condorcet (solution) winner is the candidate that beats all other candidates in a pairwise comparison. It is possible that there is no Condorcet winner because the minimum for each candidate is less than one half. We assign the winner to the candidate whose minimum number of votes is the highest; this modified criterion is called a Simpson winner (a Simpson solution). We discuss this modified version of the Condorcet method further down in this chapter.

Given the well-known failures, of voting methods (Balinski and Laraki 2014), several studies compare their relative performance within different contexts. For example, as we present in this chapter, Klamler and Pferschy (2007) relate a classical optimisation problem with voting results comparing voting preferences for local versus global "tours" (i.e., traveling through different nodes of a network) using Approval Voting, Borda, Plurality, and Simple Majority rules, with mixed results—for detailed explanation on each of these voting mechanisms, please refer to Taylor and Pacelli (2008). Using methods similar to those just mentioned, plus Relative Utilitarianism, Buenrostro et al. (2013) investigate necessary levels of agreement between voters and leading to dominance-solvable games, suggesting that Approval Voting had the best results.

Many studies focus in particular on the ability of a method to yield a Condorcet winner. For example, Merrill-III (1985) performs a simulation study of six methods to estimate the fraction of outcomes coinciding with the Condorcet winner. His statistical analysis uses both L_1 and L_2 norms, which supports our interest in this chapter in applying different norms in our matrix of (dis-)utility.

Brams and Fishburn (2001) compare results of the 1999 elections for President and Council of the Social Choice and Welfare Society using Borda count and

Approval Voting. Results of the former are somewhat inconclusive: although the Borda count coincides with the Condorcet winner, they still see merits with Approval Voting due to the small distance between first and second placed candidates. More recently, Buechel (2014) finds that an optimisation approach based on the Median problem should match the Condorcet winner. The study presented here on explores this relationship in-depth via numerical experiments in a more general setting.

Despite the important contributions of the studies above, many questions regarding comparative advantages of methods remain unsolved. This occurs even with commonly reviewed methods such as Condorcet and Approval Voting. In particular, we want to answer some interesting questions that bring to light the efficacy of democratic approaches lead by self-interested voters as problem-solving mechanisms when compared to central decision makers, also known as *benevolent dictators*, maximising the total utility when solving those same problems; after all, it is very hard to present examples of widely accepted benevolent dictators (some don't believe in their existence), but it is very easy to find voters exercising their voting rights in a myopic fashion. Our interest is in answering questions such as: How bad can solutions be when originating by voting mechanisms in general and Condorcet method in particular? When democratic methods are as effective as optimal solutions for Median and Maximum Coverage problems? Is there a worst-case bound for the performance of a democratic method compared to an optimal solution originating from a benevolent dictator? Recall that a worst-case bound of 3 has already been mentioned above. How does voting methods compare when using realistic-sized instances? Do parameters such as distance norm and number of candidates influence the quality of a democratic method as compared to a centralised method? Therefore, we need to understand better the facility location problems.

7.2 On Facility Location and Voting Problems

The comparison between the Weber solution (the Weber problem, a well-studied facility location problem, Francis et al. 1992; Love et al. 1988; Drezner et al. 2002) and voted solution has attracted interest in the past. When the problem uses either the ℓ_∞ or the ℓ_1 norms, Wendel and Thorson (1974) have shown that the Condorcet solution point is the Weber solution point. Rodríguez and Pérez (2008) present an algorithm for finding Condorcet solutions when the solution is not a singleton but a set of a given cardinality. For special topologies (i.e., tree network) Noltemeier et al. (2007) provide insights on the case of single voting location problems on trees, and optimisation algorithms are developed for the Condorcet and Simpson cases. As algorithms were developed, with focus on network frameworks, there was also the curiosity about the quality of the solution set when compared to a central decision maker.

Demand points (voters) prefer a closer candidate (facility) to a farther one, making distance a proxy for (dis-)utility. The Condorcet solution is the result of the democratic process. The Weber solution, on the other hand, is the result obtained by a central decision maker (benevolent dictator) who selects the location based on complete information. The leading question is whether democratic processes lead to high-quality solutions. We measure the quality of the Condorcet solution by evaluating the Weber objective function, at the Condorcet location, and comparing that value to the optimal Weber objective function value.

The choice of measure for quality is the ratio between the two values of the Weber objective which we call relative efficiency (RE) of the Condorcet solution. Similarly, we can use another voting process and calculate the relative efficiency by changing the numerator to reflect the new voting process. For example, Approval Voting, a voting mechanism that we will also explore further below, specifies a critical distance (or dis-utility) so that voters approve candidates within that critical distance. Baron et al. (2005) studied parochial voting and found that when the critical distance is short, only candidates with high utility to the voter or their group will be approved, and when it is long, candidates with greater average utility to all voters may also be approved.

The relative efficiency ratio is greater than or equal to one by the optimality of the Weber solution. If this ratio is close to one, the Condorcet solution is almost as good as the optimal Weber solution and is considered to be of "high quality".

In Hansen and Thisse (1981), the authors frame the problem on a network and present a worst-case bound where $RE \leq 3$. That is, in a network setting, the quality of the voted solution is very bad compared to that originating from the centralised system; moreover, they show an instance on which the bound is tight. This implies that average distance returned through voting (using the Condorcet method) can be as bad as three times that of returned by the Weber solution. Further analysis in Labbe (1985) leads to the derivation of a worst-case ratio, also bounded by 3, as a function of the total population on the network. Bandelt and Labbe (1986) investigate the worst-case bound for the Simpson solution instead of the Condorcet solution.

Menezes and Huang (2015) provide theoretical results for the quality of the Condorcet solution as a function of the solution set cardinality. They find that the relative efficiency of Condorcet solutions approaches 1 as the number of potential locations increases, and the relative efficiency of a Condorcet solution is $\sqrt{2}$ in the worst case. Departing from previous models where the problems were framed in networks, they use a continuous model and the Euclidean distance (ℓ_2 norm) as proxy for dis-utility. We will discuss the findings brought up by Menezes and Huang (2015) in detail and how it further motivates the studies on this matter.

Note that the Condorcet method solution is closely related to competitive location models, where customers "vote" for the best retail facility by patronising it, where best could be the lowest total cost for service as in Hotelling (1929) or just some other criteria of attractiveness. In competitive facility location, depending on the assumptions about consumer behaviour, there are two main branches. One employs the gravity model suggested by Reilly (1931) and applied to a competitive retailing environment by Huff (1964, 1966). Locating a facility by this model was first introduced by Drezner (1994b). In the other branch, which we will focus hereon, customers patronise the facility with highest utility (Hotelling 1929; Drezner 1982; Hakimi 1983, 1986), which in the case of Hotelling (1929) includes the price of service/goods. Attractiveness impacts the customer selection of a facility to patronise, which implies that the closest facility may not be the one getting the demand (see Drezner 1994a). For a review of competitive facility location, the reader is referred to Berman et al. (2009) and, more recently, Eiselt et al. (2015).

Rodríguez and Pérez (2000) present a variation of the classic approach to Condorcet problems in location analysis through the introduction of a (in-)sensitivity factor for voters when the difference between the distances to two candidate locations is less than a threshold value of α. This may be the correct take when utilities are almost equal to one another. Another way to model such situations is by allocating their votes depending on the relative distances to the two competitors. If the ratio between the distances is large, almost all voters vote for the closer candidate. If distances are almost the same, the votes are divided more evenly among the competitors. Such an approach is similar to the gravity model in competitive facility location (Reilly 1931; Huff 1964, 1966; Drezner 1994a).

Drezner and Menezes (2014) focus on numerical approaches, in an attempt to better understand the result presented in Menezes and Huang (2015). In that work the authors perform sensitivity analysis on the variety of voters' utility functions, the number of candidates, and different topology, or level of "voter concentration".

Menezes et al. (2016) follow up on the suggestions made in Drezner and Menezes (2014) and expand the scope in three directions. The first is the analysis of different distance norms, pushing beyond the Euclidean norm, used previously. The second direction is the inclusion of Approval Voting as choice mechanism and brings, the third direction, the comparison between Approval Voting and Maximum Coverage Problem. In Approval Voting, each voter makes an unranked list of candidates they approve. The candidates with the most approvals win. If ties occur, they are broken randomly (Brams and Fishburn 1978; Rapoport and Felsenthal 1990; Brams and Fishburn 2007; Taylor and Pacelli 2008).

Borrowing from Menezes et al. (2016), we propose the following leading questions to lead this chapter:

- What is the worst-case bound for the Condorcet solution at the optimal Weber point when distances are free from the network topology?
- How frequently is the Condorcet solution at the optimal Weber point?
- What is the average quality of the Condorcet solution?
- How important is the Euclidean norm assumption to our results? Would they hold under another norm?
- What is the impact of restricting the number of candidate solutions to both the Weber and Condorcet problems?
- What is the impact of node aggregation on the relative efficiency?
- Is the relative efficiency of the democratic approach comparable to the centralised approach?

The last question is, of course, of a very general nature, but important to be discussed from the philosophical perspective. Thus, let's start by looking into the formulations that lead our discussion so far.

7.3 The Weber and the Condorcet Problems

7.3.1 The Weber Problem

The problem is framed using sets N and M of cardinalities n and m, respectively. N is the set of demand points, while M is that of candidate solutions. Each node $i \in N$ has an associated weight w_i representing the fraction of voters in i. That is, $\sum_{i=1}^{n} w_i = 1$. The distance between $i \in N$ and $j \in M$ is d_{ij}. The Weber problem does not consider the competing candidates. The objective of the Weber problem (Francis et al. 1992; Love et al. 1988; Drezner et al. 2002) is minimising

$$W(X) = \sum_{i \in N} w_i d_i(X) \tag{7.1}$$

where $d_i(X)$ is the distance between demand point i and point X anywhere in the plane. Hence, $W(X)$, called the Weber objective, brings the value of the Weber objective at X. If X^W is a point in the plane that minimises the Weber objective, then we say that X^W is a Weber solution.

The Weber problem is one of the most researched problems in location analysis. Solving it can be done by making use of the Weiszfeld algorithm (Weiszfeld (1936) translated by Plastria (Weiszfeld and Plastria 2009)) as accelerated by Drezner (1996).

7.3.2 The Condorcet Problem

The Condorcet solution is defined by comparing all pairs of candidates where the winner is a candidate who beats, or at most ties, all other candidates in a head-to-head competition. A voter at node i prefers candidate X over candidate Y if $d_i(X) < d_i(Y)$; that is, point X is the closest of the two to node i. In order to calculate a Condorcet objective, we need to compute the minimum number (fraction) of votes v_j for $j \in M$ against all competitors one by one. Thus,

$$v_j = \min_{k \neq j \in M} \left\{ \sum_{d_{ij} < d_{ik}} w_i \right\} \tag{7.2}$$

A candidate located at $j \in M$ is a Condorcet winner (solution) if and only if $v_j \geq v_k$ for all $k \in M$ and $k \neq j$. This is equivalent to $v_j \geq 0.5$. If there is no $v_j \geq 0.5$, then the maximum among all v_j (for all $j \in M$) is termed a Simpson solution and we may use that solution when desirable—of course, we will call attention to any use of Simpson solution in substitution of a Condorcet solution.

When set M of candidates is finite, when distances are drawn from the set of real numbers it would be unlikely to have ties, but possible. That is, when preference between candidates is governed by distances, ties in distances may exist, especially when M is the set of all points in the plane. We suggest two possible rules to deal with ties between distances: (1) If there is a tie, all tying competitors lose that vote. That means to apply $d_{ij} < d_{ik}$ in (7.2); or, (2) if there is a tie, all tying competitors win that vote. That means to apply $d_{ij} \leq d_{ik}$ in (7.2).

Let X to be a candidate anywhere in M. The Condorcet objective function, to be maximised, depends on the rule being used. Let $F_r(X)$ be the objective function when Rule $r \in \{1, 2\}$ is applied. The formulations are:

$$F_1(X) = \min_{k \in M} \left\{ \sum_{d_i(X) < d_{ik}} w_i \right\} \tag{7.3}$$

$$F_2(X) = \min_{k \in M} \left\{ \sum_{d_i(X) \leq d_{ik}} w_i \right\} \tag{7.4}$$

7.4 On the Worst-Case Bound

In this section we discuss the worst-case bound for the relative efficiency, or relative error (RE) of the Condorcet solution. We will only provide the sketch of the proof for the worst-case bound, a complete analysis can be found in Menezes and Huang (2015).

Let's call $X^C \in M$, for M being any point in the plane, to be the Condorcet winner, when one exists, and X^W the Weber solution. Then, $RE = W(X^C)/W(X^W)$. Use Fig. 7.1 as a visual aid to this discussion where each disc represents a node in set N with a positive weight and two labeled nodes are solid circles. Given the set N of nodes on the plane and their coordinates, we draw a line l that passes through point X^C and is perpendicular to the line passing through X^C and X^W. Let H_L be the closed half-plane on the left side of line l (including points on line l itself) and H_R be the open half-plane on the right side of line l (not including points on line l). The situation just described can be seen in Fig. 7.1a.

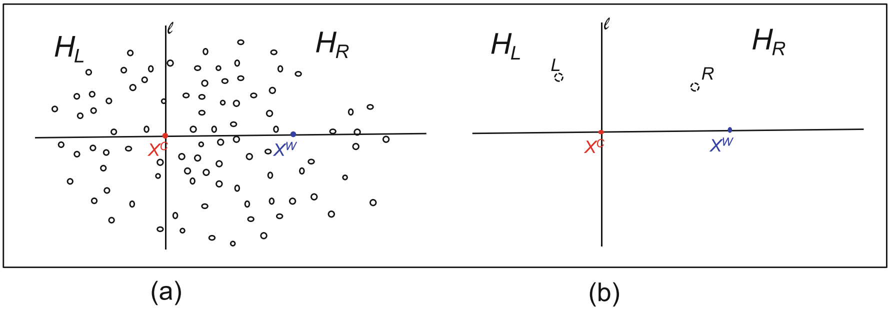


Fig. 7.1 (**a**) Line l divides the plane into two half-planes. (**b**) Surrogate nodes L and R represent aggregation of nodes in each half-plane

We state without formal proof that

Property 7.1 The total population in the closed half-plane H_L is no less than $1/2$, i.e., $\sum_{i\in N\cap H_L} w(i) \geq 1/2$.

The intuition behind Property 7.1 is that a point located just next to the right of X^C in Fig. 7.1a would beat X^C in a face-to-face competition if Property 7.1 were not true.

In Fig. 7.1b, we define the equation system $\sum_{i\in H_L} w_i\, d(L, s) = \sum_{i\in H_L} w_i\, d(i, s)$, for $s \in \{X^C,\ X^M\}$. The solution of this system of equations defines point L. Using the same arguments we can define a point R that could represent all nodes on the right half-plane. The proof for the existence of both points L and R can be found in Menezes and Huang (2015). Using the triangle inequality, one can show that $d(L, X^W) \leq d(L, X^C) + d(X^C, X^W)$.

Furthermore, one can also prove that $d(R, X^C) \geq d(R, X^W)$, which is done based on the fact that $W(X^W) \leq W(X^C)$ by the definition of the Weber objective, and also by Property 7.1.

Using $d(R, X^C) \geq d(R, X^W)$ and $d(L, X^W) \leq d(L, X^C) + d(X^C, X^W)$, and the knowledge that $d(i, X^C)+d(i, X^W) \geq d(X^C, X^W)$ for any $i \in H_R$, it is possible to carefully manipulate $RE = \frac{W(X^C)}{W(X^W)}$ until the result

$$RE \leq \frac{d(L, X^C) + d(R, X^C)}{d(L, X^W) + d(R, X^W)}, \text{ follows.} \tag{7.5}$$

In order to prove the main result, one needs to inspect two cases that make a difference. On the one hand, point R's orthogonal projection onto the line (X^C, X^W) is coinciding or to the left of point X^W and, on the other hand, that projection is to the right of it. We will show the bound for the first case and we leave to the reader to consult the original proof for the complete discussion.

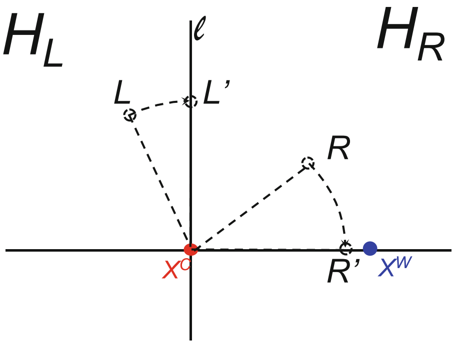


Fig. 7.2 Points L and R are projected on line l and on its orthogonal, respectively, generating L' and R', also respectively. The projections are made such that the distances of both points to point X^C remain unchanged

Using Fig. 7.2, it is straightforward to verify that, following expression (7.5),

$$\frac{d(L, X^C) + d(R, X^C)}{d(L, X^W) + d(R, X^W)} \leq \frac{d(L', X^C) + d(R', X^C)}{d(L, X^W)} \leq \frac{d(L', X^C) + d(R', X^C)}{d(L', X^W)}$$
$$\leq \frac{d(L', X^C) + d(R', X^C)}{d(L', R')} \tag{7.6}$$

Hence,

$$RE \leq \frac{d(L, X^C) + d(R, X^C)}{d(L, X^W) + d(R, X^W)} \leq \frac{d(L', X^C) + d(R', X^C)}{d(L', R')} \leq \sqrt{2}, \tag{7.7}$$

where the last inequality is guaranteed by the fact that $\triangle(X^C L' R')$ is a right triangle—for a geometric proof of last inequality see Posamentier and Lehmann (2012).

The other case, when R' is to the right of X^W, is dealt with through small changes in the manipulations leading to the same conclusion, implying that

Property 7.2 $RE \leq \sqrt{2}$.

7.5 Numerical Methods

7.5.1 Finding the Condorcet Solution Point

When the Condorcet solution belongs to a discrete set of potential locations; i.e., M is a discrete and finite set and $X^C \in M$, then finding the best Condorcet solution point is straightforward and can be done by brute force. Evaluate all $X \in M$ using $F_r(X)$ for either $r = 1$ or $r = 2$ depending on the rule used and keep the one that maximises that objective. The search process requires $O(mn)$ time and can be reduced to $O(m \log n)$ when the distance matrix is pre-computed.

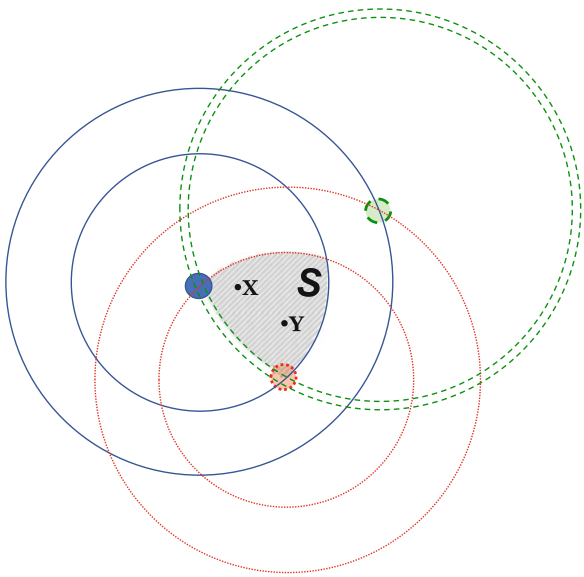
When M can be any point in the plane, then the so-called Absolute Condorcet Solution is found through more sophisticated search procedure. Let's start that explanation by supposing that m competitors exists, which for simplicity coincide with the demand points. The distance between demand point i and the m competitors named $1, \ldots, m$ is given by $d(i, j), \quad j \in \{1, \ldots, m\}$, and is sorted such as $d(i, 1) \leq d(i, 2) \leq \ldots \leq d(i, m)$.

For each demand point i draw m concentric circles of radii $d(i, j)$, for $j = 1, \ldots, m$, that form $m+1$ regions: the interior of a circle of radius $d(i, 1)$, $m-1$ rings bounded by inner radius of $d(i, j)$ and outer radius of $d(i, j+1)$ for $j = 1, \ldots, m-1$, and the exterior of the circle of radius $d(i, m)$ (see also Wendel and Thorson 1974). For each demand point there are $m + 1$ such regions. The intersection of all these regions partitions the plane into regions bounded by arcs of circles resembling polygons—see Fig. 7.3 as visual aid in which a convex polygon, labeled S, is shaded and two points (X and Y) are inside S. The three competitors and the three demand points coincide. Each such region is the intersection of some interiors of circles and exteriors of some other circles.

Let a *circular polygon* to be defined as the region enclosed by the intersection of interiors or exteriors of circles with no other circle cutting through that region. The sides of a circular polygon can be convex or concave (see Fig. 7.3) and thus, the circular polygon itself can define a convex set (or not). Note that S is a convex circular polygon.

An example of the intersection diagram of all circles is depicted in Fig. 7.3. There are three points that define both demand and competitors' location. Each demand point defines two concentric circles centred on it and passing through the other two demand points for a total of six circles. There are many circular polygons defined

Fig. 7.3 Diagram of intersecting circles



by these circles, and in general some are convex but most are not. There is only one convex circular polygon in Fig. 7.3 denoted by S. Drezner and Menezes (2014) show that convex circular polygons must be in the convex hull of the demand points. Moreover, Drezner and Menezes (2014) prove that

Property 7.3 When using Rule 1 (induced by $F_1(\bullet)$) the Condorcet solution is an interior point of a convex circular polygon, and when using Rule 2 (induced by $F_2(\bullet)$), then the optimal solution solving the Condorcet problem is on a vertex of a circular polygon.

In order to understand Property 7.3, allow X and Y as two potential locations for a new competitor. If X and Y are inside a same circular polygon (as Fig. 7.3 shows), then for all voters in any point i, $d(i, X)$ and $d(i, Y)$ are in the same range between two consecutive distances in that ordered list, and, thus, a voter at i votes the same way for a new competitor located either at X or at Y (or to any competitor at any point within the convex polygon). The objective function values at X and at Y are the same when $d(i, X)$ and $d(i, Y)$ are in the same range for all $i \in N$.

Furthermore, note that if the distances from the demand points (voters) to two potential locations X and Y are inside or outside the same circles and if one is on the circumference of a circle so is the other one, then $F_r(X) = F_r(Y)$ for Rule $r \in \{1, 2\}$.

On the other hand, consider two adjacent polygons separated by an arc induced by some demand point i. Point X is in the interior of a circular polygon and point Y in the other. That means that a path from X to Y crosses an arc separating the two circular polygons and when moving from X to Y one inequality involving demand point i changes, implying that one weight is either lost or gained by the new competitor. On the arc separating them the weight is lost and thus points located on the arc are inferior to points in the circular polygon on the convex side of that arc. Hence, when Rule 1 is applied and the set S is not empty, the value of the objective function is the same at any point in the interior of S.

However, when Rule 2 is used then points on the arcs themselves become more relevant. Consider, as before, two adjacent circular polygons and points X and Y in the interior each. The path from X to Y crosses an arc separating the two circular polygons, and **on the arc** separating them the weight is counted and thus points on the arc cannot be inferior to points in the interior on the convex side of the arc. On a vertex all weights associated with tying distances may be added to $F_2(X)$ and thus are not inferior to non-vertex points. That insight leads us to say that when Rule 2 is employed, an optimal solution exists on a vertex of a circular polygon. That implies in the existence of a finite set of candidate points when Rule 2 is employed.

The same argument employed above can be used to demonstrate that if the circular polygon has a concave arc there is a better or equal solution on a point outside that polygon. Furthermore, if the circular polygon is convex, then an optimal solution exists on the set of vertices of the convex circular polygon.

Property 7.4 An optimal solution by Rule 2 exists on a vertex of a convex circular polygon.

Drezner and Menezes (2014) use the result that for a convex norm, assuming X be a point outside the convex hull (CH) of the demand points, there exists a point $Y \in CH$ such that $d_i(Y) \leq d_i(X)\ \forall\, i$. The proof of it is given in Wendell and Hurter (1973). The result by Wendell and Hurter (1973) leads to the important property that we present without proof.

Property 7.5 For a convex norm, an optimal solution to maximising $F_r(X)$ exists in CH.

The results presented in this subsection allow for solving the problem with efficient numerical approach. An improvement of the Big Triangle Small Triangle (BTST) technique makes it very efficient to find Condorcet solutions.

7.5.2 Numerical Experiments

Drezner and Menezes (2014) propose a mechanism to find the best Condorcet location X anywhere in the plane by the global optimisation technique Big Triangle Small Triangle (BTST) proposed by Drezner and Suzuki (2004). Complete details can be found in Drezner and Suzuki (2004) and Drezner and Menezes (2014), so we turn our attention to key insights unveiled by the approach.

The results presented below were programmed in Fortran using an Intel 11.1 Fortran Compiler. Only one thread was used on a desktop with the Intel 870/i7 2.93 GHz CPU Quad processor and 8 GB RAM. In order to calculate the relative efficiency for various situations the following solution points are found.

X^M	The optimal Condorcet solution in M.
X^W	The optimal Weber solution anywhere in the plane (Drezner 1996).
X^C	The optimal Condorcet solution anywhere in the plane (Drezner and Suzuki 2004)
X^N	The node with the lowest Weber objective function.

In this section several number of n demand points were tested. The range goes over 10 values of n from $n = 10, 20, \ldots, 100$ and then from $n = 100$ to $n = 1000$ values were increased by jumps of 100 demand points. For each value of n, 1000 randomly generated sets of points (nodes) were used to fill up the tables. Locations were uniformly generated in a unit square and weights were uniformly generated in the interval [0,1]. Weights were then divided by the sum of the weights so that the sum of all weights is equal to one.

When constraining the solutions (Condorcet and Weber problems) to nodes, when a Condorcet solution exits, more than 70% of the Condorcet solutions are at a node with the minimum Weber objective for all values of n. The same result holds for the combined count of both Condorcet and Simpson solutions. The result suggests that the node with the best Weber value is a natural candidate for the

Condorcet solution. A Condorcet solution may not exist only for instances where there are small number of nodes but, as the number of nodes increases, the fraction of Simpson solutions drops substantially and vanishes when $n \geq 200$. Hence, the Condorcet approach is very sensitive to aggregation and if the one person/one vote maxim holds, then the existence of a Condorcet solution is nearly certain in large groups. On the other hand, the likelihood of a Condorcet solution is reduced when set individuals are represented by a single voter.

When the Weber optimal point is added to the set of nodes, the number of Condorcet solutions at either the Weber point or the node with the minimum Weber objective is significantly higher than these values reported when only demand points are candidate locations. For values of $n \geq 100$ there is always a Condorcet solution and the relative efficiency of the Condorcet solution is very high since the Condorcet solution is almost always on either the Weber solution point or on the minimum Weber objective node.

The results above suggest that if the number of voters is sufficiently large (i.e., greater or equal to 100) there are Condorcet solutions. The issue is whether or not those solutions are efficient (high values of relative efficiency). In order to address that matter, Table 7.1 below presents RE, the ratio between the Weber objective at the optimal Condorcet solution and the Weber optimum objective value. The table reports the minimum, average, and maximum value of this ratio as well as the standard error of the average (i.e., the standard deviation divided by $\sqrt{1000}$).

Table 7.1 Ratio of the Weber objectives at Condorcet solution and Weber optimum

	$RE = W(X^C)/W(X^W)$			
n	Min	Ave	Max	SE
10	1.000002	1.018380	1.195200	0.000724
20	1.000000	1.006737	1.092617	0.000251
30	1.000000	1.004420	1.052084	0.000162
40	1.000000	1.002901	1.041940	0.000111
50	1.000004	1.002170	1.022142	0.000084
60	1.000001	1.001830	1.022557	0.000074
70	1.000002	1.001563	1.016696	0.000061
80	1.000001	1.001292	1.013861	0.000048
90	1.000003	1.001130	1.013073	0.000042
100	1.000002	1.001004	1.007504	0.000035
200	1.000000	1.000501	1.006499	0.000018
300	1.000000	1.000332	1.004933	0.000014
400	1.000000	1.000242	1.002998	0.000009
500	1.000000	1.000205	1.001870	0.000008
600	1.000000	1.000167	1.001327	0.000006
700	1.000000	1.000150	1.001170	0.000005
800	1.000000	1.000134	1.001233	0.000005
900	1.000000	1.000115	1.000999	0.000004
1000	1.000000	1.000116	1.001404	0.000004

For $n \geq 50$ the maximum gap is less than 3% and reduces to 0.1% as the value of n increases. On average, it is very close to 1 even for values of n as small as 10 nodes. Hence, even if a large group of voters is represented by a single voter, or if the number of voters is indeed small, the efficiency of the Condorcet solution is very good. On average it is as good as the optimised solution. Clearly, if the "right" candidate (represented by the Weber solution) is included in the set of candidate solutions, the Condorcet location is almost as good as the one offered by a central decision maker.

The result presented in Table 7.1 is an important one. Nowadays, abundance of computers for processing data and the easiness of data collection provided by the internet can make a powerful combination that would allow for making decisions using only votes from voters with a stake on a particular decision. The results suggest that if done so, the solution would be as good as one coming from a central decision maker when that decision maker is honest in his/her intention to promote the common good. Subsequent studies show that, even when the Weber solution is not included in the candidate-solution set, for values on $n \geq 100$, the relative efficiency $RE \leq 10\%$ on worst cases and less than 1% on average. Hence, even when the right candidate is not available and we restrict candidates to the set of nodes, with their own positions, the relative efficiency of the Condorcet solution compared to the solution brought up by the **unconstrained** Weber solution (a solution that is possible to be a point in the plane but not on the set of nodes) is very good when the number of voters is sufficiently large.

The quality of the Condorcet solution improves (in both average and maximum ratio) as the number n of nodes increases to a point of marginal difference to the Weber objective value at the Weber point. A similar trend can be observed on all experiments made, suggesting that aggregation is a factor that reduces the relative efficiency of the Condorcet solution. Therefore, we need to further understand the impact of aggregation.

Note that aggregation, in Facility Location discipline, can be just a way of simplifying a problem, making it more tractable, but creating negative side effects through what is known as "aggregation errors". That is, differences between the level of aggregation of choice and the correct solution when every unit of interest (e.g., a single person) is the focal point. See Francis et al. (2009) for a recent review of the topic. We assume herein that the level of aggregation of choice for the centralised solution is a correct (or acceptable) level and it is the baseline for the results presented.

Aggregation is the phenomenon of a large fraction of total voters with the same voting position. It is in fact an extreme case of concentration, i.e., a high density, or large fraction of voters, within a small area when contrasted to the density elsewhere in the plane. These clustering of nodes could occur but most likely a uniform distribution of nodes inside the 100×100 square will be observed.

The study conducted by Drezner and Menezes (2014) took the original 100 distance units "big" square and divide them into 100 smaller squares with side equal to 10 distance units. The first 50% of the nodes are located in the big square by a two-dimensional uniform distribution as before. The other half of the demand nodes are allocated to Ω smaller squares as described in Drezner and Menezes (2014):

1. Choose Ω different small squares with equal probability for each.
2. For $i = (n/2) + 1 \ldots n$ do
   (a) Choose randomly one of the Ω small squares.
   (b) Randomly locate demand point i in the selected square.

The experiments are conducted similarly to those reported in Table 7.1, with additional parameter $\Omega \in \{5, 10, 20, 50\}$. That approach allows for concentrating 50% of the demand in a small cluster defined by the small squares. The clustered instances give a different perspective by generating different demand distributions. As explained in Drezner and Menezes (2014), "When $\Omega = 5$, at least 50% of the demand points are in 5% of the total area. When $\Omega = 50$ only a slightly more clustered distribution of points is created. When $\Omega = 100$ demand points are uniformly distributed with no clustering structure".

Table 7.2 reports the average relative efficiency of the Condorcet solution for different combinations of values for n and Ω. We compare the values obtained through clustering with those previously reported in Table 7.1 in order to have a benchmark.

Table 7.2 Average ratios between the Weber objectives at the Condorcet solution and Weber optimum for clustered instances

	Ω				
n	5	10	20	50	Uniform
10	1.017852	1.017967	1.017776	1.019618	1.018380
20	1.009563	1.009607	1.008909	1.009398	1.006737
30	1.007072	1.006084	1.006765	1.007463	1.004420
40	1.005458	1.005164	1.005224	1.004732	1.002901
50	1.004216	1.004105	1.004218	1.004243	1.002170
60	1.003397	1.003546	1.003550	1.003263	1.001830
70	1.003349	1.003625	1.003678	1.003098	1.001563
80	1.002975	1.002647	1.003200	1.002758	1.001292
90	1.002553	1.002753	1.002435	1.002786	1.001130
100	1.002496	1.002576	1.002562	1.002463	1.001004
200	1.001305	1.001739	1.001384	1.001507	1.000501
300	1.001190	1.001124	1.001153	1.001151	1.000332
400	1.000973	1.001006	1.000967	1.000956	1.000242
500	1.000786	1.000796	1.000874	1.000831	1.000205
600	1.000763	1.000671	1.000708	1.000708	1.000167
700	1.000613	1.000662	1.000699	1.000660	1.000150
800	1.000643	1.000644	1.000616	1.000643	1.000134
900	1.000622	1.000608	1.000536	1.000587	1.000115
1000	1.000531	1.000564	1.000521	1.000532	1.000116

Table 7.2 shows a systematically dominance, although just slightly, of RE for average ratios for clustered instances when comparing them to those presented in Table 7.1 for uniform demand (reproduced in the last column of Table 7.2). As an example, when $\Omega = 5$ and $n = 10$ Table 7.2 shows that the ratio is 1.0178 which is smaller than 1.018380 reported in the last column. The results suggest that there is no significant impact of clustering, as designed in this experiment, on average ratios.

Similar to Table 7.2, Table 7.3 reports the maximum relative efficiency of the Condorcet solution for different combinations of values for n and Ω.

For values of n and Ω relatively small, Table 7.3 shows that the maximum ratio of the two cases presents significant differences. When $n = 10$ and $\Omega = 5$ the maximum ratio is 1.323328 while the maximum for uniform demand is 1.195200. Nonetheless, when n and Ω are larger, the differences reduce. For example, when $n \geq 100$ or $\Omega \geq 10$, the differences in maximum ratios are not significant which implies high quality of the democratic process' output even in the worst case.

When clusters are present in the data (via Ω), it is somewhat similar to representing each cluster by one demand point with the total weight. In overall, the work presented in Drezner and Menezes (2014) suggests that the impact of clustering is similar to reducing the number of nodes keeping the diversity of candidate solutions. Fewer number of demand points result in lower quality Condorcet solutions mostly because there are less candidates to choose from.

Table 7.3 Maximum ratios between the Weber objectives at the Condorcet solution and Weber optimum

	Ω				
n	5	10	20	50	Uniform
10	1.323328	1.254989	1.291172	1.405066	1.195200
20	1.169552	1.198324	1.179087	1.268093	1.092617
30	1.113245	1.121800	1.098894	1.146177	1.052084
40	1.093191	1.109401	1.100839	1.115800	1.041940
50	1.128883	1.102289	1.086627	1.113002	1.022142
60	1.074025	1.067920	1.061447	1.082114	1.022557
70	1.086514	1.077868	1.090157	1.045384	1.016696
80	1.078628	1.055699	1.070291	1.068326	1.013861
90	1.057687	1.053087	1.054615	1.078387	1.013073
100	1.088925	1.072415	1.055493	1.066915	1.007504
200	1.034408	1.044634	1.036045	1.051491	1.006499
300	1.026147	1.029063	1.029973	1.035478	1.004933
400	1.034228	1.032123	1.045251	1.032863	1.002998
500	1.018115	1.022193	1.028169	1.021884	1.001870
600	1.025932	1.014086	1.021058	1.033248	1.001327
700	1.018527	1.020085	1.017664	1.018127	1.001170
800	1.019652	1.028951	1.012851	1.021849	1.001233
900	1.027214	1.022277	1.021686	1.015512	1.000999
1000	1.017988	1.019875	1.017142	1.015455	1.001404

7.6 The Effect of Different Norms as Proxy for Utility Functions

Continuing with our quest for better understanding the difference (via our measure RE) of a decision made through Condorcet voting process and that of a central decision maker with full knowledge of all voters' utility functions, calling attention to result in Menezes et al. (2016). Using the same expressions for the Weber and Condorcet problems previously defined, we introduce a variation on the notation of function $d(i, j)$, the distance from a demand point $i \in N$ to a point $j \in M$, by adding an extra parameter p representing the norm utilised for calculating the distance.

So far we have used ℓ_2 norm; that is, ℓ_p for $p = 2$, which is the Euclidean norm. Now we turn our attention to different norms and will get support from Menezes et al. (2016) for learning the effect of norms on our insights. Consider two points i and j with positions X_i and X_j; the element wise difference $\chi_{ij} = (|X_{it} - X_{jt}| : t = 1, \ldots, u)$ in $\mathbb{R}^{u+}$ represents the distance from node i to candidate j in each one of the u dimensions. We define

$$d_p(i, j) = \sqrt[p]{\sum_{t=1}^{u} \chi_{ijt}^p};$$

that is, the ℓ_p norm.

When $p = 1$, the ℓ_1 norm is the sum of distances in all dimensions, equidistant points from a given reference in this norm define a diamond in contrast to a circle defined by the ℓ_2 norm. The ℓ_∞ norm is defined by $d_\infty(i, j) = \max\{\chi_{ijt} : t = 1, \ldots, u\}$. Its equidistant points define a square.

ℓ_p norms may have behavioural or cultural meaning in our context. For example, when voters are very concerned to the overall utility of a particular solution the ℓ_2 norm may be more appropriate. When each dimension has a significance above the overall utility assessed from all dimensions together then voters use ℓ_1 norm, that is, differences in each dimension add up. The Tchebychev norm, ℓ_∞, may be a good way of representing voters assessing the utility of a candidate based on the dimension that is the farthest from the voter's position. In loosely terms, we can say that the value p, in ℓ_p norm, captures how much importance the farthest of all dimensional distances is important. As the value of p increase, the higher the weight of the farthest dimensional distance is. It goes from $p = 1$, when all have the same weight, to $p = \infty$, when only the farthest one counts. See Eiselt and Sandblom (2014, pp. 166–167), for discussion on ℓ_p norms and "differences build up" in the context of competitive models.

Menezes et al. (2016) report that changes ℓ_p norm for values of $p \in \{1, 2, 3, \infty\}$, there is very little change on the relative efficiency. Note that the authors report a quality index $Q = 1/RE$ which implies in relative efficiency higher when the value of $Q \to 1$ from below. Let the triple $(\mu_Q^{(p)}, \sigma_Q^{(p)}, n^{(p)})$ stand for mean and standard deviation of the values of Q, and the number of instances solved respectively for a given value of p.

In those three cases, according to Menezes et al. (2016), $(\mu_Q^{(p)}, \sigma_Q^{(p)}, n^{(p)}) = (1.00, 0.01, n^{(p)})$ for $p \in \{1, 2, 3\}$; and $n^{(1)} = 481$, $n^{(2)} = 511$, and $n^{(3)} = 502$. Hence, Menezes et al. (2016) reinforce previous results on the quality of the Condorcet method when compared to the benevolent dictator's decision. Moreover, it says that at least for those three values of p, the ℓ_p norm does not impact the quality of the Condorcet solution.

Menezes et al. (2016) have also tested ℓ_∞ norm and $(\mu_Q^{(\infty)}, \sigma_Q^{(\infty)}, n^{(\infty)}) = (0.99, 0.02, 474)$. A regression analysis on the value of Q as function p delivers a regression coefficient of -0.081 (with statistical-p-value < 0.001). That is, a negative regression coefficient but with little practical effect on the quality of the solution.

Another interesting result from Menezes et al. (2016) is on the impact of dimensionality of the utility function used in the problem. The authors have tested location vectors with dimensionality $\{2, 3, 4\}$. And, in fact, the higher the dimensionality of the utility function, the better the RE, i.e., the quality of the Condorcet solution. When looking into the interaction between u and p, the authors report a positive coefficient for $p \times u$.

Menezes et al. (2016) also discuss the Approval Voting method and its relation to the Weber problem. We do not discuss it here.

7.7 Conclusions

We presented studies on the relative efficiency of the Condorcet solution when compared to the Weber solution. The relative efficiency, as defined in this work, is the ratio between the value of the Weber objective function at the Condorcet solution point and at the Weber (optimal) solution point. Previous work has shown that the theoretical ratio can be as high as 3 when the problem is defined on a network. However, the impact of topology is reduced when the problem is framed in the plane and using Euclidean distances. Experiments, not based on contrived extreme examples, were tested. Contrived instances are interesting from the theoretical point of view to obtain bounds, but in the studies herein, we rather represent common situations.

The experiments have shown that the Condorcet solution and the Weber solution have a high level of coincidence, especially when the size of the set of nodes is large. The experiments suggest that the number of solutions that satisfy the Condorcet criteria is low for small values of n.

The best values of relative efficiency are obtained when the "right" candidate is included. If the optimal Condorcet solution point obtained is included (the BTST method is used herein, Drezner and Suzuki 2004), then the relative efficiency of the Condorcet solution is 1 for many instances and on the average it is very close to one. The research has found that reducing the number of candidates does not affect critically the quality of the Condorcet solution. When changing the baseline for comparing the Condorcet solution to the node with the minimum Weber objective, then the already high-quality Condorcet solution improves further.

As long as the number of candidate solutions is large, the Condorcet solution is of high quality even when a different norm is used. In fact, aggregation seems to be the main issue determining the quality of the Condorcet solution. When the population is represented by a few voting representatives the quality of the solutions deteriorates slightly, but when the reduction is on the number of potential candidates then the reduction in quality may become dramatic.

Worst-case scenarios are more likely to happen when both the number of candidates and that of the voters, are reduced suggesting that using surrogates each representing a large number of voters, like members of Congress, is inferior to systems that rely heavily on direct voting schemes. When the number of different viewpoints (candidates' profiles) is limited, such as in a bi-partisan model, then the quality can be further decreased. Thus, a referendum of all voters, with plurality of candidates, better predicts the will of the people than a vote in Congress or Parliament. That brings the need of further looking into the efficiency of two-party runs and voting mechanisms that increase relative efficiency in the case when they are adopted.

In the USA, elections of the president are a two-stage process. First, an electoral college is elected through the election of members of the electoral college in each state where, in most cases, all members elected in a state support the same candidate for president. Then, the electoral college elects the president. There are supporters for the abolishment of the electoral college and elect the president by direct vote. Reasons cited for such a change are that a president can be elected with fewer votes than his opponent as it has recently happened when President Donald Trump was elected with a smaller number of total "direct" votes received by his opponent Hillary Clinton. Additionally, there is an issue of incentive to vote since some states have a clear majority of voters supporting a certain candidate. Voters in such states do not influence the outcome and are less likely to vote. Voters in states where support for each candidate is about the same are more likely to affect the outcome and thus are more likely to vote. Every vote counts the same when elections are determined by plurality. Our results support such a change. Furthermore, a shift to direct vote would also result in different ways presidential campaigns are conducted by the candidates possibly affecting the relationship between candidates and some minority groups.

Clearly, there are reasons for an electoral college process and its existence has supporters. The main obvious reason for it is to reduce the weight of states highly populated in an election, in comparison with those less populated. According to Schulman (2018), "The Electoral College was created for two reasons. The first purpose was to create a buffer between population and the selection of a President. The second as part of the structure of the government that gave extra power to the smaller states". The author claims that the founding fathers were afraid that a tyrant could manipulate public opinion and come to power. The fear of manipulation was certainly more grounded in the time of the founding fathers when someone in a rural area of the mid-west was certainly less informed than someone in a larger city, that fear, based of lack of access to information, is less real today. In any case, the mechanism that could reduce the risk of manipulation may also bring the risk of other types of manipulations.

The insights presented herein were not affected in the presence of clustering; that is, when voters are grouped in the space discussed (the square) on a non-uniform manner. The ratio between the Weber solution and the Weber objective function at the Condorcet solution is similar to that found when there is no induced clustering *ceteris paribus*. Even when half of all nodes are located in 5% of the total area, which implies in few very dense regions, as could also be the case in real situations, all insights discussed above hold with absolute differences in the ratio appearing only beyond the second decimal place. In our work, we, the co-authors, note that clustering impacts the variability of the quality of the solution obtained. Our results show that the worst-case ratio is always inferior when clustering is present than when it is not.

This line of research can move in the direction of addressing "obnoxious" decisions (in the context of obnoxious facility location problems) for comparing quality of results originating from centralised systems and voted decisions. This particular line of research may have important practical application: in a political system, elected officials are future candidates on next elections. As such, officials avoid decisions that are reasonable but could lead to a "political suicide" that is, making a logical decision that, although right, will not satisfy voters' utility function and, therefore, not being elected later. However, if an official knows that a voted decision is optimal, or near-optimal, when compared to the reasonable decision the political should make, then allowing the decision to voters is a win–win approach: get the optimal solution without exposing herself to the risk of committing a political suicide.

Another direction of research to be investigated is when voters cannot precisely pinpoint the position of a candidate. It may be the case that candidates do not pass clearly their standpoints, either intentionally or not, and voters have some uncertainty about the impact of a candidate in his utility function.

The converse of the idea above is the positioning of a candidate when the exactly location of some groups of voters is not well understood. In other words, candidates do not fully understand the response of voters to some particular position taken by a candidate. If a candidate wants to find the position that maximises her chance of winning, then the stochastic problem gets interesting.

Also, the experiments, similar to those presented herein, can be expanded to multiple facilities. In this case, different ways of valuing multiplicity of facilities need to be carefully defined.

The model discussed herein can be easily generalised to higher dimensions. Unfortunately, the numerical models presented on cited papers are developed to work on the plane limiting their application to two possible dimensions. Problems defined on multi-dimensions, beyond the limit of 2, can better reflect the true political scene in which several considerations affect the political decision-making process making it an interesting line of research to pursue.

Another issue to investigate is related to the relationship in the way social influence affects purchasing decisions and consequently market share. For example, when one clicks on the symbol "Likes" presented in some social network websites such as Facebook, then the "like" will influence other people perceptions on the object. Social networks may bias the underlying distribution of demand for products and services.

References

Alós-Ferrer, C., & Granić, D. (2012). Two field experiments on approval voting in Germany. *Social Choice & Welfare, 39*(1), 171–205.

Arrow, K. J. (2012). *Social choice and individual values* (3rd ed.), New Haven, CT: Yale University Press.

Balinski, M., & Laraki, R. (2014). Judge: Don't vote. *Operations Research, 62*(3), 483–511.

Bandelt, H. J., & Labbe, M. (1986). How bad can a voting location be. *Social Choice & Welfare, 3*, 125–145.

Baron, J., Altman, N. Y., & Kroll, S. (2005). Approval voting and parochialism. *Journal of Conflict Resolution, 49*(6), 895–907.

Berman, O., Drezner, T., Drezner, Z., & Krass, D. (2009). Modeling competitive facility location problems: New approaches and results. In M. Oskoorouchi (Ed.), *TutORials in Operations Research* (pp. 156–181). San Diego: INFORMS.

Bhadury, J., Griffin, P. M., Griffin, S. O., & Narasimhan, L. S. (1998). Finding the majority-rule equilibrium under lexicographic comparison of candidates. *Social Choice & Welfare, 15*(4), 489–508.

Brams, S., & Fishburn, P. C. (1978). Approval voting. *American Political Science Review, 72*, 831–847.

Brams, S. J., & Fishburn, P. C. (2001). A nail-biting election. *Social Choice & Welfare, 18*(3), 409–414.

Brams, S. J., & Fishburn, P. C. (2007). *Approval voting*. New York, NY: Springer.

Buechel, B., (2014). Condorcet winners on median spaces. *Social Choice & Welfare, 42*(3), 735–750.

Buenrostro, L., Dhillon, A., & Vida, P. (2013). Scoring rule voting games and dominance solvability. *Social Choice & Welfare, 40*(2), 329–352.

Drezner, T. (1994a). Locating a single new facility among existing unequally attractive facilities. *Journal of Regional Science, 34*, 237–252.

Drezner, T. (1994b). Optimal continuous location of a retail facility, facility attractiveness, and market share: An interactive model. *Journal of Retailing, 70*, 49–64.

Drezner, Z. (1982). Competitive location strategies for two facilities. *Regional Science and Urban Economics, 12*, 485–493.

Drezner, Z. (1996). A note on accelerating the Weiszfeld procedure. *Location Science, 3*, 275–279.

Drezner, Z., Klamroth, K., Schöbel, A., & Wesolowsky, G. O. (2002). The Weber problem. In Z. Drezner & H. W. Hamacher (Eds.), *Facility location: Applications and theory* (pp. 1–36). Berlin: Springer.

Drezner, Z., & Menezes, M. (2014). The wisdom of voters: Evaluating the weber objective in the plane at the Condorcet solution. *Annals of Operations Research, 246*, 205–226.

Drezner, Z., & Suzuki, A. (2004). The big triangle small triangle method for the solution of non-convex facility location problems. *Operations Research, 52*, 128–135.

Eiselt, H. A., Marianov, V., & Drezner, T. (2015). In G. Laport, S. Nickel, & F. S. Gama (Eds.), *Location science*. Berlin: Springer International Publishing.
Eiselt, H. A., & Sandblom, C.-L. (2014). *Decision analysis, location models, and scheduling problems*. Berlin: Springer.
Francis, R. L., Lowe, T. J., Rayco, M. B., & Tamir, A. (2009). Aggregation error for location models: Survey and analysis. *Annals of Operations Research, 167*, 171–208.
Francis, R. L., McGinnis, L. F. Jr., & White, J. A. (1992). *Facility layout and location: An analytical approach* (2nd ed.). Englewood Cliffs, NJ: Prentice Hall.
Galton, F. (1907). Vox populi. *Nature, 75*, 450–451.
Gouret, F., Hollard, G., & Rossignol, S. (2011). An empirical analysis of valence in electoral competition. *Social Choice & Welfare, 37*(2), 309–340.
Hakimi, S. L. (1983). On locating new facilities in a competitive environment. *European Journal of Operational Research, 12*, 29–35.
Hakimi, S. L. (1986). p-Median theorems for competitive location. *Annals of Operations Research, 6*, 77–98.
Hansen, P., & Thisse, J.-F. (1981). Outcomes of voting and planning. *Journal of Public Economics, 16*, 1–15.
Hardin, G. (1968). The tragedy of the commons. *Science, 162*(3859), 1243–1248.
Henry, M., & Mourifié, I. (2013). Euclidean revealed preferences: Testing the spatial voting model. *Journal of Applied Econometrics, 28*(4), 650–666.
Hotelling, H. (1929). Stability in competition. *Economic Journal, 39*, 41–57.
Huff, D. L. (1964). Defining and estimating a trade area. *Journal of Marketing, 28*, 34–38.
Huff, D. L. (1966). A programmed solution for approximating an optimum retail location. *Land Economics, 42*, 293–303.
Klamler, C., & Pferschy, U. (2007). The traveling group problem. *Social Choice & Welfare, 29*(3), 429–452.
Knoblauch, V. (2001). Using elections to represent preferences. *Social Choice & Welfare, 18*(4), 823–831.
Labbé, M. (1985). Outcomes of voting and planning facility location problems in single facility location problems. *European Journal of Operational Research, 20*, 299–313.
Love, R. F., Morris, J. G., & Wesolowsky, G. O. (1988). *Facilities location: Models & methods*. New York, NY: North Holland.
Menezes, M. B. C., & Huang, R. (2015). Comparison of Condorcet and Weber solutions on a plane: Social choice versus centralization. *Computers & Operations Research, 62*, 350–355.
Menezes, M. B. C., Silveira, G. J. C., & Drezner, Z. (2016). Democratic elections and centralized decisions: Condorcet and approval voting compared with median and coverage locations. *European Journal of Operational Research, 253*, 195–203.
Merrill-III, S. (1985). A statistical model for Condorcet efficiency based on simulation under spatial model assumptions. *Public Choice, 47*(2), 389–403.
Mueller, D. C., Philpotts, G. C., & Vanek, J. (1972). The social gains from exchanging votes: A simulation approach. *Public Choice, 13*(1), 55–79.
Noltemeier, H., Spoerhase, J., & Wirth, H.-C. (2007). Multiple voting location and single voting location on trees. *European Journal of Operational Research, 181*, 654–677.
Plassmann, F., & Tideman, T. N. (2014). How frequently do different voting rules encounter voting paradoxes in three-candidate elections? *Social Choice & Welfare, 42*(1), 31–75.
Posamentier, A., & Lehmann, I. (2012). *The secrets of triangles: A mathematical journey*. Amherst, NY: Prometheus Books.
Pritchard, G., & Slinko, A. (2006). On the average minimum size of a manipulating coalition. *Social Choice & Welfare, 27*(2), 263–277.
Rapoport, A., & Felsenthal, D. S. (1990). Efficacy in small electorates under plurality and approval voting. *Public Choice, 64*(1), 57–71.
Reilly, W. J. (1931). *The law of retail gravitation*. New York, NY: Knickerbocker Press.
Rich, L. (2010). *Tapping the wisdom of the crowd*. New York: The New York Times.

Rodríguez, C. M. C., & Pérez, J. A. M. (2000). Comparison of α-Condorcet points with median and center locations. *Sociaded de Estadistica e Investigacion Operativa TOP, 8*, 215–234.
Rodríguez, C. M. C., & Pérez, J. A. M. (2008). Multiple voting location problems. *European Journal of Operational Research, 191*, 437–453.
Schulman, M. (2018). Why the electoral college – History central website. Available at http://www.historycentral.com/elections/Electoralcollgewhy.html (2018)
Sinopoli, F. D. (2000). Sophisticated voting and equilibrium refinements under plurality rule. *Social Choice & Welfare, 17*, 655–672.
Surowiecki, J. (2004). *The wisdom of crowds*. New York, NY: Doubleday.
Taylor, A. D., & Pacelli, A. (2008). *Mathematics and politics* (2nd ed.). New York, NY: Springer Science + Business Media, LLC.
Weiszfeld, E. (1936). Sur le point pour lequel la somme des distances de n points donnés est minimum. *Tohoku Mathematical Journal, 43*, 355–386.
Weiszfeld, E., & Plastria, F. (2009). On the point for which the sum of the distances to n given points is minimum. *Annals of Operations Research, 167*, 7–41.
Wendel, R. E., & Thorson, S. J. (1974). Some generalizations of social decisions under majority rules. *Econometrica, 42*, 893–912.
Wendell, R. E., & Hurter, A. P. (1973). Location theory, dominance and convexity. *Operations Research, 21*, 314–320.

Chapter 8
Neighbourhood Reduction in Global and Combinatorial Optimization: The Case of the *p*-Centre Problem

Said Salhi and Jack Brimberg

8.1 Introduction

The objective of the *p*-centre problem is to select *p* sites to locate new facilities in order to minimise the maximum distance or travel time between a set of demand points and the facilities closest to them. This problem, originally formulated on a graph by Hakimi (1964), is usually categorised as either the vertex *p*-centre problem or the absolute (or planar) *p*-centre problem. In the former, which is the discrete case, the optimal facilities are selected from a given set of potential sites (vertices) which can be either the demand points or other known sites. However, in the latter case the facilities can be located anywhere in the plane. An interesting chapter on continuous location analysis is given by Drezner (2011).

Neighbourhood reduction aims to eliminate moves or checks that cannot lead to an optimal solution for the case of exact methods, or are unlikely to affect the global best solution in the case of heuristics. It is challenging to design powerful neighbourhood reduction schemes that can cut computing time as much as possible without (or slightly) affecting the quality of the solution. Determining a good balance between the depth of the cut and the retention of solution quality at a reasonably high level is a challenging issue. Having good insight into the structure of the problem is useful as this helps in defining and designing the right neighbourhood

S. Salhi (✉)
Centre for Logistics and Heuristic Optimisation (CLHO), Kent Business School, University of Kent, Canterbury, UK
e-mail: s.salhi@kent.ac.uk

J. Brimberg
Department of Mathematics and Computer Science, The Royal Military College of Canada, Kingston, ON, Canada
e-mail: Jack.Brimberg@rmc.ca

H. A. Eiselt, V. Marianov (eds.), *Contributions to Location Analysis*, International Series in Operations Research & Management Science 281,
https://doi.org/10.1007/978-3-030-19111-5_8

reduction scheme. This mechanism, also known as reduction test, can be either dynamic which adapts as the search progresses, or deterministic which is usually defined from the outset. For more information on this research design issue, see Salhi (2017).

This chapter is organised into three parts:

1. A brief review of both vertex and planar p-centre problems with a focus on the contributions by Drezner;
2. A review of some neighbourhood reductions that are observed to be very useful for solving the vertex and planar p-centre problems; and
3. A discussion of some key research items on neighbourhood reduction that could be worth exploring.

8.2 The P-Centre Problem

We organise this section into two subsections. The first deals with the (discrete) vertex p-centre problem, while the second discusses the (continuous) planar p-centre problem.

8.2.1 The Vertex p-Centre Problem

The vertex p-centre problem, also known as the multi-facility minimax location problem, aims to optimally locate p facilities among a finite number of potential sites and to assign demand points to these open facilities in order to minimise the maximum distance between demand points and their nearest facility. There are two main formulations, as a binary linear program (BLP), and as a set covering problem (SCP).

8.2.1.1 The BLP Formulation

Let

(I, J): the set of demand points (or customers) ($i \in I = \{1, \ldots, n\}$) and set of potential facility sites ($j \in J = \{1, \ldots, m\}$),

$d(i,j)$: the distance between customer i and potential site j (Euclidian distance is used in our study);

p: the required number of facilities;

$$Y_{ij} = \begin{cases} 1 \text{ if customer } i \text{ is assigned to a facility at location } j \\ 0 \text{ otherwise} \end{cases}$$

$$X_j = \begin{cases} 1 \text{ if a facility is opened at location } j \\ 0 \text{ otherwise} \end{cases}$$

Z: the maximum distance between the customers and their closest facilities.

The problem (BLP) is then formulated as follows:

$$\text{Minimize} \quad Z \tag{8.1}$$

Subject to

$$\sum_{j \in J} Y_{ij} = 1 \quad \forall i \in I \tag{8.2}$$

$$\sum_{j \in J} X_j = p \tag{8.3}$$

$$Y_{ij} - X_j \leq 0 \ \forall i \in I, j \in J \tag{8.4}$$

$$Z \geq \sum_{j \in J} d\,(i, j)\, Y_{ij} \ \forall i \in I \tag{8.5}$$

$$X_j \in \{0, 1\} \ \ \forall j \in J \tag{8.6}$$

$$Y_{ij} \in \{0, 1\}\, \forall i \in I, j \in J \tag{8.7}$$

The objective (8.1) refers to the minimization of the maximum distance between a customer and its nearest facility. Constraints (8.2) guarantee that each customer is assigned to exactly one facility; constraint (8.3) limits the number of open facilities to be p; while constraints (8.4) ensure that a customer can only be allocated to an open facility. Constraints (8.5) define the maximum distance between a customer and its closest facility. Constraints (8.6) and (8.7) refer to the binary type of the decision variables. Note that the binary type constraints on the Y_{ij} in (8.7) can be replaced by non-negativity constraints without affecting the optimal solution, since the minimization objective will force customers to be assigned to their nearest facilities.

The p-centre problem is known to be NP-hard (Kariv and Hakimi 1979). Thus it follows that only small to medium size instances of this problem can be solved optimally using commercial optimization software such as CPLEX, LINDO, GUROBI or Xpress-MP, and it becomes more difficult to tackle for relatively large instances. One idea is to aggregate customers leading to a smaller problem which is more manageable. However, it is worth noting that such an aggregation-based approach, if not considered carefully, could lead to poorer quality solutions due to the loss of information. Another approach is to address the problem in its entirety by adopting powerful metaheuristics or mat-heuristics. For instance, Irawan et al. (2016) develop a powerful hybridisation of VNS and ILP formulations by embedding intelligent neighbourhood reduction schemes.

8.2.1.2 A Set Covering-Based Model

The minimax problem can also be solved optimally using a Set Covering Problem (SCP)-based approach. Given a covering distance (or response time) D, SCP aims to find the minimum number of facilities and their locations so that each customer is served by a facility within D distance from it.

Let

$$a_{ij} = \begin{cases} 1 \text{ if customer } i \in I \text{ can be covered by a facility sited at } j \in J \text{ (i.e., } d(i, j) \leq D) \\ 0 \text{ otherwise} \end{cases}$$

The SCP can be formulated as follows:

Minimise

$$\sum_{j \in J} X_j \tag{8.8}$$

Subject to

$$\sum_{j \in J} a_{ij} X_j \geq 1 \quad \forall i \in I \tag{8.9}$$

$$X_j \in \{0, 1\} \quad \forall j \in J \tag{8.10}$$

The objective (8.8) is to minimise the number of opened facilities. Constraints (8.9) guarantee that each customer is covered by at least one facility located within the threshold D and constraints (8.10) refer to the binary variables.

The minimax problem is optimally tackled by recursively solving a sequence of SCPs for given values of D using a binary search. For instance, Daskin (2000) adopted this approach, initially presented by Minieka (1970), on a general graph with all edge distances restricted to integers. Efficient exact algorithms for solving the vertex p-centre problem include, for example, Ilhan and Pinar (2001), Elloumi et al. (2004), Al-Khedhairi and Salhi (2005), Salhi and Al-Khedhairi (2010), and Irawan et al. (2016). The latter ones incorporate neighbourhood reductions which are discussed in Sect. 8.3.2.

8.2.2 *The Planar p-Centre Problem*

Continuous location problems are about generating sites for one or more facilities in the plane. Though the obtained solutions may not be feasible as some facilities may end up in the middle of a city or a lake, they can still be used as greenfield solutions (ideal solutions). Given that the continuous problem can be a good approximation of its discrete counterpart especially when the network has a large number of potential sites, getting the ideal solution could provide valuable information for decision makers when selecting the final sites. As the data gathering task for a large network can be very expensive to conduct, the continuous model may also be used to reduce the number of potential sites to a few promising ones, thus making the problem more manageable.

From a theoretical view point, the continuous p-centre problem is also interesting as it has a geometrical interpretation. For example, the single unweighted facility location problem (i.e., $p = 1$) reduces to finding the smallest circle that encloses all the customers, with the centre of the circle being the location of the new facility. In a similar way, the continuous p-centre problem with $p > 1$ may be interpreted as finding the centres of p circles that encompass all the customers where the radius of the largest circle is made as small as possible.

The (weighted) p-centre problem can also be described as a MinMaxMin type problem with formulation given by Drezner (1984a):

$$\text{Min } Z(X) = \underset{i=1,\ldots,n}{\text{Max}} \left[w_i \underset{j=1,\ldots,p}{\text{Min}} d\left(P_i, X_j\right) \right]$$

where the additional notation is defined as follows:

$P_i = (a_i, b_i)$: the given location of demand point i $(i = 1, \ldots, n)$
$w_i > 0$: the weight of demand point i $(i = 1, \ldots, n)$
$X_j = (x_j, y_j)$: the unknown location of new facility j with $X_j \in \Re^2$; $j = 1, \ldots, p$
$X = (X_1, \ldots, X_p)$: the vector of decision variables containing these p facility locations
$d(P_i, X_j)$: the Euclidean distance between P_i and $X_j (i = 1, \ldots, n; j = 1, \ldots, p)$

Note that the *unweighted* model is normally considered as in the discrete p-centre problem given above with equal weights ($w_i = 1, i = 1, \ldots, n$).

The single facility minimax location problem (1-centre) in continuous space has a long history. The English mathematician James Joseph Sylvester (1814–1897) first posed the problem in 1857, and then a few years later, in 1860, put forward an algorithm to solve it. The problem was dormant for over a century until Elzinga and Hearn (1972) presented an optimal geometrical-based algorithm that runs in polynomial time. Since then, other authors attempted some speed-up procedures, such as Hearn and Vijay (1982), Xu et al. (2003), and Elshaikh et al. (2015), and references therein. Some of these enhancements use simple but effective reduction schemes, which are discussed in Sect. 8.4.1. For an informative review including the history of this problem, see Drezner (2011) and references therein.

There is, however, a relative shortage of studies dealing with the problem for larger values of p (see Plastria (2002) and Callaghan et al. (2018)). Chen (1983) is among the first to tackle the p-centre problem in the plane. The problem is shown to be NP-hard in Megiddo and Supowit (1984). For a fixed value of p, the problem can be solved in polynomial time $O(n^2p)$ as shown by Drezner (1984a), though it requires an excessive amount of computational effort for larger values of n and p. Due to the non-convexity of the objective which is a function of the location variables, this problem also falls in the realm of 'global optimization'.

For the case of the 2-centre problem with Euclidean distances in the plane (i.e., $p = 2$), there is an interesting optimal algorithm by Drezner (1984b). The idea is that the entire customer set can be split into two separate sub-problems by a straight line, where each can be optimally solved as a 1-centre problem. However, as there $\frac{n(n-1)}{2}$ possibilities, the problem becomes difficult though still polynomial. The method can be extended to larger p, where more than one line would be needed, but this problem becomes much more difficult. A scheme on how to proceed from one set of p lines (p clusters) to another is an exciting exercise that could be worth exploring.

Constructive heuristics were the first to emerge for larger values of p. These use the iterative locate-allocate procedure initially proposed by Cooper (1964) for the Weber problem for local improvement, and are based on the commonly used *add*, *drop* and *swap* moves. For instance, Drezner (1984a), and Eiselt and Charlesworth (1986) were among the first to develop such methods.

Drezner (1984a) also devised a nice optimal algorithm using the idea of Z-maximal circles. For a given Z, all maximal circles are defined and either a corresponding set covering problem or a feasibility problem is solved. Starting from a lower bound for Z, successive problems are solved until a feasible solution is found (i.e., the solution has p circles with each customer being encompassed by at least one circle). This optimal algorithm was originally very slow but has since developed into a fast and powerful approach that can solve large instances to optimality. This is achieved by incorporating suitable neighbourhood reductions into the search which are discussed in Sect. 8.4.3.1.

Several years later, Chen and Chen (2009) developed a relaxation method based on Chen and Handler (1987) to optimally solve the problem. The idea is to start by solving a reduced problem containing a subset of demand points from the original problem and then gradually adding some points to the current subset until the optimal solution is feasible for the original problem. This interesting relaxation-based approach is also revisited and its efficiency is much enhanced in Sect. 8.4.3.2.

8.3 Neighbourhood Reduction for the Vertex *p*-Centre Problem

We first present basic reduction schemes which can be embedded efficiently into the brute force approach, also known as the complete enumeration technique. These are followed by those more advanced neighbourhood reductions that are adopted primarily for optimal methods, metaheuristics and mat-heuristics for the case of the *p*-centre problem.

8.3.1 Brute Force Approach

The idea is to evaluate all combinations of p possible facility sites out of the n potential sites. For each combination allocate each customer to its nearest facility, leading to p clusters, and choose the one that yields the maximum distance from the allocated customers to the centre (their nearest facility). The optimal solution is one with the minimum of these maxima. This complete enumeration technique (CET), though naive, can be used to guarantee the optimal solution for small values of p (< 5), even when n is around 100, without the use of commercial optimisers or even the use of any heuristic. This simple and rudimentary approach if applied blindly will evaluate all the $C_p^n = \frac{n!}{p!(n-p)!}$ combinations, and will fail rapidly when $p \geq 5$ even for $n = 100$.

However, with simple reduction rules, we can improve its efficiency drastically still without recourse to advanced methods. The following four rules, which are given in Al-Harbi (2010), are briefly discussed here.

(a) CET is coded in p nested loops (in an ordered fashion) leading to any two successive configurations being different by one facility only. In the allocation process, for a given customer i we have two options whether this customer has lost its initial facility and hence needs to be checked against all p facilities including the new one; otherwise one comparison between its original assignment and the new one is evaluated. For example, the instance *(Pmed1)* from the ORLIB with $n = 100$ and $p = 5$ required 506 s, whereas with this simple reduction it needed only 348 s. The CET and simple reduction were coded in C++ and performed on a PC i7 with 1.5 GHz processor and 512 MB of RAM. However, for the instance *Pmed6* ($n = 200$; $p = 5$), both versions were unable to obtain the optimal solution after 5 h of CPU time. Here, the blind approach exploited 42% of the total number of combinations, whereas the enhanced one used 70% instead, leading to better chances of obtaining an improved solution, though optimality cannot obviously be guaranteed.

(b) Besides (a), we can also record the second closest facility for each customer. Though this adds extra computational storage, it reduces the overall computing time. Using the same example with $n = 100$ and $p = 5$, this simple recording task enables the optimal solution to be found within 321 s. For $n = 200$ and $p = 5$, though the optimal solution is still not guaranteed, 77% of combinations were now evaluated.

(c) Note that both operators ((a) and (b)) are not only applicable to this problem but are also commonly used in many other combinatorial problems where an assignment is required. We take into account additional insights unique to the p-centre problem. Given the objective is to minimise the maximum coverage, it is clear that once we have one feasible solution with a value of Z, this can be used to terminate the evaluation of a given configuration if one customer happens to have its distance to its nearest facility larger than Z. In other words, there is no need to continue checking the other customers for this particular configuration. The upper bound Z can be updated as the search proceeds. This reduction scheme systematically leads to rejecting several inferior configurations early on leading to a massive reduction in the overall computing time. This basic rejection scheme obtains the optimal solutions for both instances (*Pmed1* and *Pmed6*) within the maximum 5 h allowed (18,000 s), requiring only 54 s and 11,546 s, respectively.

(d) This is an extension of (c) where for each customer $i = 1, \ldots, n$, a set of facilities $F_i = \{j = 1, \ldots, n \mid (d(i,j) \leq Z\}$ is constructed (usually updated as the search goes on). If a given configuration does not contain at least one facility in F_i, there is no need to continue the allocation of other customers as this configuration is inferior. This dynamic reduction rule, which is based on the current Z, renders the brute force even faster by obtaining the optimal solutions for *Pmed1* and *Pmed6* in 47 s and 2634 s only, respectively. This rule dominates the one that also states that for a given customer, a facility configuration that includes one of its furthest $(p - 1)$ facilities is systematically inferior and hence needs to be discarded.

The above rules demonstrate that the information in a given problem may be used to eliminate several redundant computations if appropriate neighbourhood reductions are designed. However, even with such elimination rules, the brute force approach is still limited to smaller values of n and p. Having said that, the effective use of neighbourhood reduction is still able to reduce by as much as some 90% the time attributed to unnecessary computations. The impact is even more significant when these reduction schemes are embedded within powerful metaheuristics or optimal algorithms as will be shown in the rest of this chapter.

8.3.2 Set Covering-Based Approach

The approach using SCP, as given by Daskin (2000), is shown in Fig. 8.1. In this section, we revisit some of its steps to enhance its efficiency.

Step 1- Set $L = 0$ and $U = \underset{i,j}{Max}\, d(P_i, P_j)$.

Step 2- Compute the coverage distance $D = \dfrac{L+U}{2}$

Step 3- Solve the Set Covering Problem (SCP) using D as the covering distance and let v be the optimal number of facilities obtained.

Step 4- If $v \le p$ (i.e., the solution is feasible for the p-centre problem), set $U = D$; else (ie., $v > p$, the solution is infeasible) set $L=D$.

Step 5- If $U - L \le 1$, record U as the optimal solution and stop, otherwise go to Step 2.

Fig. 8.1 The basic SCP algorithm

8.3.2.1 Revisiting Steps 1 and 5

For instance, Al-Khedhairi and Salhi (2005) proposed some basic changes in Step 1 when initialising the bounds L and U by re-defining

$$L = \underset{j}{\text{Max}}\underset{i}{\text{Min}}\, d\left(P_i, P_j\right) \text{ and } U = \underset{i}{\text{Min}}\underset{j}{\text{Max}}\, d\left(P_i, P_j\right).$$

Also, to guarantee that the elements of the distance matrix in Step 2 are used only, D is redefined slightly by setting $D = G\left(\frac{L+U}{2}\right)$ where $G(x)$ represents the nearest value to x in the distance matrix.

In addition, to terminate the search as early as possible in Step 5 and avoid redundant checks, the set $S = \{d(P_i, P_j) : L < d(P_i, P_j) < U\}$ is introduced. If $S = \varnothing$ (i.e., there are no distance values between L and U), the search terminates even if $U - L > 1$, with the optimal solution being U. In addition, if $|S| = 1$, there is one element left to assess only, say D and go to step 3. The optimal solution remains either at U or at D if the new SCP solution happens to be feasible. These schemes are tested on all instances of the ORLib ($n = 100$ to 900 and $p = 5$ to 90) where a 15% average reduction in the number of SCP calls is obtained. For the TSPLib data set ($n = 1060$ and $p = 10, 20, \ldots, 150$), a more significant average reduction of over 28% is recorded.

8.3.2.2 Further Tightening of *U* and *L* in Step 1

In step 1, a simple tightening of U can be found easily just by running a multi-start approach and choosing $U = \text{Min}\left(\underset{i}{\text{Min}}\underset{j}{\text{Max}}\, d\left(P_i, P_j\right), Z_H\right)$ with Z_H being the best solution of all the runs. For instance, when $Max(n, 500)$ multi starts are adopted, the above results for the ORLib and TSPLib improved even more leading to an average reduction in the number of iterations from the original implementation of 23% and 41%, respectively.

A further tightening of U and L can be obtained by using the solution from a powerful metaheuristic Z_H as U leading to a potential lower bound $L = \alpha U$ with $\alpha = 0.7$ or 0.8. The tighter the Z_H is, the closer the α is to 1. The power of this setting is that even if αU fails to be a 'true' lower bound, its value will automatically become the upper bound instead, and the lower bound is recomputed again as $L = \alpha U$ with α kept the same or slightly reduced until a range (L, U) is derived. It is worth noting that there are no redundant calculations when assessing the SCP for L. Salhi and Al-Khedhairi (2010) proposed this implementation with interesting results using a multi-level heuristic originally proposed by Salhi and Sari (1997). Recently, Irawan et al. (2016) adopted the above methodology with two distinct changes. The upper bound in step 1 is obtained by VNS instead (see Hansen et al. 2010), and also an ordered list of the distance matrix is constructed to easily identify the elements in the set S. As the upper bound produced by VNS in Step 1 may be close to the optimal solution, α can be set in the range [0.8–0.9] leading to an even tighter $L = \alpha U$. Note that the value of L must also exist in the distance matrix and hence is set to the nearest such value to αU.

8.3.2.3 Other Tightening of *L*

An interesting two-phase approach which shares some similarities with the SCP is based on solving the feasibility of the following covering problem (CP) instead (Ilhan and Pinar 2001). Note that the notation remains as given in Sect. 8.2.1.

$$\left\{ \sum_{j \in J} a_{ij} X_j \geq 1 \quad \forall i \in I; \sum_{j \in J} X_j \leq p; X_j \in \{0, 1\} \quad \forall j \in J \right\}$$

The idea is that if a relaxed CP (i.e., $0 \leq X_j \leq 1 \quad \forall j \in J$) which is much quicker to solve does not provide a feasible solution for a given D, there is no need to solve the integer problem. This phase one is similar to SCP in Fig. 8.1, except that Step 3 is based on solving the relaxed CP. Once a feasible solution is obtained in phase one, phase two is activated where CP is solved with the corresponding D. If the integer problem is not feasible, then a tight lower bound $L = D$ can be used. Note that phase two in Ilhan and Pinar (2001) does not follow the SCP algorithm given in Fig. 8.1 but attempts to solve the integer CP by gradually increasing D to the next minimum in the distance matrix until an integer solution of CP is found.

8.3.3 The Local Search in VNS

Irawan et al. (2016) introduce some elimination rules to avoid computing unnecessary moves in both the shaking process and the local search of the VNS. The local search, which is a vertex substitution heuristic, implements a swap move by closing

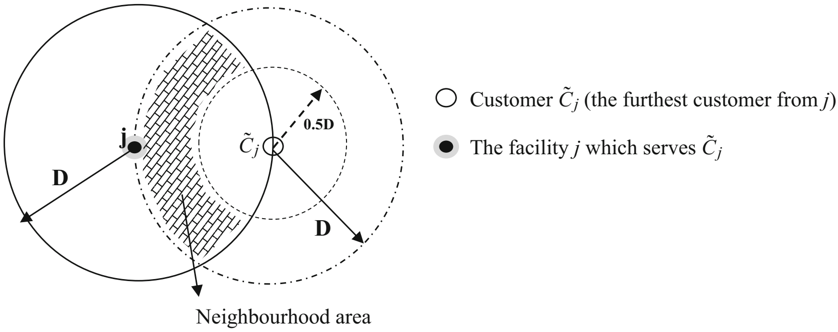


Fig. 8.2 The restricted but guided neighbourhood within VNS

an open facility and replacing it with a closed one. However, instead of removing a facility randomly from the current facility configuration, the facility (say facility j) whose radius (the distance between a facility and its furthest customer) is the largest, say D, is chosen.

Let

$E_j = \{i = 1, \ldots, n | \; d(i,j) \leq D\}$: be the set of customers served by facility j and $\tilde{C}_j = Arg\underset{i \in E_j}{\text{Max}} \left(d\left(i, j\right)\right)$: be the customer whose distance is D from facility j, $V_j = E_j \cap \left\{i = 1, \ldots, n | d\left(\tilde{C}_j, i\right) \leq D \wedge d\left(\tilde{C}_j, i\right) \geq \frac{D}{2}\right\}$: be the subset of potential sites from which to randomly choose an open facility to replace facility j.

In this case, the location of the new open facility is restricted so as not to be too close to customer $\tilde{C}_j$. This concept of using forbidden regions is shown to be effective when solving the multi-source Weber problem (Gamal and Salhi 2001) and its capacitated version (Luis et al. 2009). Here, the threshold distance is set to $\frac{D}{2}$ though this can vary; see Fig. 8.2 which is adapted from Irawan et al. (2016). This reduction scheme considers a fraction of the customer sites only, which can in turn reduce the search space considerably and lead to a substantial cut on the computing time. This can be significant given the local search in VNS is applied a large number of times.

This approach, when integrated as part of a matheuristic, shows to be effective at tackling very large instances up to $n = 71,000$ and $p = 25, 50, 75$ and 100 for both the conditional and unconditional vertex p-centre problems, see Irawan et al. (2016).

8.4 Reduction Schemes for the Planar p-Centre Problem

The planar p-centre problem has received relatively less attention compared to its counterpart the vertex p-centre problem especially when it comes to optimal methods. In this section, we first explore elimination techniques for the 1-centre problem though this can be solved optimally in polynomial time. We then present speed-up procedures for the commonly used metaheuristics followed by those devoted to the optimal and relaxation-based methods.

8.4.1 The 1-Centre Problem

The problem of determining the optimal location in continuous space for a given cluster of customers turns into finding the centre of the smallest circle that encompasses all the customers. As a circle can be identified by one, two or three critical points only, Elzinga and Hearn (1972) used this concept to develop an $O(n^2)$ geometrical-based optimal algorithm to solve the problem. The idea is to start with two demand points chosen randomly and find the corresponding optimal centre and the radius of the circle. If all demand points are encompassed by the circle, the search terminates; otherwise another point is added and a new circle with the optimal centre for the three points is constructed. If the circle does not cover all demand points, a new uncovered point is added again while one of the three points that becomes redundant is removed. This process continues until a circle that covers all points is found. There are a few studies that looked at this issue. Recently, Elshaikh et al. (2015) introduced two simple but effective reduction rules into the Elzinga and Hearn (EH) algorithm.

(a) In the starting step, instead of choosing any two points, the four points that make up the smallest rectangle that covers all the points are first identified. For each pair, the corresponding circle is found and the largest one is chosen with its critical points as the starting solution.
(b) When selecting the new point to add to the already existing circle with centre X and radius Z, instead of choosing the new point randomly, the one with the largest (weighted) distance to the current centre is selected instead, i.e., $P^* = Arg \underset{i=1,\ldots,n}{\text{Max}} (\mathrm{w}_i \mathrm{d}(\mathrm{P}_i, \mathrm{X}) > \mathrm{Z})$.

When tested on a set of randomly generated instances in a square $(0, 100) \times (0, 100)$ with $n = 100$ to 1000, the enhanced EH algorithm required about 30% and 20% of the computing time of the original version for the unweighted and the weighted cases, respectively. More details on the experiments including other less promising rules can be found in Elshaikh (2014). One may assume that there is no need to speed up such a quick polynomial procedure. This will be obviously true if the aim is to solve the problem once or a small number of times only. However, in the p-centre problem, this task embedded within a metaheuristic or an optimal method will be called upon several times, and therefore, in our view the enhancement is quite worthwhile. For example, to demonstrate the benefit of these two reduction tests, Elshaikh et al. (2015) perform an extensive experiment using a simple multi-start with 100 iterations on the $n = 1002$ *TSPLib* instance with $p = 5$ to 25 in increments of 5. It is found that over 32% less updating within EH is required leading to a reduction of over 25% in computational time.

8.4.2 Reduction within Heuristic-Based Approaches

We discuss the recent neighbourhood strategies that have proved to be promising when embedded within the powerful metaheuristics used for the p-centre problem.

8.4.2.1 VNS-Based Heuristics

In the shaking process within VNS, a certain number ($K_{\max}$) of neighbourhood structures is defined ($N_k(X); k = 1, \ldots, K_{\max}$). For the p-centre problem, these can be either (a) customer based or (b) facility based.

In (a), $N_k(X)$ can be defined by reallocating k demand points from their original facilities to other ones either randomly or following a certain strategy. Due to the characteristics of the planar p-centre problem, the number of critical points that define the largest circle obtains $K_{\max} \leq 3$. The reallocation of one of these points will in most cases reduce the radius of the largest circle (except in the case of ties). Note that other circles may increase their radii after this allocate-locate procedure, but the solution is accepted as long as Z is reduced.

In (b), $N_k(X)$ is defined by relocating k facilities ($k = 1, \ldots, K_{\max} = p$) from the current solution. Instead of randomly removing k facilities, the following removal scheme that guides the search more effectively is performed. As the problem is linked to the largest circle and its surrounding circles, it is therefore important to take into account these characteristics when designing the neighbourhood reductions. Two aspects are worth considering here. The first one is connected to the largest circle and the second is linked to those facilities deemed non-promising. For simplicity, consider the largest circle as C_1 and let

CT_k: the centre of circle C_k; $k = 1, \ldots, p$
$\delta_s = d(CT_1, CT_s)$; $s = 1, \ldots, p$: the distance between the centre of C_1 and the centre of C_s
$\gamma(k) = \underset{\substack{s=1,\ldots p \\ s \neq \gamma(l),\, l=1,\ldots k-1}}{\text{ArgMin}} \delta_s$ and
$C_{\gamma(k)}$: the kth nearest circle to C_1 with $\gamma(1) = 1$ referring to C_1.

The process starts removing the facility at CT_1 (i.e., circle C_1) and assigning it somewhere else as will be discussed shortly. If after a local search the solution is not improved, both facilities located at CT_1 and $CT_{\gamma(2)}$ (i.e., both circles C_1 and $C_{\gamma(2)}$) are then removed, and the process continues until all facilities (i. e., all circles $C_{\gamma(1)}, C_{\gamma(2)}, \ldots C_{\gamma(p)}$) are removed if necessary. If the solution is improved, the information is updated (i.e., CT_k, $\gamma(k)$, $k = 1, \ldots, p$) and as in VNS, we revert back to the removal of the facility at CT_1 again.

The second aspect is based on identifying those non-promising facilities for removal. In our case a facility is considered as non-promising if it encompasses its critical points only (i.e., there are no interior points within the circle). This identification will lead to a saving of q ($q << p$) facilities which can then be added around the critical points of the largest circle one at a time. For example, Elshaikh et al. (2015) conducted an experiment for *TSPLib* with $n = 439$ and $p = 10, 20, \ldots, 100$, where a 9% average improvement was obtained. In particular, for the case of $p = 100$, a 34% improvement was observed with seven facilities being identified as non-promising. Most of the improvements were found with $p \geq 50$.

In both cases, the re-assignment of facilities is also performed using the characteristics of the p-centre problem. Instead of inserting a removed facility either randomly anywhere in the plane or at fixed customer sites, the following attractor scheme for insertion is adopted. Using the largest circle again with radius Z, for each of its critical points (generally three or fewer), it can be shown that there is no other facility within distance Z of any of these critical points. This observation is important as it shows that at least one new facility has to be located somewhere inside this region. This statement can be shown empirically, but it is also mathematically proven in Mladenović et al. (2003). It is therefore important to consider these areas where the new locations will be sited. Within VNS, each time a facility is removed and relocated, the local search is adopted. If the solution is improved, the following updating takes place by defining the largest circle, the neighbourhood for removal and the new area where to locate. For example, when the q non-promising facilities are saved, one facility at a time is located randomly in those areas near the critical points of the largest circle, the reallocation process is then used based on avoiding unnecessary repetition of computations. That is, only affected circles have their centres and radii recalculated using their earlier respective centres and radii as the starting solution. The allocation of customers is also performed effectively by considering whether or not a customer lost his/her original facility. The process is repeated until there are no non-promising facilities remaining or the solution cannot improve anymore. This enhancement, when tested for the $n = 439$ *TSPLib* data set, shows a significant improvement over its original implementation.

The effect of this neighbourhood reduction has also helped to identify adaptively the best value of $K_{\max}$ as well as the best neighbourhood structures $N_k(X)$ that are worth examining. A VNS-based heuristic with all the above ingredients was able to obtain for the first time optimal solutions for larger TSPLib instances, see Elshaikh et al. (2015) for more details.

8.4.2.2 Perturbation-Based Metaheuristics

The idea is to perturb the current feasible solution by allowing it to have up to q facilities over or below p. This up and down shifting, which is repeated several times, has the tendency to retain those very promising facilities in the defined set. Salhi (1997) originally put forward this approach for the p median problem which is now successively adopted and extended for the p-centre problem (Elshaikh et al. 2016). The operators 'add', 'drop' and 'swap' are adopted here. The first operator applies when the solution has $p - q$ or p facilities to reach p or $p + q$ facilities, whereas

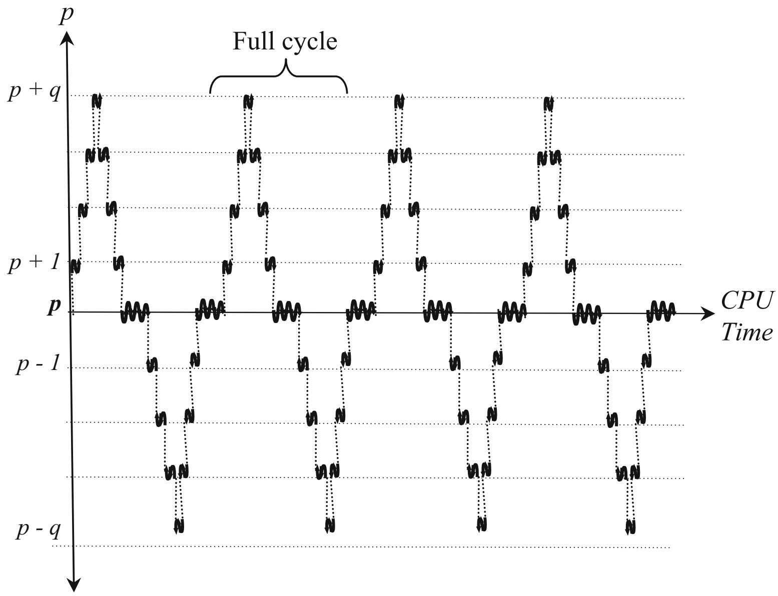


Fig. 8.3 A perturbation-based metaheuristic

the second operator is used when the solution has p or $p + q$ to reach $p - q$ or p facilities. The last operator is activated when the solution has exactly p facilities. The 'add' and 'drop' moves are implemented based on the following neighbourhood reductions.

We define the kth covering circle $CC_k = \{P = (x, y) \in \Re^2 | \, d(P, CT_1) \leq \delta_k\}$, $k = 1, \ldots, q$ with k facilities being either removed (the 'drop' move) from CC_k or added (the 'add' move) to CC_k. Both the removal and the addition mechanisms are performed using either $k = q$ in one go or gradually adding or removing one facility at a time until it reaches either $p + q$, p or $p - q$, see Fig. 8.3 for an illustration. Note that the value of q can be made dynamic using some form of learning. Besides, this value does not have to remain the same when the search goes over or under the value of p (i. e., $p - q_1; p + q_2$).

In the swap move, a facility in CC_1 is removed and relocated randomly in the continuous space also in CC_1 where a local search is then activated. Note that the Elzinga-Hearn algorithm or its equivalent (see Sect. 8.4.1) is applied at each solution to obtain the optimal centre for each cluster irrespective of whether or not the solution is feasible in terms of the number of facilities. These guided schemes reduce the computing time considerably enabling large *TSPLib* instances with over 1300 nodes and $p = 10, \ldots, 100$ to be solved efficiently; see Elshaikh et al. (2016) for these encouraging results.

8.4.2.3 Guiding the Search through Forbidden Regions

In the local search, extra care is needed when locating the removed facilities in the continuous space. Here, a new location chosen randomly on the plane may end up by chance being too close to one of the already selected facilities. Given the search is on the plane, one way forward is to construct a small area around each of the existing locations and make them tabu or forbidden. The idea of using forbidden regions is also adopted during the construction of the initial solution where the idea is to avoid having facilities that are too close to each other. This is explored successfully by Gamal and Salhi (2001) for the multi-source Weber problem (MSWP). In brief, during the construction phase of the initial solution or during the local search, any new continuous location, which lies in these forbidden regions, will be excluded from being selected. This useful information guides the search by avoiding these specified non-promising areas, thus reducing unnecessary computational efforts that would have been wasted otherwise.

In our experiments, a forbidden region is defined as the area enclosed by a circle with its centre at an already chosen location. The radius of the kth forbidden area $\left(\tilde{R}_k\right)$ is defined as $\tilde{R}_k = \alpha R_k$ with R_k being the radius of its original circle and parameter α set close to zero, say $\alpha = 0.05$. This setting could also be made adaptive by increasing α (say $\alpha \to 2\alpha$) or decreasing it (say $\alpha \to \alpha/2$) depending on whether the number of rejections is low or high, respectively, as demonstrated by Luis et al. (2009) for the capacitated MSWP. Elshaikh (2014) and Elshaikh et al. (2018) adapted the reformulation local search (RLS) which was originally proposed by Brimberg et al. (2014, 2017) for the MSWP by incorporating forbidden regions and other attributes. For example, when tested on the $n = 439$ *TSPLib* instance with $p = 10, 20, \ldots, 100$, the best average deviation is reduced from 3.114% to 2.647%, besides guaranteeing several optimal solutions, especially when $p \leq 40$. In brief, RLS is a new local search that aims to shift between discrete and continuous space while augmenting the discrete problem with the newly found continuous points (see Brimberg et al. (2014) for more details).

8.4.3 Optimal and Relaxation-Based Algorithms

We concentrate on two types of algorithms, namely the optimal method of Drezner (1984a) and the reverse relaxation technique of Chen and Chen (2009). Neighbourhood reductions are designed for each of these algorithms making them more effective and suitable for solving larger instances either optimally or by providing tight lower bounds which could also be used for benchmarking purposes if necessary.

8.4.3.1 Drezner's Optimal Algorithm

Drezner (1984a) designed an optimal algorithm based on the idea of Z-maximal circles. A circle is defined as maximal based on a given upper bound, Z.

Let the number of potential circles with n demand points be defined by $N_c(n) = |\prod_1(n)| + |\prod_2(n)| + |\prod_3(n)|$ with

$\prod_1(n)$: set of null circles$\left(|\prod_1(n)| = C_1^n = n\right)$,

$\prod_2(n)$: set of circles made up of two critical points $\left(|\prod_2(n)|=C_2^n=n\,(n-1)\,/2\right)$, and

$\prod_3(n)$: set of circles made up of three critical points making an acute triangle

$$\left(|\prod_3(n)| \leq C_3^n = n\,(n-1)\,(n-2)\,/6\right)$$

Also let

E_j be the subset of customers encompassed by circle C_j and $r(E_j)$ its radius.

C_j is said to be Z-maximal (or maximal for short) if its radius $r(E_j) < Z$ and for any demand point $i \notin E_j$; $r(E_j \cup \{i\}) \geq Z$.

First, for a given value of Z the set of maximal circles is defined by

$$\Phi_Z = \left\{C_j;\ j = 1, \ldots, N_c(n) | r\left(E_j\right) < Z \wedge r\left(E_j \cup \{i\}\right) \geq Z \forall i \notin E_j\right\}.$$

Drezner proposed two approaches; one uses the set covering problem, while the other is based on a feasibility problem. In the former, the problem is similar to the SCP given in Sect. 8.2.1.2(ii) except that for a given Z, the set of potential circles becomes Φ_Z, The feasibility problem ($P_F(Z)$) on the other hand is defined as follows:

Find

$$X_j \in \{0, 1\}, C_j \in \Phi_Z, \text{such that} \sum_{C_j \in \Phi_Z} a_{ij} X_j \geq 1 \forall i = 1, \ldots, n \text{ and} \sum_{C_j \in \Phi_Z} X_j = p$$

With

$$X_j = \begin{cases} 1 \text{ if maximal circle } \mathrm{C}_j \text{ is chosen} \\ 0 \text{ otherwise} \end{cases}$$

and

$$a_{ij} = \begin{cases} 1 \text{ if customer } i \text{ is encompassed by maximal circle } \mathrm{C}_j \ \left(\text{i.e., } i \in E_j\right) \\ 0 \text{ otherwise} \end{cases}$$

For both formulations when the solution is found, the new value of Z is set to $Z = Max\,(r(E_j)|\,X_j = 1)$ which then defines the new Φ_Z. The process continues until there is no feasible solution leading to the optimal solution having the last value of

Z. It was found empirically that the use of ($P_F(Z)$) is much more effective than using the set covering problem. For instance, very recently Callaghan et al. (2017) showed that for the $n = 439$ *TSPLib* instance with $p = 90$, the optimal solution was found in just below 3 h (and using 393 calls to the model) using ($P_F(Z)$), whereas the SCP-based method needed about 38 h and 4580 calls. This observation was noted even more emphatically for the $n = 575$ *TSPLib* instance where the optimal solution was obtained after 30 h using $P_F(Z)$, whereas the former stopped after 2 days of running with one feasible solution only and a 20% gap from the optimal solution. Callaghan et al. (2017) were able to speed up considerably the optimal method of Drezner (1984a), namely the one using $P_F(Z)$. This was achieved by incorporating efficient neighbourhood reduction mechanisms, thereby enabling several larger instances to be solved for the first time to optimality.

To respond to this challenge, a look-alike p-centre formulation $P_{op}(Z)$, which also considers ($P_F(Z)$), is first proposed.

$$\text{Minimize } R$$
$$s.t.\ \{ \sum_{C_j \in \Phi_Z} a_{ij} X_j \geq 1 \forall i = 1, \ldots, n;\ \sum_{C_j \in \Phi_Z} X_j = p;\ X_j \in \{0, 1\};\ C_j \in \Phi_Z \} \equiv (P_F(Z))$$
$$\text{and } X_j r\left(C_j\right) \leq R \text{ for all } j \in \Phi_Z$$

When testing the $n = 439$ *TSPLib* instance for $p = 90$ using $P_{op}(Z)$, the optimal solution was obtained about seven times faster than with $P_F(Z)$. Though it is relatively harder to solve $P_{op}(Z)$ than $P_F(Z)$, the former produces tighter Z values leading to less calls, each requiring the long computational burden in defining Φ_Z. These results demonstrate that though a reduction in computing time is achieved, there are two issues that could help to speed up the search. These include

1. An efficient identification of Φ_Z from one iteration (or call) to the next and
2. A scheme to find a good compromise solution as a feasible solution may not reflect the quality of the solution while an optimal solution may take too long to find.

The following neighbourhood reductions aim to respond to (1) and (2).

(a) To check if C_j is Z-maximal we normally need to determine for every demand point i if $r(E_j \cup \{i\}) \geq Z$ by using the EH algorithm or similar. However, in our situation if at any iteration k of EH the radius found, say $r(E_j \cup \{i\})_k \geq Z$, then we exit EH and check for the next i. There is no need to complete EH until the end as $r(E_j \cup \{i\}) > r(E_j \cup \{i\})_k \geq Z\ \forall\ k$. In addition, when EH is applied, in our situation, the search starts with the critical points forming circle C_j instead. This implementation cuts the computational burden considerably. These critical points are stored in a data structure when determining circles in $\prod_1$, $\prod_2$ and $\prod_3$ at the very beginning, so no extra computational time is really required.

(b) It is also observed that a large number of Z-maximal circles remain maximal from one iteration to the next. For instance, for $n = 439$ *TSPLib* and $p = 100$, on average <20% of the circles need to be tested at each iteration (Callaghan et al. 2017). It is therefore crucial to identify these circles as quickly as possible. Let Z_t be the value of Z at iteration t; then a maximal circle at iteration t

remains maximal at iteration $t + 1$ if $r(E_j) < Z_{t+1}$ according to Lemma 1 in Callaghan et al. (2017). This leads to not checking the expensive part which is $r(E_j \cup \{i\}) \geq Z_{t+1}\ \forall\ i$ as $Z_{t+1} < Z_t$.

(c) It is important to detect whether a circle is maximal or not quickly so as to avoid performing an unnecessary full check. According to Lemma 2 in Callaghan et al. (2017), if $\exists i \notin E_j \mid d(P_i, CT_j) < Z$ then C_j is not Z-maximal. Also there are some points that do not need to be checked. For instance if $i \notin E_j$ and $d(P_i, CT_j) \geq 2Z$, there is no need to find $r(E_j \cup \{i\})$ as $r(E_j \cup \{i\}) > Z$ (see Lemma 3 in Callaghan et al. 2017). In other words, if $d(P_i, CT_j) \geq 2Z\ \forall\ i \notin E_j$, C_j is systematically Z-maximal. This leads to performing the check for those points in $\{i \notin E_j | Z \leq d(P_i, CT_j) < 2Z\}$ only. If this set is empty and Lemma 2 does not apply, then C_j is Z-maximal. In other words, there is no need to check it.

(d) It is also useful to identify information from previous non-maximal circles. For instance, if a circle C_j at iteration t is found non-maximal, it means there was at least one point $i_1 \notin E_j$, say the qth point to be evaluated, that led to $r(E_j \cup \{i_1\}) < Z_t$. In the next iteration, it is important to start with i_1 to check whether or not $r(E_j \cup \{i_1\}) < Z_{t+1}$. This means the first $(q - 1)$ points are ignored leading to a saving of $(q - 1)$ unnecessary checks each involving the use of EH or its equivalent. This is considerable when applied to all circles.

For example, when these neighbourhood reduction schemes (a–d) are implemented for the $n = 439$ *TSPLib* instance for $p = 70$, 80, 90 and 100, a massive reduction in computational time is recorded. Individually (c) yields about 84% reduction, followed by (d) with a similar amount of 83%, with (a) producing just below 51% and finally (b) resulting in 26%. When all four are combined together following the ranked order of their individual performances *c-d-a-b*, the following cumulative percentage reductions of 84%, 90%, 96% and 97% are recorded. This shows that only a tiny 3% of the total time is required there by demonstrating the power of these neighbourhood reductions, which also enable the enhanced algorithm to solve to optimality larger instances.

To tackle (b) a detailed analysis showed that CPLEX consumes approximately 30% and 80% on average of the CPU time (Callaghan et al. 2017) for $n = 439$ *TSPLib* and $n = 575$ *TSPLib*, respectively. The higher values are found with larger values of p, reaching 99% for the largest problem. This was also noted to occur at the latter iterations mainly to guarantee optimality of $P_{op}(Z)$ at a given Z. In order to alleviate this issue, a scheme that adaptively guides the level of usage of CPLEX is added. This scheme aims to terminate CPLEX earlier if a compromise solution is considered to be good. To achieve this, a moving average over the last m iterations is recorded for both the computing time for identifying the maximal circles which we denote by $CPU(\max)$ and the time for running CPLEX denoted by $CPU(cplex)$. We define the ratio of these two times as $\lambda = \frac{CPU(\max)}{CPU(cplex)}$. If $\lambda \geq 1$, this shows that the time for identifying maximal circles is relatively higher. In this case, we solve the problem to optimality. However, if it is not the case, we set two additional levels for the duality gap as $GAP\ (\%) = \begin{cases} 1.0 & \text{if } \lambda \leq 0.4 \\ 0.5 & \text{if } 0.4 < \lambda < 1 \\ 0 & \text{otherwsie} \end{cases}$.

The above reduction schemes have contributed significantly in determining several optimal solutions for large instances up to $n = 1323$ and $p = 10, 20, ...100$ for the first time.

8.4.3.2 Relaxation-Based Approaches

The idea is to relax the original problem by successively solving small sub-problems that gradually increase in size until an optimal solution is found. Handler and Mirchandani (1979) originally discussed this idea of relaxation, but Chen and Handler (1987) proposed an algorithm where at each iteration, a demand point is added and the new augmented sub-problem is then solved again. The search continues until an optimal solution for the sub-problem happens to be feasible for the original problem. Chen and Chen (2009) revisited the problem by adding k demand points at a time. Three relaxation-based algorithms known as the improved, the binary and the reverse relaxation algorithms were presented. Callaghan (2016) performed an extensive experiment and concluded that the reverse relaxation algorithm is the most promising. This led to the design of three neighbourhood reductions to speed up this algorithm, so it can be used to solve larger instances either optimally or by providing tight lower bounds (Callaghan et al. 2018). For convenience, the reverse relaxation algorithm is briefly summarised in Fig. 8.4 as some of its steps form the basis of the following neighbourhood reductions.

(a) In Step 1, the initial subset is randomly chosen which may not be easy to replicate and may lead to either fast or slow convergence. One way forward is to construct such a subset deterministically reflecting the characteristics of the p-centre problem. As shown by Chen and Handler (1987), the smallest

1- Set the lower bound LB, the value of k, the set of potential circles J ($|J| = N_c$) and choose randomly a subset of demand points SUB ($|SUB| << n$).

2- Compute a_{ij} ($i = 1,...,|SUB|; j = 1,...,N_c$) based on LB and solve the corresponding set covering problem.

3- If the solution X is feasible (i.e., $\sum_{j \in J} X_j \leq p$) go to step 4;

Otherwise set the new value of LB to the smallest radius of a circle in SUB that is larger than LB, and go to Step 2.

4- If X is feasible for the original problem, the optimal solution is X and stop.

Otherwise add k furthest demand points to SUB and go to step 2.

Fig. 8.4 Main steps of Chen and Chen's algorithm (Chen and Chen 2009) possible value of $|SUB|$ required to yield a solution of p circles is to have at least $\text{Min}_{SUB} = \underset{r \geq 3}{\text{Min}} N_c(r) \geq p$ with $N_c(r) = C_1^r + C_2^r + C_3^r$.

The idea is to use the vertices of the convex hull as a guide and let CH denote such a subset. It can be shown that these points are not all necessarily critical points. We construct SUB as follows: Let i_1 be the furthest point to CH and set $SUB = \{i_1\}$. If $|SUB| < p$, allocate all demand points to their nearest point in SUB and identify the largest cluster. Choose the next point to add to SUB as the furthest point from this largest cluster, say i_2 and set $SUB = SUB \cup \{i_2\}$. This mechanism is repeated until $|SUB| = Min_{SUB}$ where the construction of the circles is performed. If the solution is not feasible in SUB (i.e., there are not enough circles), continue the addition of new points in the same way until a feasible solution is found.

(b) In Step 4, the added k points need not be necessarily the furthest points. For instance in the worst scenario, all or most of the furthest k points may belong to the same elongated cluster as shown in Fig. 8.5 where four points (P_1, P_2, P_3, P_4) are close to each other forming a small cluster, denoted by cluster 1. Once P_1 (the furthest from the solution) is added (in bold), a new much improved solution can be found showing that its contribution is important. However, the addition of the other three points, P_2, P_3 and P_4, will not affect this new solution and their inclusion will only add unnecessary computations. It is therefore important to identify such a cluster, so these three points do not need to be part of SUB. This rule can be applied to other clusters to identify the more representative points to add. The addition of the k new points can be performed either by including one point at a time followed by the evaluation of the new solution configuration or by adding all the k points in one go.

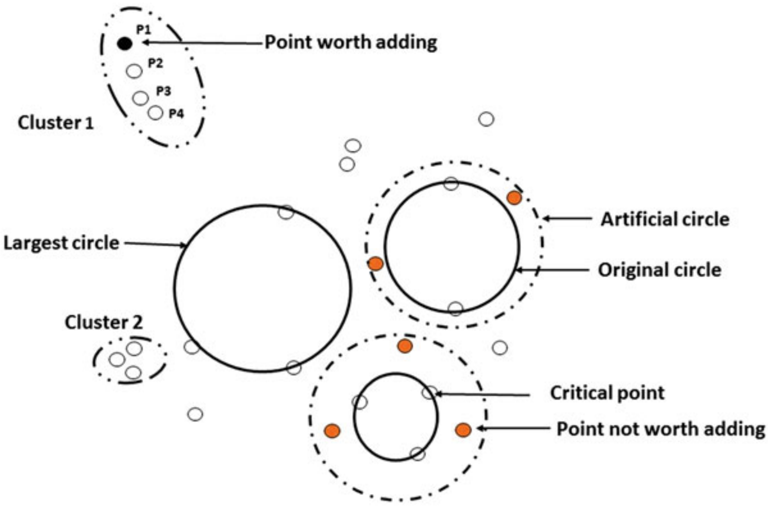


Fig. 8.5 Effect of point clusters and artificial circle in the addition of the new points

To speed up the search even more, artificial circles are constructed whose centres remain the centres of the existing circles but whose radii are increased to Z. These are the dotted circles shown in Fig. 8.5. The checking is based on those uncovered points away from the artificial circles instead of using all the initial uncovered points. As an example when tested on the $n = 439$ *TSPLib* instance for $p = 10, 20, . \ldots, 100$, this neighbourhood reduction eliminates over 10% of computing time.

(c) Also, the value of k in Step 4 does not have to remain constant at each iteration and for all instances. This parameter can be made dynamic at a given iteration t. Let N_{unc}^t denote the number of uncovered points at iteration t. We can then define $k(t) = f\left(p, N_{unc}^t\right)$.

(d) In Step 3, *LB* is updated by taking the next radius larger than the current value of *LB*. Though this is mathematically correct and will end up with the final value, the search may use too many updates, many of which turn out to be unnecessary. Instead of choosing the next largest, we adopt a jump-based scheme to select the $jump^{th}$ largest. A similar but simpler idea was initially proposed and successfully implemented for the vertex p-centre problem by Al-Khedhairi and Salhi (2005) where a jump of two was used. This is extended by defining the jump as a function of a predefined maximum jump size, and the ratio $\frac{N_{unc}^t}{n}$. This jump-based scheme systematically learns as the search progresses. Note that the obtained solution may provide an upper bound instead of a lower bound. In this case, a backtracking is required by evaluating the values of *LB* between $LB(t)$ which was definitely a lower bound and $LB(t + 1)$ which happens to be an upper bound. Here at most $(jump(t) - 1)$ jumps may be required to guarantee optimality.

These neighbourhood reductions were found very promising when compared to the original implementation. For instance, when tested on the $n = 439$ *TSPLib* instance with $p = 10, 20, \ldots, 100$, an average reduction in computing time of nearly 90%, with the smallest being just over 50% and the largest nearly 97%, was observed.

The incorporation of all the above reduction schemes enables the algorithm to solve most of the larger instances optimally within 3 h of computing time. For those instances where the optimal solution could not be guaranteed, a tight lower bound was recorded, which may be used in the future for assessing new heuristics.

8.5 Neighbourhood Reduction Highlights and Conclusions

In this chapter, a brief review of both the vertex and planar p-centre problems is given with an emphasis on contributions made by Zvi Drezner. Several neighbourhood reductions especially designed for these two location problems are then discussed with the aim to enhance the efficiency of existing algorithms or in assisting at designing a more effective heuristic or optimal algorithm. For the vertex

p-centre problem, a series of neighbourhood reduction rules are presented that have enhanced the performance of existing optimal algorithms considerably thus enabling the exact solution of well-known ORLib instances ($n = 100$ to 900) and the $n = 1060$ TSPLib instance to be obtained in much faster time than before. Similarly, for the continuous p-centre problem, four TSPLib instances varying in size from $n = 439$ to 1323 with $p = 10, 20, \ldots, 100$ are used as a platform to demonstrate the effectiveness of the proposed neighbourhood reduction schemes. Enhanced VNS and perturbation heuristics are now much more effective than before. Also, Drezner's optimal algorithm and the relaxation-based methods of Chen and Chen are now able to provide optimal solutions for the first time for many of the largest instances tested, and tight lower bounds for the rest.

The implementation of these exact methods using the neighbourhood reduction schemes discussed in this chapter can be made even faster if a tighter initial solution is provided, say by a powerful metaheuristic. In addition, as these schemes tend to cut on computational time by avoiding time wastage, if the same allowed computing time is used as the stopping criterion for the enhanced version, the new solution might easily improve on the original one as many more iterations would be performed leading to more moves being evaluated.

It is also interesting to observe that in Sect. 8.4.3.1(c), the checking area in Lemma 3 and the recording of the points that define non-maximal circles can be made slightly tighter as recently pointed out by Plastria (2017).

There exist a few variations of the p-centre problem. In the conditional p-centre problem some (say q) facilities already exist and the objective is to locate p new facilities in addition to the existing q facilities. Minieka (1980) presented the problem while Drezner (1989, 1995) defined it formally as the (p, q) centre problem, and put forward a binary search to solve it. Chen and Chen (2010) also adapted their algorithm discussed in Sect. 8.4.3 to tackle this problem. Another related problem is when each demand point needs to be covered by at least α facilities. This problem, initially proposed by Krumke (1995), is known as the α-neighbourhood p-centre problem, and has its applications in the case of facility disruption. Chen and Chen (2013) used Minieka's algorithm and modified their relaxation method described in Sect. 8.4.3 to solve this problem. Very recently, Callaghan et al. (2018) studied both variants by adapting the powerful reduction schemes discussed in Sect. 8.4.3 so that larger problems can now be solved to optimality for the first time.

One possible extension is to adapt the neighbourhood reductions used for the continuous problem in Sect. 8.4.3 that rely on maximal circles, cluster points and artificial circles, to the discrete problem though this case can be solved by other means. It is also worth noting that reduction schemes do exist for other combinatorial and global optimisation problems. In general the more constrained the problem is, the more significant the impact of neighbourhood reduction can be. For example, in the vehicle routing problem, a saving on CPU time of up to 85% was recorded without a significant loss in solution quality (Salhi and Sari 1997; Sze et al. 2016, 2017).

It is necessary to mention in conclusion that the use of neighbourhood reduction techniques may adversely affect the solution quality. The aim is therefore to construct such schemes which only exclude moves that have a high probability of not harming the quality of the solution. This risk presents an exciting challenge of finding the right balance between a strong neighbourhood reduction (remove as much as possible) and maintaining solution quality.

Acknowledgments The first author is grateful to his PhD students, namely Drs AI. Al-Khedhairi, AM. Alharbi, CA. Irawan, AM. Elshaikh and B. Callaghan. Their contributions in this area through the last 15 years are very much appreciated.

References

Al-Harbi, A. M. (2010). *Combining heuristics and exact approaches for the vertex p-centre problem and other related location problems*. PhD Dissertation, University of Kent, Canterbury, UK.

Al-Khedhairi, A., & Salhi, S. (2005). Enhancements to two exact algorithms for solving the vertex p-center problem. *Journal of Mathematical Modelling Algorithms, 4*, 129–147.

Brimberg, J., Drezner, Z., Mladenović, N., & Salhi, S. (2014). A new local search for continuous location problems. *European Journal of Operational Research, 232*, 256–265.

Brimberg, J., Drezner, Z., Mladenović, N., & Salhi, S. (2017). Using injection points in reformulation local search for solving continuous location problems. *Yugoslav Journal of Operations Research, 27*, 291–300.

Callaghan, B. (2016). *An investigation into exact methods for the continuous p-centre problem and its related problems*. PhD thesis, University of Kent, Canterbury, UK

Callaghan, B., Salhi, S., & Nagy, G. (2017). Speeding up the optimal method of Drezner for the p-centre problem in the plane. *European Journal of Operational Research, 257*, 722–734.

Callaghan, B., Salhi, S., & Brimberg, J. (2018). Optimal solutions for the continuous p-centre problem and related α-neighbourhood and conditional problems: A relaxation-based algorithm. *Journal of the Operational Research Society, 70*, 192–211. https://doi.org/10.10880/01605682.2017.1421854.

Chen, R. (1983). Solution of minisum and minimax location-allocation problems with Euclidean distances. *Naval Research Logistics Quarterly, 30*, 449–459.

Chen, D., & Chen, R. (2009). New relaxation-based algorithms for the optimal solution of the continuous and discrete *p*-center problems. *Computers & Operations Research, 36*, 1646–1655.

Chen, D., & Chen, R. (2010). A relaxation-based algorithm for solving the conditional p-center problem. *Operations Research Letters, 38*, 215–217.

Chen, D., & Chen, R. (2013). Optimal algorithms for the α-neighbor *p*-center problem. *European Journal of Operational Research, 225*, 36–43.

Chen, R., & Handler, G. Y. (1987). Relaxation methods for the solution of the minimax location-allocation problem in Euclidean space. *Naval Research Logistics Quarterly, 34*, 775–788.

Cooper, L. (1964). Heuristic methods for location-allocation problem. *SIAM Review, 6*, 37–53.

Daskin, M. S. (2000). A new approach to solving the vertex *p*-center problem to optimality: Algorithm and computational results. *Communications of the Operations Research Society of Japan, 45*, 428–436.

Drezner, Z. (1984a). The p-center problem-heuristic and optimal algorithms. *Journal of the Operational Research Society, 35*, 741–748.

Drezner, Z. (1984b). The planar two-center and two-median problems. *Transportation Science, 18*, 351–361.

Drezner, Z. (1989). Conditional *p*-center problems. *Transportation Science, 23*, 51–53.

Drezner, Z. (1995). On the conditional p-median problem. *Computers and Operations Research, 22*, 525–530.

Drezner, Z. (2011). Continuous center problems. In H. A. Eiselt & V. Marianov (Eds.), *Foundations of location analysis* (pp. 63–78). Boston, MA: Springer.

Eiselt, H. A., & Charlesworth, G. (1986). A note on *p*-center problems in the plane. *Transportation Science, 20*, 130–133.

Elloumi, S., Labbe, M., & Pochet, Y. (2004). A new formulation and resolution method for the *p*-center problem. *INFORMS Journal on Computing, 16*, 84–94.

Elshaikh A. (2014). *Adaptive heuristic methods for the continuous p-centre problem*. PhD Dissertation, University of Kent, Canterbury, UK.

Elshaikh, A., Salhi, S., & Nagy, G. (2015). The continuous p-centre problem: An investigation into variable neighbourhood search with memory. *European Journal of Operational Research, 241*, 606–621.

Elshaikh, A., Salhi, S., Brimberg, J., Mladenovic, N., Callaghan, B., & Nagy, G. (2016). An adaptive perturbation-based heuristic: An application to the continuous p-center location problems. *Computers & Operations Research, 75*, 1–11.

Elshaikh, A., Salhi, S., & Brimberg, J. (2018). *A reformulation local search for the continuous p-center location problem*. Submitted (internal report).

Elzinga, J., & Hearn, D. W. (1972). Geometrical solutions for some minimax location problems. *Transportation Science, 6*, 379–394.

Gamal, M. D. H., & Salhi, S. (2001). Constructive heuristics for the uncapacitated continuous location-allocation problem. *Journal of the Operational Research Society, 52*, 821–829.

Hakimi, S. L. (1964). Optimum locations of switching centers and the absolute centers and medians of a graph. *Operations Research, 12*, 450–459.

Handler, G. Y., & Mirchandani, P. B. (1979). *Locations on networks: Theory and algorithms*. Cambridge, MA: MIT Press.

Hansen, P., Mladenovic, N., Brimberg, J., & Moreno-Perez, J. A. (2010). Variable neighbourhood search. In M. Gendreau & J.-Y. Potvin (Eds.), *Handbook of metaheuristics* (Vol. 61–86, 2nd ed.). New York, NY: Springer.

Hearn, D. W., & Vijay, J. (1982). Efficient algorithms for the (weighted) minimum circle problem. *Operations Research, 30*, 777–795.

Ilhan, T., Pinar, M. (2001). *An efficient exact algorithm for the vertex p-center problem*. http://www.ie.bilkent.edu.tr/~mustafap/pubs/

Irawan, C. A., Salhi, S., & Drezner, Z. (2016). Hybrid metaheuristics with VNS and exact methods: Application to large unconditional and conditional vertex p-centre problems. *Journal of Heuristics, 22*, 507–537.

Kariv, O., & Hakimi, S. L. (1979). An algorithmic approach to network location problem, I: The p-centers. *SIAM Journal on Applied Mathematics, 37*, 539–560.

Krumke, S. O. (1995). On a generalization of the p-centre problem. *Information Processing Letters, 56*, 67–71.

Luis, M., Salhi, S., & Nagy, G. (2009). Region-rejection based heuristics for the capacitated multi-source Weber problem. *Computers & Operations Research, 36*, 2007–2017.

Megiddo, N., & Supowit, K. J. (1984). On the complexity of some common geometric location problems. *SIAM Journal of Computing, 13*, 182–196.

Minieka, E. (1970). The m-center problem. *SIAM Review, 12*, 138–139.

Minieka, E. (1980). Conditional centers and medians on a graph. *Networks, 10*, 265–272.

Mladenović, N., Labbé, M., & Hansen, P. (2003). Solving the p-center problem with tabu search and variable neighbourhood search. *Networks, 42*, 48–64.

Plastria, F. (2002). Continuous covering location problem. In Z. Drezner (Ed.), *Facility location: Applications and theory* (pp. 37–79). New York, NY: Springer.

Plastria, F. (2017). Private communication.

Salhi, S. (1997). A perturbation heuristic for a class of location problems. *Journal of the Operational Research Society, 48*, 1233–1240.

Salhi, S. (2017). *Heuristic search: The emerging science of problem solving*. Cham, Switzerland: Palgrave McMillan (published by Springer).

Salhi, S., & Al-Khedhairi, A. (2010). Integrating heuristic information into exact methods: The case of the vertex *p*-centre problem. *Journal of the Operational Research Society, 61*, 1619–1631.

Salhi, S., & Sari, M. (1997). A multi-level composite heuristic for the multi-depot vehicle fleet mix problem. *European Journal of Operational Research, 103*, 95–112.

Sze, J. F., Salhi, S., & Wassan, N. (2016). A hybridisation of adaptive variable neighbourhood search and large neighbourhood: Application to the vehicle routing problem. *Expert System with Applications, 65*, 383–397.

Sze, J. F., Salhi, S., & Wassan, N. (2017). The cumulative capacitated vehicle routing problem with min-sum and min-max objectives: An effective hybridisation of adaptive search and large neighbourhood search. *Transportation Research Part B, 101*, 162–184.

Xu, S., Freund, R., & Sun, J. (2003). Solution methodologies for the smallest enclosing circle problem. *Computational Optimization and Applications, 25*, 283–292.

Chapter 9
Innovations in Statistical Analysis and Genetic Algorithms

Taly Dawn Drezner

9.1 Opening: Background and a Role Model

I am very lucky to have Zvi Drezner as a father. He is a warm, dedicated, and engaged dad, as well as a successful researcher and a true role model in both life and work. Curious about the world, I followed his (and my mother's) footsteps into academia, though with a focus in ecology (more precisely, the life science component of Geography, called biogeography). Through the years we have conversed about many topics from natural history and science, to astronomy, statistics, and genetic algorithms, among many others. From these conversations, we were both exposed to new ideas through stimulating and fun conversations. I am very fortunate to have these very special father–daughter times, which I cherish.

Zvi's family discussions about research don't end with wife Tammy and daughter Taly; one day my 9-year old son sat on Grandpa Zvi's lap looking at an image on the computer of population and facility locations in Orange County, California, which my son curiously asked about. After grandpa's explanation, young Ryan said, "There should be more facilities in denser areas!" Grandpa was so impressed, that he developed and together we wrote up the answer to that comment, which was published (Drezner et al. 2019) shortly after Ryan's 10th birthday.

T. D. Drezner (✉)
Department of Geography, York University, Toronto, ON, Canada
e-mail: drezner@yorku.ca

H. A. Eiselt, V. Marianov (eds.), *Contributions to Location Analysis*, International Series in Operations Research & Management Science 281,
https://doi.org/10.1007/978-3-030-19111-5_9

9.2 Overview

Through the many conversations Zvi Drezner and I have had, several innovations have emerged. Two papers represent the intersection of a biologically trained scientist with genetic algorithm approaches that mimic biological principles, resulting in new approaches and the identification of weaknesses such as terminology that is used incorrectly in genetic algorithms relative to the biological counterparts that those concepts mimic.

Our other innovations involve improvement in methods, data collection, and statistics. I developed a new approach rather intuitively for collecting data to estimate transition points in populations (Drezner 2008). Statistically unconfirmed, my father and I developed the statistical foundation with the help of an order statistics specialist. We also developed an improvement to the Bonferroni statistical correction, which adjusts output for false positive results that are generated as a byproduct of running multiple statistical tests, a common occurrence in ecology and other disciplines. We updated the Bonferroni approach, creating a far less conservative version for more practical use.

9.3 Innovations in Genetic Algorithms

Zvi Drezner has often focused on optimization problems that typically have too many solutions to check individually or solve by a branch and bound algorithm. Thus, heuristic approaches are needed to find good solutions so as to solve problems with a reasonable amount of computing time and resources. There are several commonly used heuristic approaches including tabu search (Glover and Laguna 1997), simulated annealing (Kirkpatrick et al. 1983), and variable neighborhood search (Hansen and Mladenović 2001). All of these approaches select a starting solution and then seek to improve that solution with a possibly better one in the next iteration. By comparison, the genetic algorithm heuristic approach includes multiple solutions at a time and seeks good solutions through a merging process of other solutions, and isolating better solutions for further improvement. In many cases, better results are obtained through a merging process.

Zvi Drezner has worked extensively with genetic algorithms (GAs). The GA approach mimics biological processes for algorithm development to solve optimization problems. GAs take a set of solutions and "mate" them to find improved solutions. Mating involves combining elements of each parent to create a new solution, the offspring. Each solution is an individual in the population. The process of mating, producing offspring, and adding them into the population is repeated a given number of times. Fitness in GAs is the value of the objective function and the goal is to find the solution with the best objective function value. The best population member is the GA outcome. All GAs include: (1) rules for parent selection, (2) procedures for merging or combining the two parents and generating offspring, and (3) a decision rule to determine which offspring are to be kept (at the expense of an existing member) and which should be removed (death or becoming non-reproductive). Any of these three components can be modified to improve the performance of the algorithm. Being biologically trained, I discussed biological phenomena and applications with Zvi Drezner and we developed new approaches for GAs based on biological principles. We also identify terms used in GAs that are inconsistent with their usage in biology.

In both of our GA papers, we pursue new approaches for parent selection. Most GAs select two parents at random, though in nature this is not a random process. My parents (Drezner and Drezner 2006) first designated about half of the population as males and half females. Parents are then selected randomly, one from each gender. Although very basic, this replication of nature yielded improved results. Drezner and Marcoulides (2003) suggested selecting one parent at random, and then selecting the second parent that is most dissimilar to the first parent from a random subset. In our work, we proposed two non-random parent selection rules, one mimicking the alpha male phenomenon in nature (one male to several females). The other was inspired by ideas of female choice, where one (the female) is randomly selected but the better of two randomly selected males is chosen for mating with a pre-specified probability of π.

9.3.1 Biological Background

A gene is a piece of DNA that influences a trait in that organism (Freeman et al. 2014). The same gene may have several forms, called alleles (Freeman et al. 2014). For example, in Mendel's famous pea experiments, the gene for seed shape included an allele for round seeds and an allele for wrinkled seeds (Mendel 1866; Freeman et al. 2014).

Every individual possesses a unique combination of genetically determined traits, some of which may benefit it through its life. If those alleles translate into beneficial traits that result in the production of more offspring, then more individuals in the next generation will carry those traits (alleles). By comparison, other traits may confer disadvantages that reduce an individual's fitness. Fitness is the number of offspring an individual can produce relative to other individuals (Freeman et al. 2014). If these disadvantages translate to reduced reproductive success, those traits will be represented in proportionally fewer individuals in the next generation. Thus the next generation will typically have more individuals that carry beneficial traits and fewer individuals with the disadvantageous traits. Through generations and time, the genetic make-up of a population changes. Natural selection occurs in populations through time. These principles of inheritance and the success of more fit individuals have been applied in GAs to solve optimization problems.

9.3.2 The Female Choice Approach

In nature, females may choose their mate in a variety of ways, including through visual cues such as coloring and appearance, as these can be signs of health and access to food, or females may choose mates through preference of a particular male's territory (e.g., better resources) (Rosser 1992; McGraw and Ardia 2003). In our female choice-inspired study (Drezner and Drezner 2018), two random individuals (males) are selected. The better individual is selected as the first parent at a pre-specified probability $0 \leq \pi \leq 1$ of the time. Otherwise, the other individual is selected as the first parent. When $\pi = 1$, the better population member is always selected as the first parent, and when $\pi = 0$ the inferior one is always selected. The other sampled individual is returned to the population, and then the second parent (the female) is selected by the Drezner and Marcoulides (2003) principle that selects a more dissimilar mate, also consistent with the biological principle of inbreeding depression (Edmands 2007; Fenster and Galloway 2000). The GA process then commences, mating these two parents to produce an offspring. This was tested on π with values of 0, 25, 50, 75, and 100% to find the best value of π.

Extensive experiments were performed on the planar p-median problem (the multi-source Weber problem) (Brimberg et al. 2000), and the quadratic assignment problem (Drezner 2015). The planar p-median was tested on three problems with $n = 654$, 1060, and 3038 demand points (Reinelt 1991) for a total of 57 instances. Each instance was run 10 times and the best and average solution was recorded. The best solutions were obtained for $\pi = 0$. For example, the best known solution was found for the $n = 654$ instances for all 17 instances in all 10 runs. Unlike our original biological premise, the quality of the solution generally deteriorates as π increases. Four new best known solutions were obtained for the $n = 3038$ instances (for $p = 250$, 350, 450, 500). The quadratic assignment problem was tested on the (de Carvalho and Rahmann 2006) 14 instances. The best results were obtained for $\pi = 0.25$. One new best known solution was obtained for quadratic assignment problem instance BL144 (Drezner and Drezner 2018).

9.3.3 The Alpha Male Approach

In nature, many species have social structures with a dominant male that sires many or all offspring in a group of females (Freeman et al. 2014). Species with dominant males include various mammals such as baboons (Galbany et al. 2015), rodents (Farentinos 1980), horses (Wolter et al. 2014), and seals (Hoelzel et al. 1999), as well as other animal groups including birds (Polak 2006), insects (McDermott et al. 2014), and fish (Solomon-Lane et al. 2014). Male to male competition or combat may determine which male gets to mate, sometimes with multiple females (Haley et al. 1994). In such situations the stronger, healthier male is most likely to be the victor. Initially, we followed this principle by selecting the best population members as alpha males in developing the approach. However, randomly selected alpha males yielded better results. In Drezner and Drezner (2019) we began with a population of 100 and randomly selected k individuals of the population as alpha males, with the rest of the individuals defined as female ($100 - k$). The value of k is a parameter of the algorithm. Each female was randomly paired with one of the alpha males for mating and then an offspring was generated. Thus, $100 - k$ new offspring were produced, resulting in $200 - k$ individuals (the original 100 plus $100 - k$ offspring). The 100 best individuals were carried forward to constitute the next population. We tested many values of k to find the best performing one. This process was repeated and each time k members were randomly selected as alpha males (they may have been, e.g., females, previously). This process was repeated a pre-specified number of times so that run times are comparable to previous experiments (Drezner and Misevičius 2013).

We tested the alpha male approach on the de Carvalho and Rahmann (2006) 14 instances of the quadratic assignment problem. We tested fixed values of k in every iteration and randomly generated the value of k in a range in every generation. Generating the value of k in a range provided better results than a fixed value of k. The best results were obtained around $k = 25$ in a range of 5, for example, $20 \leq k \leq 25$. For $k = 25$ there are, on average, about 3 females associated with each alpha male. Values vary in the animal world, but are often one male to a single digit number of females (Lukas and Clutton-Brock 2014). Two new best known solutions were obtained for the quadratic assignment problem instances BL100 and BL121 (Drezner and Drezner 2019). Our results show that randomly selected alpha males rather than better fit ones yielded the best results for solving our quadratic assignment instances. We also observed that when the female mates with an alpha male that is more dissimilar to her, the results are better, consistent with biological observations that breeding with close relatives produces less fit offspring, termed inbreeding depression in populations (Freeman et al. 2014). We also observe that when the number of females per alpha male fluctuates (e.g., over time), the results are better than with a fixed number of females, which necessarily occurs in animal populations.

9.3.4 Genetic Algorithm Terms and Parallel Biological Principles

In biology, the term *fitness* refers to how many viable offspring any given individual can produce relative to other members of the same species (Freeman et al. 2014). Fitness is related to the environment and to interactions with other species, both positively (e.g., mutualisms, facilitation) and negatively (e.g., competition). By comparison, in GAs fitness is not related to offspring production (foundational in biology) nor to the environment, as the environment is not a component of GAs. Rather, in GAs fitness represents the value of the objective function and the goal is to find the most fit population member (best solution), i.e., how good an individual is in its value of the objective solution. Thus the term fitness strongly diverges from its biological origins.

The term *invasion* (or invasive species) relates to the establishment of a species in a new place beyond its native range, where it negatively impacts (sometimes dramatically) the native species and community (e.g., Kent et al. 2018). *Introduced* species fall under a similar definition but their effect on the native community is less destructive. Both of these refer to *new* species to an area. In contrast, *immigration* is when individuals of the *same* species move from one population to another already established population. Immigration is important for increasing gene flow and fitness in recipient populations, while the term invasion is inherently negative and involves new, competing species that do not contribute to the gene pool of another species, in this case the one of interest (Whiteley et al. 2015). In GAs, a few outsiders may be added to increase genetic diversity, akin to the process of *immigration* in biology, but the term *invasion* (perhaps derived from the human idea of an attack by a foreign army) has been misapplied to this process. Since GAs only involve one species, they do not include interspecific (between species) interactions.

Both in biology and in common usage, the term *generation* refers to a large segment of a population made up of similarly aged individuals, or a cohort. In GAs, however, the term is essentially used in place of the term "birth" or to describe a single offspring, which is "generated." Thus a family with three children would be described as being composed of two generations biologically (parents, children), while in GAs, each child represents a different generation.

There are also examples of GA results that parallel biological phenomena more closely. For example, individuals that are too similar or too different yield poor offspring in both GA and in the natural world (inbreeding and outbreeding depression). For example, fitness can decline when reproduction occurs between genetically distant members of the same species (*outbreeding depression*) (Edmands 2007; Fenster and Galloway 2000). In GAs, the equivalents of both inbreeding and outbreeding depression also show reduced success. Drezner and Drezner (2018) review many of these applications and many more parallels between GAs and the biological processes that inspire them.

9.4 Innovative Statistical Analysis

Zvi Drezner and I also have two statistical innovations. (1) We developed the statistical underpinnings of a new field approach designed to estimate transition points in a population (of any species). The principle is not specific to the life sciences and can be used for numerous applications that require transition point (from one stage to another) assessment or quantification in a population where individual transition age is variable. (2) Inspired by our exposure to order statistics, we developed a new and far less conservative approach for dealing with type I (false positive) errors that result when multiple statistical tests are run.

9.4.1 Estimating Transition Between Stages

Life cycle transition points (such as the age when juveniles transition to adults) are foundational in ecology. I estimated the juvenile–adult transition age (when reproduction begins) in a long-lived species (e.g., 150–250 years). Estimating the mean age at which a population becomes reproductive is complicated; sampling for the youngest reproductive individual yields an outlier (whose value is related to sample size), thus not representing a measure of central tendency for the population. Sampling individuals over many years can be difficult in long-lived species. In a population of 400 individuals whose lifespans average 200 years, only two individuals per year will transition, requiring decades of observations. Even sample size, foundational in statistics, is itself difficult to establish as very young juveniles are irrelevant, as are older adults. Distinguishing those individuals that are statistically meaningful and which are not is unclear. I now briefly introduce the species I used to develop the new approach for context and then I discuss the methodology and its statistical foundations.

The saguaro (pronounced "swah-roh") cactus (*Carnegiea gigantea*, Fig. 9.1) is a keystone species and a charismatic plant that symbolizes the desert, with branches ("arms") that seem to reach up to the sky (Drezner 2014). The age of transition from juvenile to reproductive adult varies with environmental conditions. Several of the

Fig. 9.1 Different life stages of the saguaro cactus (*Carnegiea gigantea*). Left: single-stemmed reproductive adults (the two taller stems, each with reproductive structures visible at their apex). Middle: juvenile, non-reproductive plant (no reproductive structures observed). Right: a plant displaying the branched form, with reproductive structures visible on the branches as well. Image taken by TD Drezner at the Kofa National Wildlife Refuge, Arizona, in the Sonoran Desert

plants in Fig. 9.1 show the presence of fruits (irregular small features at the tops of the plants). As individuals get older, branches eventually develop, where each additional branch essentially doubles the number of seeds that can be produced each season by that plant (Steenbergh and Lowe 1983). Just as the age when reproduction starts varies over the species' range, branching is also variable. For example, in dry areas, plants tend to be under-branched, while where more water is available, plants use that water to increase the number of seeds they produce by branching (Yeaton et al. 1980).

In order to establish the age at which these plants begin to produce offspring, the oldest juveniles (e.g., 5 oldest) with no reproductive structures (flowers, fruits), and the same number of the youngest adults that are reproductive, were sampled (Drezner 2008). These are the individuals that are near the transition point (Drezner 2008). Individual age was estimated from height using a site-specific model developed for this particular species (Drezner 2003). I sampled the oldest pre-transition and the youngest post-transition members in four environmentally distinct populations for comparison. Each population was extensively searched such that the number of plants examined to isolate the oldest pre- and youngest post-transition individuals was large. The average age of these (e.g., 10) plants yielded the estimate for the mean transition age in that population. This methodology was later employed in a second study to find the transition from columnar form (no branches) to the branched form (Drezner 2013b). In the juvenile–adult transition study, the five (k) shortest flowering individuals and the five tallest non-flowering individuals were extracted. For the follow-up study on the transition to branched form, I used $k = 10$ (Drezner 2013b). When originally published, this approach was presented as it was carried out, but it was developed only intuitively (Drezner 2008) and was lacking statistical justification. We developed the statistical underpinnings of the technique. We assume the following (Drezner et al. 2015):

1. Once an individual has transitioned to the next stage, it does not revert back.
2. The transition age is normally distributed. Other distributions can be analyzed in the same way.
3. There are about the same number of individuals at a given age in the range covered by the $2k$ observations.

The total number of individuals observed in a population (typically identified as the "sample size"), those assessed, and then included or excluded from the final $2k$ list is not relevant to our analysis. For example, if a million very young juveniles or very old adults were also present in the population of interest, it would not change our analysis or our results. We found the transition distribution as follows:

Let the k expected smallest values of the standardized Normal distribution for a sample of n be $a_1(n) \le a_2(n) \le \ldots \le a_k(n)$. These values were found by extensive simulations. The unknown mean and standard deviation of the distribution of the transition age are μ and σ. The data consist of the k youngest post-transition individuals with ages $a_1 \le \ldots \le a_k$ and k oldest pre-transition individuals with ages $b_1 \ge \ldots \ge b_k$.

To fit the data for a specific sample size n to the order statistics requires finding the μ and σ that satisfy the following set of equations as closely as possible:

$$a_j(n)\sigma + \mu = a_j; \; -a_j(n)\sigma + \mu = b_j \text{ for } j = 1, \ldots, k. \tag{9.1}$$

The solution that minimizes the sum of squares of errors in these equations can be obtained by solving a simple linear regression where μ is the $y-$intercept and σ is the slope. The solution to this simple linear regression is based on several values:

$$\mu = \frac{\sum_{j=1}^{k} a_j + \sum_{j=1}^{k} b_j}{2k}; \; S_x = 2\sum_{j=1}^{k} a_j^2(n); \;\; S_y = \sum_{j=1}^{k} \left\{ (a_j - \mu)^2 + (b_j - \mu)^2 \right\};$$

$$S_{xy} = \sum_{j=1}^{k} a_j(n) \left(a_j - b_j \right). \tag{9.2}$$

In addition to the mean μ calculated in (9.2)

$$\sigma = \frac{S_{xy}}{S_x}. \tag{9.3}$$

The standard errors of μ and σ are

$$SE(\mu) = \sqrt{\frac{S_y - \sigma S_{xy}}{4k(k-1)}}; \;\; SE(\sigma) = SE(\mu)\sqrt{\frac{2k}{S_x}}. \tag{9.4}$$

We calculated the correlation coefficient r and found the p-value of the regression

$$r = \sigma\sqrt{\frac{S_x}{S_y}}. \tag{9.5}$$

When the analysis is repeated for various values of n, the n that yields the largest value of r is selected for calculating the values in Eqs. (9.2) and (9.3). A spreadsheet that automatically calculates μ, σ, and their standard errors for the transition distribution for any $k \leq 10$ is available at http://onlinelibrary.wiley.com/doi/10.1002/env.2351/suppinfo (Drezner et al. 2015). The spreadsheet calculates the values by Eqs. (9.2)–(9.5) for every $10 \leq n \leq 200$, selects the sample size resulting in the largest value of the correlation coefficient r, and records the parameters of the transition distribution for the selected n. Researchers can insert their observations to obtain the results. For complete details see Drezner et al. (2015).

The main results (Drezner et al. 2015) include:

1. The originally developed methodology is statistically sound and offers a new, practical, and robust approach for estimating transition points such as in years of age.
2. We calculated the underlying statistics and confirm that the measure of central tendency originally estimated (Drezner 2008, 2013b) is indeed the mean (Drezner et al. 2015).
3. While the original means were correct, the standard errors reported in the two original studies were not, and interestingly, all eight erroneous values originally reported (Drezner 2008, 2013b) were higher than the updated, correct values. Even with means as high as 139 years, the updated SE values in all 8 field-collected datasets were less than 1 year, including those based on only 5 pre- and 5 post-transition values (Drezner 2008; Drezner et al. 2015).
4. Despite the small number of values used in the final calculations (recognizing that those are derived from a much larger number of measurements), the results are robust and insensitive to changes in k. The start of branching study used $k = 10$; we compared those results with $k = 5$, 6, 7, 8, and 9 individuals from each stage. The estimated mean ages for the branching transition in the four sites with $k = 10$ ($k = 5$ in parentheses following) were: 77.8 (79.1) years of age, 95.9 (97.2), 102.9 (102.8), and 139.2 (139.9) (Drezner et al. 2015).
5. This technique makes assessing transitions fast and efficient, requiring only one field season, and can be easily calculated with our spreadsheet http://onlinelibrary.wiley.com/doi/10.1002/env.2351/suppinfo (Drezner et al. 2015).

9.4.2 *The Correlated Bonferroni Technique*

Running multiple statistical tests yields multiple p-values, creating a statistical challenge as the more tests that are run, the more likely a significant result will be obtained by chance. Twenty results would be expected to have one significant ($p < 0.05$) result by chance alone. In fact, there is a 64% chance that at least one result in 20 would be statistically significant. The Bonferroni technique (BT) was developed to correct for false positive (type I) errors. It approximately divides α (e.g., 0.05) by the number of significant results (k), and only results with p-values lower than the new threshold pass the test (Bonferroni 1936; Miller 1981). The BT was updated by Holm (1979) who proposed the less conservative sequential Bonferroni technique (SeqBT) (Rice 1989) where all significant p-values are placed in ascending order, recalculating $\frac{\alpha}{k}$ for each test anew in sequence. Thus, the tenth smallest p-value must be less than approximately $\frac{\alpha}{10}$ (see Table 9.1), the ninth smallest p-value less than $\frac{\alpha}{9}$, etc. The SeqBT has since been adopted in many studies (e.g., Drezner 2013a; Gittman et al. 2016; Snyder and Stepien 2017).

Concerns have been expressed about even the less conservative SeqBT. Not only is it used inconsistently, but the decision-making process for applying it is

Table 9.1 Critical values for $3 \leq s \leq 10$ significant results compared with BT

k	$s = 3$ $\rho = 0.383$	$s = 4$ $\rho = 0.677$	$s = 5$ $\rho = 0.800$	$s = 6$ $\rho = 0.862$	$s = 7$ $\rho = 0.897$	$s = 8$ $\rho = 0.919$	$s = 9$ $\rho = 0.933$	$s = 10$ $\rho = 0.943$	SeqBT/BT[a]
1	0.0500	0.0500	0.0500	0.0500	0.0500	0.0500	0.0500	0.0500	0.0500
2	0.0263	0.0290	0.0320	0.0343	0.0359	0.0370	0.0378	0.0383	0.0253
3	0.0183	0.0221	0.0263	0.0294	0.0316	0.0331	0.0342	0.0349	0.0170
4		0.0179	0.0223	0.0256	0.0279	0.0294	0.0305	0.0313	0.0127
5			0.0195	0.0228	0.0250	0.0266	0.0276	0.0284	0.0102
6				0.0207	0.0229	0.0244	0.0254	0.0262	0.0085
7					0.0212	0.0227	0.0238	0.0245	0.0073
8						0.0214	0.0224	0.0232	0.0064
9							0.0214	0.0221	0.0057
10								0.0212	0.0051

[a]For the BT the value for $k = s$ significant results is used and for SeqBT the values up to k are used

uncertain (Cabin and Mitchell 2000). For example, should two tables, each with 10 significant results, be pooled (thus $k = 20$), or does each set of analyses stand on its own (Cabin and Mitchell 2000)? In a survey of editors of three highly respected life science journals, there was no consensus on the usage and application of the SeqBT (Cabin and Mitchell 2000). Also troubling is that more in-depth analysis is discouraged with the SeqBT (Moran 2003). As more significant results are obtained, the cut-off significance level declines, potentially eliminating results that had been previously viable with a smaller k (Moran 2003). This is even more problematic in cases where results are consistently significant, but in all cases they are very close to α (e.g., 0.05) (Moran 2003). In such cases, the consistency across tests, rather than demonstrating a reliable pattern worth reporting, instead becomes a liability as many or all results are rejected by the SeqBT. If 5 of 10 tests are significant ($p < 0.05$) but only marginally, the SeqBT will fail to reject all null hypotheses (i.e., none would remain significant) (Moran 2003). However, the likelihood that 5 of 10 tests yield $p < 0.05$ by random chance is less than 1 in 10,000! Thus, some of these 5 results must be significant, yet their significance would be reversed with the BT or SeqBT corrections. The Bonferroni technique and its modifications have been used to reduce false positive results, but at the cost of rejecting viable results. The SeqBT remains rather conservative (Cabin and Mitchell 2000). The reduction of potential false positive results should not lead to excessive false negative results from BT-type corrections that may be too stringent.

The Bonferroni tests assume that these ordered statistics (p-values in ascending order) are essentially independent, yet they are not. Even using randomly generated data, the correlation coefficient ρ may be surprisingly high (e.g., for k=10, ρ is greater than 0.9, Table 9.1) (Arnold et al. 1992). We developed the correlated BT (CorBT) (Drezner and Drezner 2016) that incorporates the inherent correlation that exists among ordered data. The CorBT can be used in place of the BT, or it can be used sequentially in place of the SeqBT, as our proposed sequential correlated BT (SeqCorBT). When $\rho = 0$, the CorBT is equivalent to the BT, and the SeqCorBT is equivalent to the SeqBT. The correlation values ρ and the cut-off p-values for 3–10 significant results are given in Table 9.1. Note that for a different value of α, the critical values can be approximated by multiplying by $\frac{\alpha}{0.05}$. By adjusting for the natural correlation among p-values, much less conservative cut-offs result, but researchers can soundly correct for false positive results associated with multiple tests. For 10 significant ($p = 0.05$) results, the lowest significance level must be lower than 0.0212 to be significant using our SeqCorBT, compared to <0.0051 with the BT or SeqBT tests. We provide a user-friendly spreadsheet at http://onlinelibrary.wiley.com/doi/10.1002/bes2.1214/suppinfo which is available for readers who wish to apply this technique to their own research.

Drezner and Drezner (2016) found that for s significant results

$$\rho \approx 1 - \frac{1.329}{s} + \frac{6.396}{s\sqrt{s}} - \frac{12.646}{s^2}, \tag{9.6}$$

with significance 3.5×10^{-169} for the regression analysis.

Let $\theta(\rho, k)$ be the critical value for the kth significant result. If the smallest p-value of k significant results is less than $\theta(\rho, k)$, the kth null hypothesis can be rejected with significance α. It is shown in Drezner and Drezner (2016) that

$$\theta(\rho, k) = \frac{\alpha}{k} + \left(\alpha - \frac{\alpha}{k}\right) \rho^{\lambda(\rho,k)}, \tag{9.7}$$

where $\lambda(\rho, k)$ is

$$\lambda(\rho, k) \approx 3.928 + \frac{1}{1-\rho}\left(1.101 - \frac{3.811}{k} + \frac{4.765}{k^2}\right) - (1-\rho)^2\left(3.009 + \frac{1.783}{k}\right). \tag{9.8}$$

9.5 Summary

We proposed new techniques in statistics and in genetic algorithms. We used biological principles to innovate new approaches in genetic algorithms that yielded improved solutions to optimization problems, finding improved best known results for multiple instances. These mimicked patterns in the natural world, including female choice of mates, as well as alpha male social structures. We also highlight inconsistencies between biological processes and their genetic algorithm counterparts.

Two other innovations in methodology and statistics include our development of the sequential correlated Bonferroni test which controls for false positive results that occur from running multiple statistical tests. It incorporates the correlation between significant p-values, thereby resulting in a less conservative filter. We also developed the statistical underpinnings of a new approach for estimating transition points (in species or any other defined population) between stages. Transition from one stage to the next is a natural part of life, yet it can be difficult to estimate, particularly in cases where only a few transitions occur in every measurement period. We confirmed the validity and applicability of this new approach demonstrating low standard errors and robust output.

Interacting with a researcher in a very different field resulted in unique problem solving opportunities. The intersection of a genetic algorithm researcher with a life scientist helped to expose inconsistencies and fuel new avenues of investigation, while the mathematical and statistical foundations offered by a mathematician helped to solidify novel approaches for scientists. Such free form and synergistic collaborations are too few in research, but offer great potential for new directions and perspectives.

References

Arnold, B. C., Balakrishnan, N., & Nagaraja, H. N. (1992). *A first course in order statistics*. New York: Wiley.

Bonferroni, C. E. (1936). *Teoria statistica delle classi e calcolo delle probabilita*. Libreria internazionale Seeber.

Brimberg, J., Hansen, P., Mladenović, N., & Taillard, E. (2000). Improvements and comparison of heuristics for solving the uncapacitated multisource Weber problem. *Operations Research, 48*, 444–460.

Cabin, R. J., & Mitchell, R. J. (2000). To Bonferroni or not to Bonferroni: When and how are the questions. *Bulletin of the Ecological Society of America, 81*, 246–248.

de Carvalho, S. A. Jr., & Rahmann, S. (2006). Microarray layout as a quadratic assignment problem. In D. Huson, O. Kohlbacher, A. Lupas, K. Nieselt, & A. Zell (Eds.), *Proceedings of the German Conference on Bioinformatics* (Vol. 83, pp. 11–20), Bonn: Gesellschaft für Informatik.

Drezner, T., & Drezner, Z. (2006). Gender specific genetic algorithms. *Information Systems and Operations Research, 44*, 117–127.

Drezner, T. D. (2003). Saguaro (C*arnegiea gigantea*, Cactaceae) age–height relationships and growth: The development of a general growth curve. *American Journal of Botany, 90*, 911–914.

Drezner, T. D. (2008). Variation in age and height of onset of reproduction in the saguaro cactus (C*arnegiea gigantea*) in the Sonoran Desert. *Plant Ecology, 194*, 223–229.

Drezner, T. D. (2013a). The paradoxical distribution of a shallow-rooted keystone species away from surface water, near the water-limited edge of its range in the Sonoran Desert: Seed-seedling conflicts. *Acta Oecologica, 47*, 81–84.

Drezner, T. D. (2013b). Variability in reproductive effort of a keystone species: Age and height of branch establishment. *Physical Geography, 34*, 136–148.

Drezner, T. D. (2014). The keystone saguaro (C*arnegiea gigantea*, Cactaceae): A review of its ecology, associations, reproduction, limits, and demographics. *Plant Ecology, 215*, 581–595.

Drezner, T. D., Drezner, Z., & Balakrishnan, N. (2015). Estimating the transition of individuals between life stages. *Environmetrics, 26*, 526–533.

Drezner, Z. (2015). The quadratic assignment problem. In G. Laporte, S. Nickel, & F. S. da Gama (Eds.), *Location science* (pp. 345–363). Chum: Springer.

Drezner, Z., & Drezner, T. D. (2016). A remedy for the overzealous Bonferroni technique for multiple statistical tests. *Bulletin of the Ecological Society of America, 97*, 91–98.
Drezner, Z., & Drezner, T. D. (2018). Biologically inspired parent selection in genetic algorithms (in review)
Drezner, Z., & Drezner, T. D. (2019). The alpha male genetic algorithm. *IMA Journal of Management Mathematics, 30*, 37–50.
Drezner, Z., & Marcoulides, G. A. (2003). A distance-based selection of parents in genetic algorithms. In M. G. C. Resende & J. P. de Sousa (Eds.), *Metaheuristics: Computer decision-making* (pp. 257–278). Boston: Kluwer Academic Publishers.
Drezner, Z., & Misevičius, A. (2013). Enhancing the performance of hybrid genetic algorithms by differential improvement. *Computers & Operations Research, 40*, 1038–1046.
Drezner, Z., Gelfand, R. J., & Drezner, T. D. (2019). Sensitivity of large scale facility location solutions. *Journal of Supply Chain and Operations Management, 17*(2).
Edmands, S. (2007). Between a rock and a hard place: Evaluating the relative risks of inbreeding and outbreeding for conservation and management. *Molecular Ecology, 16*, 463–475.
Farentinos, R. (1980). Sexual solicitation of subordinate males by female tassel-eared squirrels (*Sciurus aberti*). *Journal of Mammalogy, 61*, 337–341.
Fenster, C. B., & Galloway, L. F. (2000). Inbreeding and outbreeding depression in natural populations of C*hamaecrista fasciculata* (Fabaceae). *Conservation Biology, 14*, 1406–1412.
Freeman, S., Harrington, M., & Sharp, J. C. (2014). *Biological science* (2nd Canadian ed.). Toronto: Pearson.
Galbany, J., Tung, J., Altmann, J., & Alberts, S. C. (2015). Canine length in wild male baboons: Maturation, aging and social dominance rank. *PloS One, 10*, e0126415.
Gittman, R. K., Peterson, C. H., Currin, C. A., Joel Fodrie, F., Piehler, M. F., & Bruno, J. F. (2016). Living shorelines can enhance the nursery role of threatened estuarine habitats. *Ecological Applications, 26*, 249–263.
Glover, F., & Laguna, M. (1997). *Tabu search*. Boston: Kluwer Academic Publishers.
Haley, M. P., Deutsch, C. J., & Le Boeuf, B. J. (1994). Size, dominance and copulatory success in male northern elephant seals, *Mirounga angustirostris*. *Animal Behaviour, 48*, 1249–1260.
Hansen, P., & Mladenović, N. (2001). Variable neighborhood search: Principles and applications. *European Journal of Operational Research, 130*, 449–467.
Hoelzel, A. R., Le Boeuf, B. J., Reiter, J., & Campagna, C. (1999). Alpha-male paternity in elephant seals. *Behavioral Ecology and Sociobiology, 46*, 298–306.
Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics, 6*, 65–70.
Kent, A., Drezner, T. D., & Bello, R. (2018). Climate warming and the arrival of potentially invasive species into boreal forest and tundra in the Hudson Bay Lowlands, Canada. *Polar Biology, 41*, 2007–2022.
Kirkpatrick, S., Gelat, C. D., & Vecchi, M. P. (1983). Optimization by simulated annealing. *Science, 220*, 671–680.
Lukas, D., & Clutton-Brock, T. (2014). Costs of mating competition limit male lifetime breeding success in polygynous mammals. *Proceedings of the Royal Society of London B: Biological Sciences, 281*(1786), 20140418.
McDermott, D. R., Chips, M. J., McGuirk, M., Armagost, F., DiRienzo, N., & Pruitt, J. N. (2014). Boldness is influenced by sublethal interactions with predators and is associated with successful harem infiltration in Madagascar hissing cockroaches. *Behavioral Ecology and Sociobiology, 68*, 425–435.
McGraw, K. J., & Ardia, D. R. (2003). Carotenoids, immunocompetence, and the information content of sexual colors: An experimental test. *The American Naturalist, 162*(6), 704–712.
Mendel, G. (1866). Versuche über pflanzenhybriden. *Verhandlungen des naturforschenden Vereines in Brunn 4*3, 44.
Miller, R. G. (1981). *Simultaneous statistical inference*. New York: McGraw Hill.
Moran, M. D. (2003). Arguments for rejecting the sequential Bonferroni in ecological studies. *Oikos, 100*, 403–405.

Polak, M. (2006). Booming activity of male Bitterns *Botaurus stellaris* in relation to reproductive cycle and harem size. *Ornis Fennica, 83*, 27–33.

Reinelt, G. (1991). TSLIB a traveling salesman library. *ORSA Journal on Computing, 3*, 376–384.

Rice, W. R. (1989). Analyzing tables of statistical tests. *Evolution, 43*, 223–225.

Rosser, A. M. (1992). Resource distribution, density, and determinants of mate access in puku. *Behavioral Ecology, 3*, 13–24.

Snyder, M. R., & Stepien, C. A. (2017). Genetic patterns across an invasion's history: A test of change versus stasis for the Eurasian round goby in North America. *Molecular Ecology, 26*, 1075–1090.

Solomon-Lane, T. K., Willis, M. C., Pradhan, D. S., & Grober, M. S. (2014). Female, but not male, agonistic behaviour is associated with male reproductive success in stable bluebanded goby (*Lythrypnus dalli*) hierarchies. *Behaviour, 151*, 1367–1387.

Steenbergh, W. F., & Lowe, C. H. (1983). *Ecology of the saguaro III: Growth and demography*. Washington: National Park Service.

Whiteley, A. R., Fitzpatrick, S. W., Funk, W. C., & Tallmon, D. A. (2015). Genetic rescue to the rescue. *Trends in Ecology & Evolution, 30*(1), 42–49.

Wolter, R., Pantel, N., Stefanski, V., Möstl, E., & Krueger, K. (2014). The role of an alpha animal in changing environmental conditions. *Physiology & Behavior, 133*, 236–243.

Yeaton, R. I., Karban, R., & Wagner, H. B. (1980). Morphological growth patterns of saguaro (C*arnegiea gigantea*: Cactaceae) on flats and slopes in Organ Pipe Cactus National Monument, Arizona. *The Southwestern Naturalist, 25*, 339–349.

Chapter 10
Hub Location and Related Models

Sibel A. Alumur

10.1 Introduction

This chapter discusses hub location, round trip location, transfer point location, and collection depots location problems. We define each problem and present mathematical formulations. We aim to reflect upon Zvi Drezner's contributions to the literature in each of the problem settings. In particular, we highlight the difference between hub location and classical facility location problems. We present mathematical formulations of the single and multiple allocation p-hub median, uncapacitated and capacitated hub location, and p-hub center problems.

We also define and present mathematical models for the round trip location, collection depots location, and transfer point location problems. Moreover, for each of these problems, we discuss its relation to the hub location problem and reflect upon Zvi Drezner's contributions. We lastly present some extensions to these problems and offer research prospects.

10.2 Hub Location Problem

Hubs are facilities that serve as switching, sorting, connecting, consolidation, or break-bulk points to transport traffic between many origins and destinations. The advantages of using hubs stem from (a) lower transportation or transmission costs from consolidated flows that exploit economies of scale, especially between hubs,

S. A. Alumur (✉)
Department of Management Sciences, University of Waterloo, Waterloo, ON, Canada
e-mail: sibel.alumur@uwaterloo.ca

H. A. Eiselt, V. Marianov (eds.), *Contributions to Location Analysis*, International Series in Operations Research & Management Science 281,
https://doi.org/10.1007/978-3-030-19111-5_10

(b) reduced costs from establishing a sparser network to connect many dispersed origin–destination pairs, and (c) better service from allowing more frequent connections (Alumur et al. 2019).

Hub location problems address the location of hub facilities. A distinguishing feature of hub location problems from the classical facility location problems is that demand is specified between origin–destination pairs, rather than at points or nodes of a network. In hub location problems, each point/node may have demand with every other point/node in the network. Demand can be for movement of passengers, freight, or information. Having demand between pairs of nodes necessitates interaction between the new facilities, i.e., hubs, to be located. It is usually assumed in classical facility location problems that new facilities would not interact with each other nor would they interact with the demand points assigned to other new facilities. In other words, the classical location problems involve the location of *non-interacting* new facilities.

Hub location problems have various applications in transportation and telecommunication network design such as airline passenger and freight transportation, maritime transportation, public transportation, express shipment delivery, postal operations, trucking (truckload and less-than-truckload), and computer networks design. The common phenomenon in designing such networks is to determine where to consolidate and distribute flows or data, i.e., to determine the locations of hubs.

Hub location problems link facility location and network design decisions; hence, they are very challenging set of problems. Most of the classical hub location problems are proved to be NP-Hard. The reader may refer to Campbell et al. (2002), Alumur and Kara (2008), Campbell and O'Kelly (2012), Farahani et al. (2013), Contreras (2015), and Alumur et al. (2019) for reviews and discussions on hub location problems.

Hub location problems are introduced to the literature by O'Kelly (1986). The first mathematical formulation of the problem is a quadratic integer program presented in O'Kelly (1987). The decision variables of this model are

$$x_{ij} = \begin{cases} 1, & \text{if node } i \text{ is allocated to a hub at node } j, \\ 0, & \text{otherwise.} \end{cases}$$

The quadratic formulation can then be stated as:

$$\text{Min} \quad \sum_{i \in N} \sum_{j \in N} w_{ij} \left(\sum_{k \in N} c_{ik} x_{ik} + \sum_{m \in N} c_{mj} x_{jm} + \sum_{k \in N} \sum_{m \in N} \alpha c_{km} x_{ik} x_{jm} \right) \tag{10.1}$$

$$\text{s.t.} \quad \sum_{j \in N} x_{jj} = p, \tag{10.2}$$

$$\sum_{j \in N} x_{ij} = 1 \quad i \in N, \tag{10.3}$$

$$x_{ij} \leq x_{jj} \quad i \in N, j \in N, \tag{10.4}$$

$$x_{ij} \in \{0, 1\} \quad i \in N, j \in N, \tag{10.5}$$

where N is the set of nodes, w_{ij} is the demand to be transported from node i to node j, c_{ij} is the unit transportation cost from node i to node j, and α is the economies of scale discount factor.

The objective function (10.1) calculates total cost of transportation. Constraint (10.2) ensures that exactly p hubs are to be located. By constraints (10.3) each demand node is allocated to a single hub. Constraints (10.4) allow demand nodes to be allocated only to located hubs and constraints (10.5) define the binary decision variables. This model is referred as the (*single allocation*) *p-hub median problem* in the literature.

Initial models of hub location were very much inspired from the facility location literature. In addition to the median (minisum) version, center (minimax), and covering type hub location problems have been defined and modeled (Campbell 1994). Moreover, uncapacitated and capacitated hub location models with fixed costs have also been widely studied.

In addition to the location decisions, access network and inter-hub network design decisions are to be made in hub location problems. Access network consists of the connections that link demand points to hubs, whereas inter-hub network consists of the network connections only between hubs. Another network design decision can be for having direct connections between (non-hub) demand points.

Access network design options presented in the literature include *single*, *multiple*, and *r-allocation* models. In single allocation, each demand node is connected to a single hub, as stated in constraints (10.3). In multiple allocation, a demand node can be allocated to as many hubs as necessary, and finally, in r-allocation models each demand node can be allocated to at most r hubs.

We would like to note that for the single allocation version, assigning a demand node to its nearest hub does not necessarily provide optimal solutions to the problem unlike other uncapacitated facility location problems because allocation decision depends on origin–destination flows. O'Kelly (1987) worked on a data set based on the airline passenger interactions between 25 U.S. cities in 1970 as evaluated by the Civil Aeronautics Board (CAB). In an optimal single allocation solution using the CAB data set, for example, the city Denver is allocated to a hub at Chicago located 907 miles away, rather than the hub at Dallas-Fort Worth located only 664 miles away. On the other hand, even though allocation to the nearest hub may not be optimal, as noted in Alumur et al. (2019), the nearest hub may provide a good approximation to the optimal single allocation solution which is a property used in some heuristics.

Similar to the access network, different network topologies are possible for the design of the inter-hub network, such as a star, tree, cycle, or a mesh network. Most of the literature focuses on building *complete* hub networks in which each hub has a direct connection with another. In most of the models though, this is an implicit assumption of having triangle inequality in the distance (or cost) data and also having no costs associated with establishing the inter-hub connections.

After the introduction of the quadratic integer program by O'Kelly (1987), different linear formulations of hub locations problems have been presented in the literature. In the sequel, we present linear integer programming formulations for single and multiple allocation p-hub median, uncapacitated and capacitated hub location, and p-hub center problems. In addition to the previously defined decision variables, we will use the following 4-index continuous variables:

x_{ijkm} = Fraction of flow from origin i to destination j that is routed via hubs at locations k and m in that order.

The *single allocation p-hub median problem* can be modeled as (Skorin-Kapov et al. 1996):

$$\text{Min} \quad \sum_{i \in N} \sum_{j \in N} \sum_{k \in N} \sum_{m \in N} w_{ij} c_{ijkm} x_{ijkm} \tag{10.6}$$

$$\text{s.t.} \quad (10.2) - (10.5),$$

$$\sum_{m \in N} x_{ijkm} = x_{ik} \quad i, j, k \in N, \tag{10.7}$$

$$\sum_{k \in N} x_{ijkm} = x_{jm} \quad i, j, m \in N, \tag{10.8}$$

$$x_{ijkm} \geq 0 \quad i, j, k, m \in N, \tag{10.9}$$

where $c_{ijkm} = c_{ik} + \alpha c_{km} + c_{mj}$.

Using the same set of decision variables, a linear integer programming formulation of the *multiple allocation p-hub median problem* can be stated as (Skorin-Kapov et al. 1996):

$$\text{Min} \quad \sum_{i \in N} \sum_{j \in N} \sum_{k \in N} \sum_{m \in N} w_{ij} c_{ijkm} x_{ijkm}$$

$$\text{s.t.} \quad (10.2), (10.5), (10.9),$$

$$\sum_{k \in N} \sum_{m \in N} x_{ijkm} = 1 \quad i, j \in N, \tag{10.10}$$

$$\sum_{m \in N} x_{ijkm} \leq x_{kk} \quad i, j, k \in N, \tag{10.11}$$

$$\sum_{k \in N} x_{ijkm} \leq x_{mm} \quad i, j, m \in N. \tag{10.12}$$

In the uncapacitated/capacitated hub location problem, total number of hubs to be located is no longer restricted to p, instead, there is a fixed cost term in the objective function associated with locating a hub. Let f_k denote the cost of opening a hub at node k, the objective function of the uncapacitated/capacitated hub location problem can then be written as:

$$\text{Min} \quad \sum_{i \in N} \sum_{j \in N} \sum_{k \in N} \sum_{m \in N} w_{ij} c_{ijkm} x_{ijkm} + \sum_{k \in N} f_k x_{kk}. \tag{10.13}$$

To model the *uncapacitated hub location problem*, constraint (10.2) is to be removed from each of the single and multiple allocation p-hub median models and objective function (10.13) is to be employed.

For the *capacitated hub location problems*, additional capacity constraints are introduced into the models. If Γ_k denote the capacity of hub k, the capacity constraint on the total flow passing through a hub can be stated as:

$$\sum_{i \in N} \sum_{j \in N} \sum_{m \in N} w_{ij} (x_{ijkm} + x_{ijmk} - x_{ijkk}) \leq \Gamma_k x_{kk} \quad k \in N, \tag{10.14}$$

where the third them in the parenthesis is included to avoid the double-count (Campbell 1994).

Different types of capacity constraints are modeled in the literature such as having a capacity constraint only on the total inflow/outflow to/from a hub. It is also possible to have capacity constraints for limiting the flow on the access or inter-hub network connections. Modeling of capacity constraints depends very much on intended applications of the problem.

The *p-hub center problems*, on the other hand, are minimax type hub location problems where the maximum cost is to be minimized. Different types of p-hub center problems are defined in the literature (Campbell 1994). The most commonly employed p-hub center objective is

$$\min \max_{i,j,k,m \in N} \{c_{ijkm} x_{ijkm}\}. \tag{10.15}$$

A number of different formulations, for example, with less number of decision variables or tighter constraints, are presented in the literature for all versions of hub location problems. For a comparison of formulations in terms of computational efficiency and a list of effective solution methodologies, we refer the reader to the abovementioned surveys and reviews on hub location.

Most of the literature has focused on network hub location problems where the locations of the hubs are restricted to the nodes of the network, i.e., the set N. It is possible to model and solve continuous hub location problems on the plane as well. Initial work on planar hub location problems are by O'Kelly (1986, 1992), Aykin (1988, 1995), Campbell (1990), and O'Kelly and Miller (1991).

Zvi Drezner's contributions to hub location literature focused on a special case of the multiple allocation p-hub median problem. Motivated from the operation of a domestic airline network of a relatively small country such as Japan, Sasaki et al. (1999) defined and studied the *1-stop multiple allocation p-hub median problem*. In this problem, each origin–destination flow can make at most one stop at a hub facility, compared to having at most two hub-stops in classical hub location problems. To have at most one hub-stop, each demand point needs to be connected to all hubs. This is a special case of the multiple allocation p-hub median problem in the presence of triangle inequality and when there is no cost reduction due to economies of scale ($\alpha = 1$). Optimal solution of this problem provides a lower bound on the objective function value of the multiple allocation p-hub median problem.

Sasaki et al. (1997) referred the same problem as the *relay point location problem*. Sasaki et al. (1999) formulated the problem as a p-median problem and proposed a branch-and-bound algorithm and a greedy-type heuristic. Suzuki and Drezner (1997), on the other hand, studied the continuous version of this problem when the demand is evenly spread in a given area to provide insights on the solution patterns.

An interesting approach to the 1-stop multiple allocation p-hub median problem was later developed by Drezner and Drezner (2001). The authors modeled the case where passengers do not necessarily select the hub providing the shortest distance. The portion of the passengers who select a particular hub is taken to be inversely proportional to a power of the total distance to the destination through that hub, which is referred as the gravity rule (Huff 1964, 1966).

We would like to conclude this section by presenting a critical assessment of the basic assumptions made in many hub location models. As noted before, exploiting economies of scale is one of the major advantages of using hub networks. In preliminary models of hub location, economies of scale is incorporated by using a constant discount factor, usually referred as α, that is independent of flows. However, as pointed out initially by O'Kelly and Bryan (1998), and later on by Kimms (2006) among many others, that the reduction in costs due to economies of scale should depend on the amount of flow carried and, hence, this simple cost model is not valid in practice. Alternatives for better modeling economies of scale in hub location models include modeling cost of flows on the links by using a non-linear function dependent on flow (e.g., O'Kelly and Bryan 1998), allowing for building incomplete hub networks (e.g., Alumur et al. 2009), and operating different types of vehicles by incorporating fixed and variable costs of vehicles in the models (e.g., Serper and Alumur 2016; Masaeli et al. 2018). The reader is referred to Alumur et al. (2019) for an extensive discussion on this topic.

10.3 Round Trip Location Problem

The *round trip location problem* is to find the location of a facility that minimizes the maximum weighted distance between the facility and pairs of demand points. This is a minimax type single facility planar location problem introduced by Chan and Hearn (1977).

There are n pairs of demand points with fixed locations denoted by $P_i = (a_i, b_i)$ and $Q_i = (c_i, d_i)$, $i = 1, \ldots, n$. $X = (x, y)$ denotes the coordinates of the new facility to be located. The ith trip, which is weighted by a positive constant w_i, starts at the facility X, visits two demand points P_i and Q_i sequentially, and returns to the facility. Alternatively, due to symmetrical distances, the ith trip can be deemed between P_i and Q_i that needs to go through facility X. The "cost" of each round trip distance is therefore

$$F_i(X) = w_i[d(P_i, X) + d(X, Q_i) + g_i], \tag{10.16}$$

where $d(X, Y)$ is the distance between locations X and Y, and $g_i = d(Q_i, P_i)$ or any constant.

The cost function with the general (l_p) distance norm can be written as:

$$F_i(X) = w_i[(|x - a_i|^p + |y - b_i|^p)^{1/p} + (|x - c_i|^p + |y - d_i|^p)^{1/p} + g_i], \tag{10.17}$$

where $p \geq 1$.

The problem is to minimize the maximum round trip distance, that is,

$$\min_X \max_{1 \leq i \leq n} \{F_i(X)\}. \tag{10.18}$$

It is also possible to model the minisum version of this problem minimizing the total distance traveled. However, we would like to note that the problem with the minisum objective would be the same as the standard Weber problem.

The round trip location problem is equivalent to the *single facility minimax hub location problem* on the plane (*1-hub center problem* on a network). In this setting, the starting point, i.e., the origin, of delivery i is P_i and the receiving point, the destination, is Q_i. Each delivery needs to make a stop at the hub facility X which is to be located. The weight w_i may represent the demand between ith pair of demand points or the average time to travel one unit distance.

Chan and Hearn (1977) modeled and solved this problem only for the rectilinear distance case. Drezner and Wesolowsky (1982) proposed a solution method applicable for the Euclidean as well as general (l_p) distances. This method involves the numerical solution of differential equations by standard means. Faster algorithms with improved complexity were later on proposed by Drezner (1982) for general distances, and Drezner (1985) for the rectilinear distance case.

O'Kelly and Miller (1991) studied the single facility minimax hub location problem and visualized the "ellipse" enclosing the feasible locations for a hub to serve pairs of demand points. The authors compared different solution techniques and concluded through their computational experiments that the algorithm of Drezner and Wesolowsky (1982) is extremely efficient. More recently, O'Kelly (2009) extended this problem to three dimensional space, where the nodes are permitted to be on different layers, by formulating and solving the 3-D single facility minimax hub location problem.

10.4 Collection Depots Location Problem

The *collection depots location problem* is to find the location of a facility that minimizes total distance traveled to provide service to a given set of demand points using known locations of collection depots. (Although the name of the problem may imply that the facilities to be located are the “collection depots,” note that the locations of the collection depots are fixed.) In this problem, each service consists of a trip from the facility to the customer to collect materials, then dropping the materials at one of the available collection depots, and returning to the facility to wait for the next call. Alternatively, the first leg of the trip can be from the facility to the collection depot, the second from the depot to the customer, and the last from the customer to the facility to be located. If the distances are symmetrical, total distance traveled would be the same.

The collection depots location problem is introduced by Drezner and Wesolowsky (2001), where they also provided a number of potential applications of the problem. Possible applications of the collection depots location problem include a septic tank cleaning service, garbage collection or tree pruning service, a delivery service that requires a stop at some available warehouse, a quality inspection of delivered items on the way to the customers. If fueling is required either on the way to the customer or on the way back, gas stations can be modeled as collection depots as well.

The difference of this problem from the round trip location problem is that the stops are not predetermined. Travel from/to the customer to/from the depot is dependent on the location of the facility which is to be determined. More precisely, each demand point is to be allocated to a collection depot and this allocation depends on where the facility is located. Hence, each customer is not necessarily allocated to its closest depot. In this sense, the round trip location problem is a special case of the collection depots location problem where the allocations of demand points to the collection depots are fixed.

A related problem, not involving return trips, is the traveling salesman location problem (Berman et al. 1995). In this problem, a new facility needs to be built to serve customers who are visited several at a time. The collection depots location problem is different than the traveling salesman location problem because of the need to allocate the nearest depot to one leg of each return trip.

Consider n demand points denoted by P_i, $i = 1, \ldots, n$, and m collection depots denoted by C_j, $j = 1, \ldots m$, with fixed locations. Let w_i be the weight of demand point P_i and X denote the location of the new facility. The function $d(X, Y)$ denotes the distance between locations X and Y. The collection depots location problem can then be stated as:

$$\min_X \sum_{i=1}^{n} w_i \left\{ d(X, P_i) + \min_{1 \le k \le m} \{ d(P_i, C_k) + d(C_k, X) \} \right\}. \tag{10.19}$$

Drezner and Wesolowsky (2001) extensively analyzed this problem and proved a number of solution properties. Berman et al. (2002) investigated the collection depots location problem on a network. The authors additionally considered the minimax version of the problem. Minimax objective can be used for the same potential applications when the decision maker wishes to minimize the largest service time rather than the total. The minimax version of the problem is later on referred as the *round trip center problem* by Tamir and Halman (2005). The minimax collection depots location problem can be stated as:

$$\min_{X} \max_{1 \leq i \leq n} \left\{ w_i \left\{ d(X, P_i) + \min_{1 \leq k \leq m} \{d(P_i, C_k) + d(C_k, X)\} \right\} \right\}. \tag{10.20}$$

Berman and Huang (2004) and, more recently, Drezner et al. (2018) studied the *multifacility collection depots location problem* on a network and in the plane, respectively. Multiple facilities are to be located in this problem setting. The multifacility problem in the plane can be modeled as:

$$\min_{X_j, j=1,\ldots,p} \sum_{i=1}^{n} w_i \min_{1 \leq j \leq p} \left\{ d(X_j, P_i) + \min_{1 \leq k \leq m} \left\{ d(P_i, C_k) + d(C_k, X_j) \right\} \right\}, \tag{10.21}$$

where p facilities are to be located and X_j denotes the location of the jth facility.

The multifacility collection depots problem can be viewed as an extension of the p-median problem (multi-source Weber problem). When there are no collection depots ($m = 0$) or when all n demand points also serve as collection depots ($m = n$), the objective function of the multifacility collection depots problem is two times that of the *p-median problem*. Moreover, if the number of collection depots is the same as the number of facilities ($m = p$), then the optimal solution is to locate depots and facilities at the same locations and the problem again converts to the p-median problem (Drezner et al. 2018).

Berman and Huang (2004) presented an integer programming formulation of this problem on a network. The decision variables of the model are

$$x_{ijk} = \begin{cases} 1, & \text{if demand from node } i \text{ is assigned to the facility at node } j \\ & \text{and depot } k \text{ is selected,} \\ 0, & \text{otherwise.} \end{cases}$$

$$y_j = \begin{cases} 1, & \text{if there is a facility located at node } j, \\ 0, & \text{otherwise.} \end{cases}$$

The multifacility collection depots location problem on a network can then be formulated as follows:

$$\text{Min} \quad \sum_{i=1}^{n}\sum_{j=1}^{\bar{n}}\sum_{k=1}^{m} w_i[d(j,i)+d(i,k)+d(k,j)]x_{ijk} \tag{10.22}$$

$$\text{s.t.} \quad \sum_{j=1}^{\bar{n}} y_j = p, \tag{10.23}$$

$$\sum_{j=1}^{\bar{n}}\sum_{k=1}^{m} x_{ijk} = 1 \quad i=1,\ldots,n, \tag{10.24}$$

$$x_{ijk} \le y_j \quad i=1,\ldots n;\ j=1,\ldots,\bar{n};\ k=1,\ldots,m, \tag{10.25}$$

$$x_{ijk}, y_j \in \{0,1\}, \quad i=1,\ldots n;\ j=1,\ldots,\bar{n};\ k=1,\ldots,m. \tag{10.26}$$

The objective function (10.22) minimizes the total weighted travel distance. By constraint (10.23) exactly p facilities are to be located. Constraints (10.24) ensure that each demand node is assigned to exactly one collection depot-facility combination. Constraints (10.25) state that a demand node can only be assigned to an open facility. Lastly, constraints (10.26) establish binary restrictions on the decision variables.

The difference of the collection depots location problem from the hub location problem is that there is no flow to be transported between the demand points, collection depots, or facilities. Hence, we cannot talk about an origin–destination flow in this problem setting. Thus, the collection depots location problem is not a special case of the hub location problem.

10.5 Transfer Point Location Problem

Transfer point location problems are hierarchical type facility location problems in which demand nodes are allocated to transfer points and transfer points are connected to a central facility. An example of this problem setting is from emergency services: The patients are first transferred to a transfer facility, such as a helicopter pad, at normal speed and from there they are transferred to the central facility, such as a hospital, at increased speed. Different types of the problem are defined in the literature (Berman et al. 2007, 2008, 2005):

The Transfer Point Location Problem (TPLP): In this problem setting, the location of the central facility is given and only one transfer point is to be located to serve a given set of demand points.

The Multiple Transfer Points Location Problem (MTPLP): In this setting, the location of the central facility is again given, and multiple transfer points are to be located. Each demand point is to be served by a single transfer point and direct connection of the demand points to the central facility is allowed.

The Facility and Transfer Points Location Problem (FTPLP): This is the most general setting where the MTPLP model is extended to also find the optimal location for the central facility.

It is possible to model both minisum and minimax versions of all the above problems. Berman et al. (2007) analyzed properties of the solutions to the TPLP model. They studied both planar and network variants, as well as the minisum and minimax objectives. Berman et al. (2008) investigate the MLTP model both in the plane and on a network. Berman et al. (2005), on the other hand, propose heuristic solution procedures for solving the FTPLP. Below, we present the FTPLP model under both objective functions.

FTPLP involves the location of p transfer facilities and a single central facility. There are n demand points to serve with known locations $P_i, i = 1, \ldots, n$. Let w_i be the weight associated with demand point i. Let X denote the location of the central facility and H_j the locations of the transfer points with $j = 1, \ldots, p$. The function $d(X, Y)$ denotes the distance between locations X and Y. The cost (or time) per unit distance of traveling from the transfer point to the facility is to be multiplied by a reduction factor of $\alpha < 1$. The *minisum FTPLP* is then formulated as:

$$\min_{X;H_j, j=1,\ldots,p} \left\{ \sum_{i=1}^{n} w_i \min \left\{ d(P_i, X), [d(P_i, H_j) + \alpha d(H_j, X)], j = 1, \ldots, p \right\} \right\}. \tag{10.27}$$

The *minimax FTPLP* model is

$$\min_{X;H_j, j=1,\ldots,p} \left\{ \max_{i=1,\ldots,n} \left\{ w_i \min \left\{ d(P_i, X), [d(P_i, H_j) + \alpha d(H_j, X)], j = 1, \ldots, p \right\} \right\} \right\}. \tag{10.28}$$

Assuming that there are n nodes in the network and a facility or a transfer point can be located at any node of the network, Berman et al. (2005) formulated the minisum FTPLP on a network with the use of the following decision variables:

$$x_{ijk} = \begin{cases} 1, & \text{if demand from node } i \text{ is assigned to a facility at node } k \text{ via a} \\ & \text{transfer point from node } j, \\ 0, & \text{otherwise.} \end{cases}$$

$$x_{ik} = \begin{cases} 1, & \text{if demand from node } i \text{ is assigned directly to the facility at node } k, \\ 0, & \text{otherwise.} \end{cases}$$

$$z_k = \begin{cases} 1, & \text{if the facility is located at node } k, \\ 0, & \text{otherwise.} \end{cases}$$

$$y_j = \begin{cases} 1, & \text{if the transfer point is located at node } j, \\ 0, & \text{otherwise.} \end{cases}$$

Let $C_{ijk} = d(i, j) + \alpha d(j, k)$, and $C_{ik} = d(i, k)$. The minisum FTPLP on a network can then be formulated as follows:

$$\text{Min} \quad \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} C_{ijk} x_{ijk} + \sum_{i=1}^{n} \sum_{k=1}^{n} C_{ik} x_{ik} \tag{10.29}$$

$$\text{s.t.} \quad x_{ijk} \le y_j \quad i, j, k = 1, \ldots, n, \tag{10.30}$$

$$x_{ijk} \le z_k \quad i, j, k = 1, \ldots, n, \tag{10.31}$$

$$x_{ik} \le z_k \quad i, k = 1, \ldots, n, \tag{10.32}$$

$$\sum_{j=1}^{n} \sum_{k=1}^{n} x_{ijk} + \sum_{k=1}^{n} x_{ik} = 1 \quad i = 1, \ldots, n, \tag{10.33}$$

$$\sum_{j=1}^{n} y_j = p, \tag{10.34}$$

$$\sum_{k=1}^{n} z_k = 1, \tag{10.35}$$

$$x_{ijk}, x_{ik}, y_j, z_k \in \{0, 1\}, \quad i, j, k = 1, \ldots, n. \tag{10.36}$$

Objective function (10.29) minimizes total distance traveled. Constraints (10.30) and (10.31) ensure together that a node can be assigned to the central facility at node k via a transfer point at node j only if there is a facility located at node k and a transfer point located at node j. By constraints (10.32) direct connection of node i to central facility k is not possible unless there is a facility located at node k. Constraints (10.33) ensure that a node is either assigned to a transfer point or has a direct connection with the central facility. Constraint (10.34) establishes p transfer points, and constraint (10.35) limits the number of central facilities to 1. Lastly, constraints (10.36) define the domain of the binary decision variables.

The FTPLP can be considered to be a special case of the *hierarchical hub location problem* where the demand originates from many origin points and all destined to a single destination point, i.e., the central facility, which happens to be a hub.

The hierarchical hub location problem was introduced by Yaman (2009) motivated from the design of cargo delivery and telecommunication networks. The aim is to serve the demand between every pair of nodes in the network with minimum total cost. More specifically, given a set of demand nodes, a set of possible locations for hubs, a set of possible locations for central hubs, the number of hubs and central hubs to be opened, the traffic demand and the routing cost between pairs of nodes, and the discount factors due to economies of scale, the hierarchical hub location problem determines the locations of hubs and central hubs, the assignment of demand nodes to hubs, and the assignment of hubs to central hubs to minimize the total routing cost in the network.

In a classical hub network, a path connecting an origin–destination pair visits at most two hubs. In a hierarchical hub network, it is possible to have paths visiting four hubs between origin–destination pairs. A version of the hierarchical hub location problem considering multiple transportation modes and time-definite deliveries has been studied by Alumur et al. (2012).

Lastly, we would like to note that although FTPLP reduces to the hierarchical hub location problem, FTPLP should not be regarded as a hub location problem. A distinguishing feature of hub location problems is to have flow between hub facilities, whereas in transfer point location problems no flow is to be transported between the transfer points as the demand is originated from many points but destined to only one.

10.6 Extensions and Research Prospects

Travel distance or travel time is one of the most important input parameter used in all location models. Many location problems implicitly assume symmetric travel distances. However, as pointed out by Drezner and Wesolowsky (1989), distance from point A to point B can be different than that of the distance from point B to A. With this observation, Drezner and Wesolowsky (1989) introduced and solved asymmetric Weber and minimax location problems with rectilinear and Euclidean distances. A recent note by Drezner and Drezner (2019) addresses asymmetric Weber and round trip Weber location problems when the distance or time to get to a destination is affected by winds.

In many location models, distance and travel time are considered to be substitutable. This is a consequence of having the implicit assumption of a constant travel speed. However, travel speed is not constant in practice because there is acceleration and deceleration portions involved in each trip. A more accurate modeling of travel time, when it is not necessarily proportional to the distance, is proposed by Drezner et al. (2009). They introduced and solved the Weber problem incorporating Euclidean acceleration–deceleration distances.

If there is a discrete set of available facility locations, then it is possible to take care of asymmetric travel distances or realistic travel times, taking acceleration–deceleration into account, for example, by computing these parameters a priori for each demand point–candidate location pair. Thus, for network location problems, actual distances and travel times between each pair of nodes can be provided as an input parameter. For planar location problems, on the other hand, this is not the case as there are infinitely many possibilities for facility locations in the plane. Hence, these issues need to be considered explicitly in planar location models as done in Drezner and Wesolowsky (1989), Drezner et al. (2009), and Drezner and Drezner (2019).

Most hub location models focus on network models containing a discrete set of available hub locations. However, as pointed out in Alumur et al. (2019), planar hub location models may provide deeper geometric insights into the problem. Multifacility planar hub location problems have not received much attention compared with single facility models. Solution techniques developed for other multifacility planar location models may be useful in solving their hub location counterparts. For example, algorithms provided for the round trip location problem (Drezner and Wesolowsky 1982; Drezner 1982, 1985) may be useful in solving the multifacility planar minimax hub center or hub covering problems.

As discussed before, one major drawback of many hub location models is having simplified assumptions for modeling economies of scale, for example, using a constant discount factor that is independent of flow (i.e., α). Given that economies of scale is a *raison d'etre* for using hub networks, hub location research should model economies of scale better. However, accurate models of economies of scale that incorporate flow-dependent discounts may lead to non-linear formulations, and consequently, to longer solution times and intractable models for large instances (Alumur et al. 2019).

Another possible extension of hub location models is to incorporate the time dimension such as by integrating frequency of service and scheduling considerations into the models. How to incorporate time depends highly on intended applications of the models, such as for passenger versus freight transportation networks. Modeling the time dimension brings out synchronization issues which are generally very challenging to model and solve as in service network design problems.

Accurate estimation of demand is very important in solving hub location problems. Handling continuous demand (e.g., Campbell 1990; Suzuki and Drezner 1997) as well as modeling endogenous effects of hub locations on the demand (e.g., O'Kelly 1986) remains to be challenging, while such extensions will certainly make hub location models more realistic.

Lastly, multifacility versions of the round trip location, collection depots location, and facility and transfer points location problems offer further research prospects.

References

Alumur, S. A., Campbell, J. F., Contreras, I., Kara, B. Y., Marianov, V., & O'Kelly, M. E. (2019, under review). On modeling hub location problems. Technical report.

Alumur, S. A., & Kara, B. Y. (2008). Network hub location problems: The state of the art. *European Journal of Operational Research, 190*, 1–21.

Alumur, S. A., Kara, B. Y., & Karasan, O. E. (2009). The design of single allocation incomplete hub networks. *Transportation Research Part B, 43*, 936–951.

Alumur, S. A., Yaman, H., & Kara, B. Y. (2002). Hierarchical multimodal hub location problem with time-definite deliveries. *Transportation Research Part E, 48*, 1107–1120.

Aykin, T. (1988). On the location of hub facilities. *Transportation Science, 22*, 155–157.

Aykin, T. (1995). The hub location and routing problem. *European Journal of Operational Research, 83*, 200–219.
Berman, O., Drezner, Z., & Wesolowsky, G. O. (2002). The collection depots location problem on networks. *Naval Research Logistics, 49*, 15–24.
Berman, O., Drezner, Z., & Wesolowsky, G. O. (2005). The facility and transfer points location problem. *International Transactions in Operational Research, 12*, 387–402.
Berman, O., Drezner, Z., & Wesolowsky, G. O. (2007). The transfer point location problem. *European Journal of Operational Research, 179*, 978–989.
Berman, O., Drezner, Z., & Wesolowsky, G. O. (2008). The multiple location of transfer points. *Journal of the Operational Research Society, 59*, 805–811.
Berman, O., & Huang, R. (2004). Minisum collection depots location problem with multiple facilities on a network. *Journal of the Operational Research Society, 55*, 769–779.
Berman, O., Jailet, P., & Simchi-Levi, D. (1995). Location-routing problems with uncertainty. In Z. Drezner (Eds.), *Facility location: A survey of applications and methods* (pp. 427–452). New York: Springer.
Campbell, J. F. (1990). Locating transportation terminals to serve an expanding demand. *Transportation Research B, 24*, 173–192.
Campbell, J. F. (1994). Integer programming formulations of discrete hub location problems. *European Journal of Operational Research, 72*, 387–405.
Campbell, J. F., Ernst, A. T., & Krishnamoorthy, M. (2002). Hub location problems. In Z. Drezner & H. Hamacher (Eds.), *Facility location: Applications and theory* (pp. 373–407). Berlin: Springer.
Campbell, J. F., & O'Kelly, M. E. (2012). Twenty-five years of hub location research. *Transportation Science, 46*, 153–169.
Chan, A. W., & Hearn, D. W. (1977). A rectilinear distance round-trip location problem. *Transportation Science,11*, 107–123.
Contreras, I. (2015). Hub location problems. In G. Laporte, S. Nickel, & F. S. da Gama (Eds.), *Location science* (pp. 311–344). Heidelberg: Springer.
Drezner, T., & Drezner, Z. (2001). A note on applying the gravity rule to the airline hub problem. *Journal of Regional Science, 41*, 67–73.
Drezner, T., & Drezner, Z. (2019, in review). Asymmetric distance location model. INFOR: Information Systems and Operational Research.
Drezner, T., Drezner, Z., & Kalczynski, P. (2018). The planar multifacility collection depots location problem. *Computers & Operations Research, 102*, 121–129. https://doi.org/10.1016/j.cor.2018.09.010
Drezner, Z. (1982). Fast algorithms for the round trip location problem. *IIE Transactions, 14*, 243–248.
Drezner, Z. (1985). O(NlogN) algorithm for the rectilinear round trip location problem. *Transportation Science, 19*, 91–93.
Drezner, Z., Drezner, T., & Wesolowsky, G. O. (2009). Location with acceleration-deceleration distance. *European Journal of Operational Research, 198*, 157–164.
Drezner, Z., & Wesolowsky, G. O. (1982). A trajectory approach to the round trip location problem. *Transportation Science, 16*, 56–66.
Drezner, Z., & Wesolowsky, G. O. (1989). The asymmetric distance location problem. *Transportation Science, 23*, 201–207.
Drezner, Z., & Wesolowsky, G. O. (2001). On the collection depots location problem. *European Journal of Operational Research, 130*, 510–518.
Farahani, R. Z., Hekmatfar, M., Arabani, A. B., & Nikbakhsh, E. (2013). Hub location problems: A review of models, classification, solution techniques, and applications. *Computers & Industrial Engineering, 64*, 1096–1109.
Huff, D. L. (1964). Defining and estimating a trade area. *Journal of Marketing, 28*, 34–38.
Huff, D. L. (1966). A programmed solution for approximating an optimum retail location. *Land Economics, 42*, 293–303.

Kimms, A. (2006). Economies of scale in hub & spoke network design models: We have it all wrong. In *Perspectives on operations research* (pp. 293–317). Wiesbaden: Deutscher Universities.

Masaeli, M., Alumur, S. A., & Bookbinder, J. H. (2018). Shipment scheduling in hub location problems. *Transportation Research Part B, 115*, 126–142.

O'Kelly, M. E. (1986). The location of interacting hub facilities. *Transportation Science, 20*, 92–105.

O'Kelly, M. E. (1987). A quadratic integer program for the location of interacting hub facilities. *European Journal of Operational Research, 32*, 393–404.

O'Kelly, M. E. (1992). A clustering approach to the planar hub location problem. *Annals of Operations Research, 40*, 339–353.

O'Kelly, M. E. (2009). Rectilinear minimax hub location problems. *Journal of Geographical Systems, 11*, 227–241.

O'Kelly, M. E., & Bryan, D. L. (1998). Hub location with flow economies of scale. *Transportation Research Part B, 32*, 605–616.

O'Kelly, M. E., & Miller, H. J. (1991). Solution strategies for the single facility minimax hub location problem. *Papers in Regional Science, 70*, 367–380.

Sasaki, M., Suzuki, A., & Drezner, Z. (1997). On the selection of relay points in a logistics system. *Asia-Pacific Journal of Operational Research, 14*, 39–54.

Sasaki, M., Suzuki, A., & Drezner, Z. (1999). On the selection of hub airports for an airline hub-and-spoke system. *Computers & Operations Research, 26*, 1411–1422.

Serper E. Z., & Alumur, S. A. (2016). The design of capacitated intermodal hub networks with different vehicle types. *Transportation Research Part B, 86*, 51–65.

Skorin-Kapov, D., Skorin-Kapov, J., & O'Kelly, M. (1996). Tight linear programming relaxations of uncapacitated p-hub median problems. *European Journal of Operational Research, 94*, 582–593.

Suzuki, A., & Drezner, Z. (1997). On the airline hub problem: The continuous model. *The Journal of the Operations Research Society of Japan, 40*, 62–74.

Tamir, A., & Halman, N. (2005). One-way and round-trip center location problems. *Discrete Optimization, 2*, 168–184.

Yaman, H. (2009). The hierarchical hub median problem with single assignment. *Transportation Research Part B, 43*, 643–658.

Chapter 11
Gravity Models in Competitive Facility Location

Tammy Drezner

11.1 Introduction

I dedicate this chapter to my husband, Zvi Drezner, who has been the inspiration for my work. As a Ph.D. student in Urban Planning, I was looking for a topic for my dissertation. One evening, I asked Zvi to tell me more about his area of research. He told me about facility location, a topic about which I knew little. The conversation meandered through the different branches of location theory until Zvi mentioned one of his papers about retail location and the "fact" that people patronize the facility closest to them. I thought that this was not the case, people patronize a more attractive facility at a greater distance, and a discussion about facility attractiveness ensued. My advisor was intrigued by this new approach to modeling competitive facility location. My dissertation was born. Zvi has been an inspiration ever since. We have co-authored many papers about competitive facility location and other location topics. It has been a pleasure and honor working with him.

The underlying theme running through all competitive location models is the existence of an interrelationship between four variables: buying power (demand), distance, facility attractiveness, and market share, with the first three variables being independent variables and market share is the dependent variable. It is implicitly assumed that revenue and profit are an increasing function of market share. Therefore, maximizing market share is equivalent to maximizing revenue or profit. Buying power, or effective buying income, is known (for example, in *Sales and Marketing Management* magazine). Distance from demand points to facilities can be measured. Attractiveness is usually estimated by surveys.

T. Drezner (✉)
Steven G. Mihaylo College of Business and Economics, California State University-Fullerton, Fullerton, CA, USA
e-mail: tdrezner@fullerton.edu

H. A. Eiselt, V. Marianov (eds.), *Contributions to Location Analysis*, International Series in Operations Research & Management Science 281,
https://doi.org/10.1007/978-3-030-19111-5_11

Reilly (1931) suggested the gravity model. The model involves the hypothesis that two cities attract retail trade from an intermediate town approximately in direct proportion to the populations of the two cities and in inverse proportion to the square of the distances from the two cities to the intermediate town (Buffa 1976). This rule imitates the physics law of gravity. The city size represents the body mass. The gravitational force declines by the square of the distance.

Huff (1964, 1966) proposed the use of the gravity model for estimating market share suggesting that the probability that customers patronize a retail facility is proportional to its attractiveness multiplied by a distance decay function $f(d)$. The distance decay is the rate at which facility attraction declines as a function of the distance. Huff suggested the decay function $\frac{1}{d^{\lambda}}$, where λ is a parameter.

Another competitive facility model was suggested by Hotelling (1929). He proposed that competitors compete by charging different mill prices and customers select the facility that provides the lowest total price of mill prices plus the cost of travel. This approach led to many papers (for example, Hakimi 1981; Drezner 1982; Hakimi 1983, 1986, 1990; ReVelle 1986; Ghosh and Rushton 1987; Serra and ReVelle 1995) that apply the proximity rule, i.e., customers patronize the closest facility. The proximity rule implies that all facilities charge the same price and thus are equally attractive.

Drezner (1994a) assumed that customers are willing to travel an extra distance to a more attractive facility. A utility function is defined and the customer selects the facility with the highest utility. This is a generalization of the proximity rule where only distance is considered in the utility function. These two approaches were summarized in Drezner (1995). Drezner and Drezner (1996) generalized the utility model to the random utility model. The components of the utility function are assumed to have a random distribution. A similar approach was proposed by Leonardi and Tadei (1984).

Drezner et al. (2011, 2012, 2015, 2016) proposed a new non-stochastic approach for estimating market share captured by competing facilities. Each competing facility has a "sphere of influence" determined by its attractiveness level. More attractive facilities have a larger radius of the sphere of influence. The buying power spent by a customer in the sphere of influence of several facilities is equally divided among the competing facilities. The buying power of a customer in the sphere of influence of no facility is lost. Lost demand is discussed in Sect. 11.3.3. Note that ReVelle (1986) coined the term "sphere of influence." However, ReVelle's model is based on the proximity rule, not on a "radius of influence." In his model all demand is satisfied and all the buying power is distributed among the facilities. The idea of a radius of influence is at the core of *central place theory* (Lösch 1954; Christaller 1966). According to central place theory there is a maximum *range of the good*, depending on retail category, that customers are willing to travel to obtain the good. This concept is further considered by Zeller et al. (1980), Black (1984), and Ghosh and Craig (1986, 1991) in the marketing literature and in Clark (1968) and Clark and Rushton (1970) in the geographic literature. The reader is referred to Ghosh and Craig (1991) for a comprehensive discussion of central place theory.

There are many applications to competitive location models. The location of shopping malls, grocery stores, general merchandise stores, specialty stores (clothing, children apparel, shoes, men's suits, jewelry, toys, appliances, computers and computer supplies, books, gifts, hardware stores, office supplies, furniture stores, food specialties, pharmacies, etc.), restaurants (fast food, coffee shops, ethnic food, steak houses, ice cream parlors, sandwich places, etc.), gas stations, bank branches, movie theaters, car dealerships, car repair shops, and many others. For recent reviews of competitive facilities' location problems see Berman et al. (2009), Drezner (2014), Eiselt et al. (2015), Drezner and Eiselt (2002), and Eiselt (2011).

In Sect. 11.2 the gravity model formulation is presented. The three components of the gravity model are the distance, the distance decay function, and the attractiveness. These components and variants thereof are discussed in detail.

In Sect. 11.3 many aspects of the gravity models are detailed.

1. The leader–follower model, also termed the Stackelberg equilibrium, is anchored in game theory. The leader anticipates a follower's reaction to his location decision.
2. Allocating a given budget to facilities in order to maximize the increase in total market share captured.
3. Most competitive facility location models assume that all available buying power is partitioned among the competing facilities. Models that consider lost demand are discussed.
4. Various scenarios in models that incorporate changing market conditions are discussed. The minimax regret objective is minimized.
5. The threshold objective is minimizing the probability that a minimum threshold market share is not met. If a facility fails to meet the threshold, it will have to be closed.
6. Consistent and inconsistent rules. Whether a customer changes his/her mind on the way to the selected facility when passing by another facility.
7. New facilities belonging to the same franchise cannibalize existing franchise facilities.
8. The order of sequentially locating two facilities which belong to the same chain is investigated.

In the following section I discuss gravity based models for non-competitive location models. In Sect. 11.5, I describe four solution methods that concentrate on competitive models but can be applied to many optimization models as well. I conclude the chapter with ideas for future research.

11.2 The Gravity Competitive Facility Location Model

The gravity formulation for one new facility is as follows: Let n be the number of demand points; the buying power at demand point $1 \le i \le n$ is b_i; p competing facilities with attractiveness A_j are located at $X_j = (x_j, y_j)$ for $1 \le j \le p$. The distance decay function at a distance d from the facility is $f(d)$.

A new facility of attractiveness A is to be located at an unknown location $X = (x, y)$. The distance between demand point i and a location Z is $d_i(Z)$. The market share (total buying power attracted by the new facility) $M(X)$ is

$$M(X) = \sum_{i=1}^{n} b_i \frac{Af[d_i(X)]}{Af[d_i(X)] + \sum_{j=1}^{p} A_j f[d_i(X_j)]} \tag{11.1}$$

This expression can be easily generalized to locating several new facilities and to chain facilities where some of the existing facilities belong to one's chain and are not competitors. Drezner (1994b) found the best location for a new facility based on the gravity rule using the decay function of $f(d) = \frac{1}{d^2}$.

Implementing the gravity model requires the determination of (a) the distance between demand points and facilities, (b) the distance decay function, and (c) the attractiveness level of each competing facility.

11.2.1 The Distance

Hillsman and Rhoda (1978) and Hodgson and Neuman (1993) discussed three types of errors in measuring distance from demand areas to facilities. Type A error is the error in the estimation of distance between the demand point and the facility when the facility is located outside the area represented by the demand point. Type B error is the error in the estimation of distance between the demand point and the facility when the facility is located inside the area represented by the demand point. Type C error is the error caused by incorrect allocation of demand points to facilities. Because the problem discussed below does not involve allocation of demand, only Types A and B errors exist.

11.2.1.1 The Distance Correction

Drezner and Drezner (1997) proposed a distance correction to the gravity model that may apply to other location models as well. In most location models demand is represented by demand points. However, in reality, customers live in neighborhoods. Data may be available by zip codes or census tracts. Listing all individual customers is impractical. It is inaccurate to assume that all customers at a demand "point" are at the same distance from a facility. The distance correction incorporates these considerations. Drezner and Drezner (1997) suggested that if the area of a demand

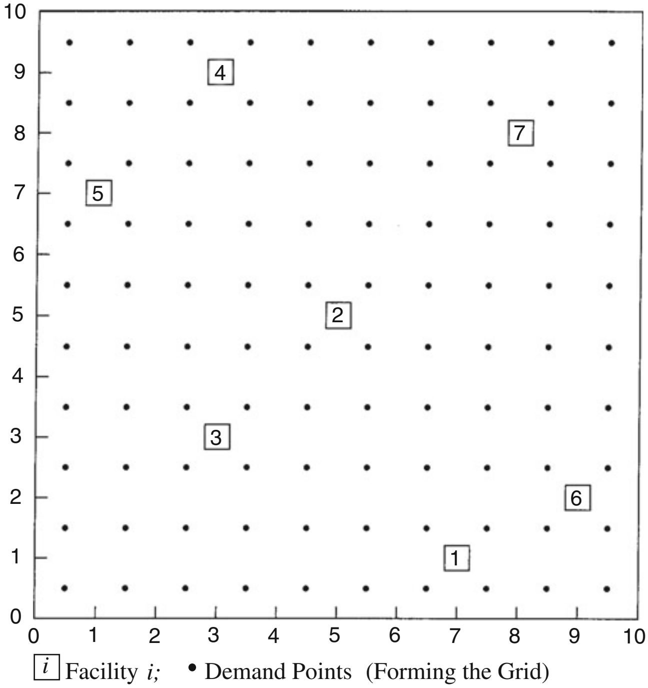
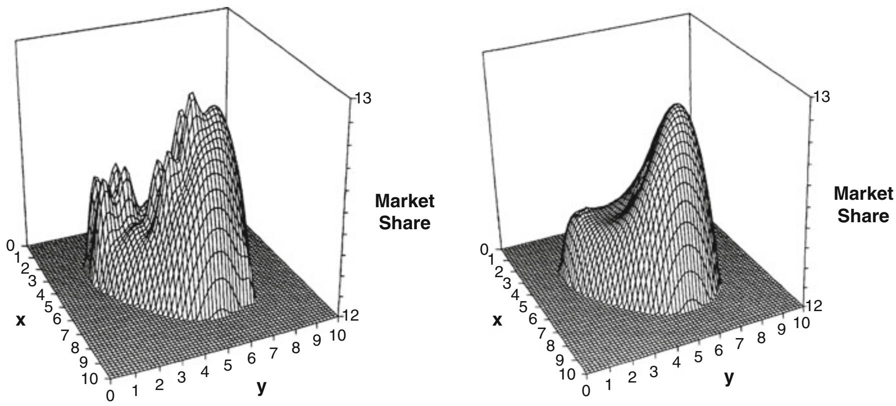


Fig. 11.1 The example problem

"point" is A, the distance to a facility from the center of the area (the demand point) is d, then the corrected distance to be used in the gravity model is about $\sqrt{d^2 + 0.24A}$.

Drezner and Drezner (1997) used an example problem of 100 demand points in a square of size 10 by 10 with seven existing facilities as depicted in Fig. 11.1. This example problem was used in many papers (for example, O'Kelly 1995; Drezner 1995).

Each demand point has an area of 1. The market share captured by the new facility is depicted in Fig. 11.2. On the left, the surface plot of the "standard" Huff model using $f(d) = \frac{1}{d^2}$ as the decay function is depicted. On the right, the market share captured when demand is continuous in the 10 by 10 square is depicted. When demand is generated at demand "points" there are many peaks at various locations. In the continuous case the surface is "smooth" with two local maxima. When a decay function of $f(d) = \frac{1}{d^2+0.24}$ (distance correction) is used (as suggested in Drezner and Drezner 1997), the surface is very close to the continuous surface with two local maxima. See the figure in Drezner and Drezner (1997).



Demand generated at points. **Continuous demand in the square.**

Fig. 11.2 Discrete and continuous market share surfaces

11.2.2 The Distance Decay Function

The distance decay function represents the decline in the probability that a customer patronizes a facility as a function of his distance from it. In the original gravity model (Reilly 1931) it is assumed that the distance decay parallels gravity decay and thus $f(d) = \frac{1}{d^2}$. Huff (1964, 1966) suggested a decay function of $f(d) = \frac{1}{d^\lambda}$, where the power λ depends on the retail category. Huff found $\lambda = 3$ for grocery stores (Huff 1966), $\lambda = 3.191$ for clothing, and $\lambda = 2.723$ for furniture stores (Huff 1964). Drezner (2006) found that for shopping malls $\lambda = 1.27$. This indicates that distance is not as important when selecting a shopping mall as it is when selecting a grocery store to patronize.

Wilson (1976) suggested an exponential decay $e^{-\lambda d}$ which was used in many subsequent papers (Fernandez et al. 2007; Sáiz et al. 2009; Aboolian et al. 2007a,b, 2009; Drezner and Drezner 2008; Hodgson 1981). Drezner (2006) compared power and exponential decay on a real data set and showed that exponential decay fits the data better. Other distance decay functions were successfully implemented. The function $f(d) = e^{-1.705d^{0.409}}$ was used in Bell et al. (1998). A logit function $f(d) = \frac{1}{1+e^{\alpha+\beta d+\gamma d^2}}$ was used in Drezner et al. (1998b). A general decay function in the context of gradual cover was analyzed in Berman and Krass (2002).

11.2.3 The Attractiveness

Huff (1964, 1966) suggested facility floor area as a surrogate for attractiveness. A major improvement on Huff's approach was suggested by Nakanishi and Cooper

(1974) who introduced the multiplicative competitive interaction (MCI) model. The MCI coefficient replaces the floor area with a product of factors, each an attractiveness component. Each factor in the product is raised to a power. Thus, the attractiveness of a facility is a composite of a set of attributes rather than the floor area alone. The MCI model was elaborated on and applied by Jain and Mahajan (1979) to food retailing using specific attractiveness attributes. Drezner (2006) report on a study conducted by Drezner et al. (1998a) who surveyed 272 mall visitors. The survey asked respondents to rate the malls they were familiar with on the nine mall attractiveness attributes that had been identified in previous research:

1. Mall prices
2. Distance to mall
3. Adequate parking
4. Variety of stores
5. Mall safety
6. Food court/restaurants
7. Mall appearance
8. Favorite brand names
9. Movies/entertainment

Of the nine attributes, three were identified by a structural equations modeling analysis (see Drezner et al. (1998a) for details) as predictors of malls' overall attractiveness: (a) variety of stores, (b) mall appearance, and (c) favorite brand names. These three attributes were found to be comparable in their relative importance.

Another extensive grocery stores study is by Bell et al. (1998) who interviewed 520 households over a 2 year period and recorded 30,012 shopping trips. Such approaches to estimating the attractiveness level of competing facilities, based on the MCI concept, require two surveys: one for determining the components of the attractiveness level and a second one for estimating the weight of each component.

Two additional techniques, the inferred attractiveness and the derived attractiveness, were proposed (Drezner 2006; Drezner and Drezner 2000, 2002b). Both were tested for estimating shopping malls' attractiveness levels. Drezner and Drezner (2000, 2002b) proposed the inferred attractiveness approach for estimating the attractiveness of competing facilities, when locations of p existing competing facilities in an area are known. The attractiveness levels are treated as p unknown variables, one variable for each facility. The total buying power captured by each facility is the facility's total annual sales which is available in secondary data sources. Each facility's market share is a function of the attractiveness levels of all competing facilities since all other parameters (buying power, distances to demand points) are known and a specific distance decay function is assumed. The attractiveness levels are then estimated by minimizing the sum of squares of the differences between the actual market shares and the calculated ones. This is similar to the least squares multiple regression. Various distance decay functions were tested in Drezner and Drezner (2002b) yielding comparable results for the facilities' attractiveness levels.

Drezner (2006) proposed the derived attractiveness technique for estimating the attractiveness of competing facilities using distance traveled to the facility. The approach was tested on shopping malls in Orange County, California. 3112 shoppers were intercepted in ten different malls. Shoppers were asked where they reside and whether they came from home. All those who came from home were counted. A list of the shoppers' origin was constructed for each mall. In a manner similar to the inferred attractiveness approach (Drezner and Drezner 2000, 2002b), each mall's attractiveness level was treated as an unknown variable and the expected number of customers visiting each mall from each origin was estimated based on the population at each origin and the distance between the origin and the mall.

To validate the findings of the inferred and derived attractiveness techniques, two surveys were conducted: A survey of 272 respondents was reported in Drezner and Drezner (2002b) and described above. A second survey was conducted in conjunction with the derived attractiveness study (Drezner 2006). The 3112 patrons who were intercepted at the ten malls were also asked to rate the mall they were patronizing. They were asked to rate the mall on the three attributes: variety of stores; mall appearance; and favorite brand names. An attractiveness level was then calculated for each mall.

All four approaches (inferred attractiveness, derived attractiveness, and the two surveys) are compared in Drezner (2006). All methods yielded similar attractiveness level values confirming their validity. The derived attractiveness approach using exponential decay function provided better results than the derived attractiveness level using the power decay function.

Drezner et al. (2018c) introduced random attractiveness to the gravity model. Customers' attractiveness perception of a particular facility is likely to be varied due to varying assessment of facilities' attributes and varying levels of information about the facilities. However, existing competitive facility location models assume that facilities' attractiveness levels are fixed.

Two solution methods are proposed. One method is based on discretizing the attractiveness distribution (Drezner and Zerom 2016). A second method is based on the concept of "effective" attractiveness of a facility. Effective facility attractiveness is defined as the level of fixed attractiveness whose corresponding optimal market share is approximately equal to the optimal expected market share under random attractiveness. When a facility's attractiveness is random, it tends to lose some competitive advantage. The effective attractiveness of every facility is lower than its average attractiveness. The decline in attractiveness for a given mean is approximately proportional to the variance of the attractiveness distribution. The accuracy of the results of both solution methods is confirmed by simulations.

11.2.4 The Relationship Between Attractiveness and Distance Decay

Drezner et al. (2018d) proposed a new gravity based competitive facility model which is easier to implement than existing models. In existing gravity models, all facilities have the same distance decay function. Drezner et al. (2018d) proposed to replace attractiveness multipliers, see (11.1), by varying decay functions for different facilities.

The model is based on the following observation: more attractive facilities attract shoppers from larger distances. Facilities' attractiveness is estimated by actual customers' behavior rather than by complex opinion surveys. For example, for the exponential decay function $f(d) = e^{-\lambda d}$, facilities have different λs rather than the same decay function for all facilities. Demand attracted by more attractive facilities has a slower decay and thus has smaller λ values. No modifications are required in order to apply existing solution algorithms to the new model. The effectiveness of the new approach is demonstrated using a real data set. For complete details see Drezner et al. (2018d).

11.3 Additional Considerations

In this section we present additional refinements to the basic competitive facility location gravity model as well as investigate properties of competitive location models. These considerations apply to gravity models as well as to other competitive location models.

11.3.1 Leader–Follower Models

Drezner and Drezner (1998) considered the leader–follower problem which is anchored in game theory and is also termed the Stackelberg equilibrium (Stackelberg 1934; Drezner and Drezner 2017). The leader plans to locate a facility knowing that a competitor will locate his facility in the future at the best location for him, knowing where the leader located his facility. The follower's problem is identical to the standard competitive location problem because the follower has all the necessary information. The leader's problem is more difficult. He needs to incorporate the follower's reaction in his location decision. For every possible leader's strategy, the follower's problem has to be solved. The leader–follower location model in a competitive environment was also investigated in Drezner (1982), Küçükaydın et al. (2012), Plastria and Vanhaverbeke (2008), Redondo et al. (2013, 2010), Saidani et al. (2012), Sáiz et al. (2009), Drezner et al. (2015), Eiselt (2011), and Eiselt et al. (2015). A recent review of leader–follower models is found in Drezner and Drezner (2017).

Drezner and Drezner (1998) suggested three heuristic procedures for solving the leader's problem:

1. The brute force approach proposes to evaluate the market share on a dense enough grid and hopefully identify the neighborhood of the global optimum. Once this neighborhood is identified, a second finer grid search in this neighborhood is performed.
2. The pseudo mathematical programming approach which consists of formulating a different problem. The leader's market share is maximized subject to the derivative of the follower's market share equal to zero.
3. The gradient search approach which finds a local maximum by applying a gradient search from a randomly selected starting location for the leader's new facility. The gradient search requires the determination of the best location for the follower's competing facility in every evaluation of a possible location.

For complete details see Drezner and Drezner (1998). The computational results indicate that the gradient search performed best. The pseudo mathematical programming approach is recommended for users who do not wish to code a special program but rather use standard software. The brute force approach is recommended if the surface of the market share captured as a function of location is of value to the user.

11.3.2 Budget Constraints

The location of several competing facilities with a given budget is considered in several papers. Facility's attractiveness depends on the investment allocated to it. The attractiveness values of the facilities are variables. In addition to finding the best locations, the purpose of this paper is to investigate a competitive location problem to determine how to allocate a budget to expand company's chain by either adding new facilities, expanding existing facilities, or a combination of both actions. Solving large problems may exceed the computational resources currently available. The authors treat a special case when the market can be divided into mutually exclusive sub-markets. These can be markets in cities around the globe or markets far enough from each other so that it can be assumed that customers in one market do not patronize retail facilities in another market, or that cross-patronizing is negligible. The company has a given budget to invest in these markets. Three objectives are considered: maximizing profit, maximizing return on investment (ROI), and maximizing profit subject to a minimum ROI. An illustrative example problem of 20 sub-markets with a total of 400 facilities, 4800 potential locations for new facilities, and 5000 demand points is optimally solved in less than 2 h of computing time or the facilities the planner also has to decide how to allocate the budget to the facilities. The problem is also termed "location and design" of competing facilities (Aboolian et al. 2008, 2007a; Fernandez et al. 2007). A cover based competitive model for this problem is in Drezner et al. (2012).

Drezner (1998) proposed the limited budget problem. The problem was formulated in a continuous space. An interesting conclusion of the analysis is that new franchises should invest the total budget in one new facility, while established franchises should divide the budget about equally among several new facilities.

This conclusion is similar to the recommendations suggested in the Colonel Blotto game (Roberson 2006). A given number of troops needs to be allocated among several battlefields in order to win the most battlefields in a fight against an enemy. Commanders/colonels with a small number of troops should put all their resources in one battlefield, while commanders with many available troops should allocate about the same number of troops to each battlefield.

Drezner et al. (2016) investigated a competitive location problem to determine how to allocate a budget to expand company's chain by either adding new facilities, expanding existing facilities, or a combination of both actions. Solving large problems may exceed the computational resources currently available. A special case when the market can be divided into mutually exclusive sub-markets is analyzed. These can be markets in cities around the globe or markets far enough from each other so that it can be assumed that customers in one market do not patronize retail facilities in another market, or that cross-patronizing is negligible. The company has a given budget to invest in these markets. Three objectives are considered: maximizing profit, maximizing return on investment (ROI), and maximizing profit subject to a minimum ROI. An illustrative example problem of 20 sub-markets with a total of 400 facilities, 4800 potential locations for new facilities, and 5000 demand points is optimally solved in less than 2 h of computing time. This approach can be implemented in other facilities' location objectives. For example, suppose that in a p-median model, the demand area can be partitioned into mutually exclusive subsets far enough from one another. The number of facilities in each subset is determined so that the total cost is minimized.

11.3.3 Lost Demand

Most competitive location models assume that the whole buying power is distributed among the competing facilities. In reality, if the closest facility is too far, customers may select substitute facilities. For example, if a customer wishes to patronize a Chinese restaurant and there is no such restaurant nearby, he may patronize a different kind of restaurant or prefer to eat at home. His buying power is lost to the competing Chinese restaurants. If the product is not essential, customers may forgo the purchase rather than drive to a far location. The issue of lost demand was analyzed in Berman et al. (2006). The decline in attracted buying power is modeled as a declining step function.

Drezner and Drezner (2008) considered two objectives. One of the objectives is the minimization of the lost buying power, and the second is the maximization of the buying power captured by one's chain.

Drezner and Drezner (2012) suggested to model lost demand by establishing a virtual competitor who is located at a distance D from all potential customers. The virtual competitor is not located at a physical point. The distance D represents the maximum distance customers are willing to travel to the competing facility. The buying power captured by the virtual facility is actually the buying power lost by "real" facilities. The virtual facility is closest to customers that have no "real" facility within a distance D.

11.3.4 Changing Market Conditions

Drezner and Drezner (2002a) considered several ways in which market conditions may change in the future. For example,

1. Buying power changing in time and can be different for different communities which was also assumed in Drezner and Wesolowsky (1991), Farahani et al. (2009, 2014, 2015), and Rezapour et al. (2011).
2. The attractiveness of one's new facility may be different for different time intervals.
3. A new competitor enters the market at some point in the future.
4. A competitor exits the market at some point in the future.
5. A competitor (or one's own facility) is renovated at some point in the future, thereby changing its overall attractiveness.

A scenario can incorporate more than one change in market conditions by combining any number of the scenarios above. One can, for example, incorporate continuous changes in buying power, entry of two competitors, and exit of another competitor in one scenario. The objective is the minimax regret objective, see, for example, Puerto et al. (2009).

Drezner (2009a) analyzed the model for a set of possible scenarios which may exist in the future. The best location for a new retail facility such that the market share captured at that location is as close to the maximum as possible regardless of the future scenario is analyzed. This is also the minimax regret objective.

11.3.5 The Threshold Objective

The objective is minimizing the probability that the facility does not meet a given market share, representing profit. This model is useful when a facility will have to be closed if it does not reach a certain sales level.

Drezner et al. (2002b) considered a new store is to be located. Demand at each demand point follows some probabilistic distribution. The total market share captured at a certain location has some probability density function. The problem is formulated assuming that the generated demand (buying power) at various demand points follows a multivariate normal distribution. It is possible that a correlation exists between the actual demand generated at various demand points. For example, if demand at one demand point is higher than average, then it is likely that it is higher at other demand points because there may be a common cause for the demand to be higher than average.

Drezner and Drezner (2011b) applied the threshold objective to the Weber problem. The objective is to minimize the probability of exceeding a cost threshold.

11.3.6 Consistent and Inconsistent Rules

Drezner et al. (1996) analyzed whether customers can change their selection on the way to a retail facility. Drezner et al. (1996) define a consistency property. A selection rule is consistent if the selection of a facility to patronize does not change along the way to the selected facility. A rule is inconsistent if the selection may change along the way. The distances to the competing facilities change along the way and at some point the selection dictated by the rule may be different.

A practical example is the location of a mom-and-pop store. If it is located on the way to an attractive mall at the outskirts of town, customers who travel to the mall may pass by the mom-and-pop store, change their mind and patronize it. The mom-and-pop store attempts to capture the traffic flow of customers on the way to the mall. However, the mom-and-pop store should not be located too close to the mall. Customers may not change their mind since they are already close to the more attractive mall.

11.3.7 Cannibalization

Cannibalization occurs at the retail level of chain facilities, especially in the case of franchises. When opening a new retail outlet in close proximity to an existing outlet, the new facility cannibalizes the sales of the existing one. Though not a franchise, this applies to Starbucks coffee and other chain retailers. Unlike cannibalization in new product development and introduction that is well researched, cannibalization at the retail level has been overlooked for the most part. With the growth of franchise operations, cannibalization emerges as an important and timely issue. Since companies wish to grow and expand, managers are faced with the strategic decision of optimally locating new, additional facilities such that cannibalization of existing chain members is minimized. There are cases of lawsuits regarding cannibalization in fast food franchise systems such as Arby's, Burger King, KFC, McDonald's, Subway, and Taco Bell. This phenomenon is referred to as encroachment.

A similar problem is observed and documented in the hospitality/lodging industry for such franchise systems as Holiday Inn, Days Inn, Howard Johnson, Ramada, Comfort Inn, and Quality Inn. Franchise contracts contain detailed provisions governing the relationship between franchisors and franchisees for the conduct of business at specific locations but they usually do not restrict the franchisor's ability to expand the franchise systems within a territory. Many franchisees believe they have lost business as a result of cannibalization from new units in the same chain, a phenomenon referred to in the lodging industry as "impact." Disputes between franchisees and franchisors over territorial encroachment have elicited responses from state legislators, who enacted laws to protect franchisees from encroachment and from franchisors themselves who institute policies for managing the impact of system expansion on existing franchised units.

Cannibalization is analyzed in Drezner (2011) and in Plastria (2005). Drezner (2011) found the efficient frontier of the market share versus the cannibalization. The analysis is based on the theorem: "When the limit on the allowed cannibalization is increased, the maximum market share cannot decrease." First, the range of the cannibalization limits is established. Then, an equally spaced list of cannibalization limits in this range is constructed. The maximum value of the market share is found for each cannibalization limit, thus establishing the efficient frontier.

11.3.8 Sequential Location

Drezner and Drezner (2016) investigated sequential location of two facilities belonging to one chain. There are two strategies for locating the first facility:

1. locating it at its single facility optimum,
2. randomly locating it.

The second facility is then located at its optimal location given the first facility's location. Three objectives are tested: minisum, minimax, and market share captured by the gravity rule (the competitive objective). For the competitive objective it was found that optimally locating the first facility is better than locating it at random. On the other hand, for the minisum and minimax objectives it is better to locate the first facility at random.

11.4 Applying the Gravity Rule to Other Objectives

The gravity rule can be applied to other commonly used non-competitive location objectives. Rather than assuming that a user gets services from the closest facility, he chooses a facility according to the gravity rule. The probability of patronizing a facility is proportional to the facility's attractiveness and to some decay function of the distance.

11.4.1 Gravity *p*-Median

In the standard p-median model (Daskin 1995) it is assumed that each user travels to the closest facility. This implicitly implies that facility choice is centrally controlled or that all facilities charge the same price for the service. Drezner and Drezner (2007) proposed the gravity p-median model. It is assumed that users choose from among the facilities providing services according to the gravity rule rather than from the closest facility. Users consider facilities' attractiveness in their choice. Similar to the standard p-median problem, the objective is to minimize the sum of the expected weighted distances.

11.4.2 Gravity Hub Location

Drezner and Drezner (2001) applied the gravity rule to the hub location problem. A traveler needs to fly from one airport to another. Several potential hubs are available. If the origin or the destination is a hub airport, the traveler chooses a non-stop flight. Otherwise, the probability that a certain hub is selected is proportional to the hub's attractiveness (price, walking distance from the arrival gate to the connecting one, chance of inclement weather, etc.) and to a distance decay function such as the total travel distance (or time) raised to a given inverse power. Such a model can be generalized to selecting a sequence of two or more hubs.

11.4.3 Gravity Multiple Server

Drezner and Drezner (2011a) considered the gravity rule version of the multiple server location problem (Berman and Drezner 2007). Total service time consists of travel time to the facility, waiting time in line, and service time. There is a given number of servers to be distributed among the facilities. Each facility acts as an M/M/k queuing system. In Drezner and Drezner (2011a) customers select a server with a probability proportional to its attractiveness and to a decay function of the distance, not necessarily the closest one. Two models are proposed: a stationary one and an interactive one. In the stationary model it is assumed that customers do not consider the expected waiting time in line and service time at the facility in their facility selection decision simply because they do not know these values. In the interactive model it is assumed that customers know the expected waiting time in line and service time at the facility and do consider them in their facility selection decision.

11.4.4 Planar Gravity Models

Drezner and Drezner (2006) considered two problems of locating p facilities in the plane using the gravity rule for customers' facility selection. The first problem is the p median where the total distance traveled by customers is minimized. The second problem focuses on equalizing demand across facilities by minimizing the variance of total demand attracted to each facility. In addition, a multi-objective approach, which combines the two objectives, is considered. Heuristic solution procedures are proposed and tested.

11.5 Solution Approaches

11.5.1 Single Facility

Drezner and Drezner (2004) optimally solved the single facility competitive location problem by applying the big triangle small triangle global optimization algorithm (BTST, Drezner and Suzuki 2004). The procedure BTST requires an effective upper bound on the market share captured when the facility is located anywhere in a triangle. The procedure is very efficient and finds the optimal solution for 10,000 demand points in less than 6 min of computer time. The generalized Weiszfeld algorithm (Drezner 2009b) repeated from 1000 different starting solutions required about the same time for all 1000 runs. The optimum was obtained at least 17 times (for $n = 50$) and the average for $n = 10{,}000$ problems is 726 times out of 1000.

11.5.2 Multiple Facilities

Drezner et al. (2002a) proposed five heuristic procedures (H1–H5) for the maximization of the market share by locating p new facilities with given attractiveness levels using the gravity rule.

H1: Finding a good location for each facility, one at a time, by the generalized Weiszfeld algorithm (Drezner 2009b). This is repeated 100 times from randomly generated starting solutions and the best one is selected.

H2: Same as H1 but the 100 randomly generated starting solutions are generated in sparse configurations considering the previous starting and final solutions obtained so far. For complete details see Drezner et al. (2002a).

H3: Selecting a grid of points that covers the demand area and applying a steepest ascent algorithm when the locations of the new facilities are restricted to grid points.

H4: Applying a simulated annealing approach (Kirkpatrick et al. 1983) for locations restricted to grid points.

H5: Using the final solutions of H4 as starting solutions for the heuristic H1.

Following extensive computational experiments, heuristic H5 is recommended.

11.5.3 The TLA Method

The TLA (tangent line approximation) method (Aboolian et al. 2007b) can find a solution to the gravity model within a given accuracy. For its implementation, the objective function should be a concave function, twice differentiable, and non-decreasing of a linear functional. These conditions hold for the gravity model. The idea is to replace the objective function by a piece-wise linear function. The feasible range is divided into segments and a tangent line is constructed in each segment touching the objective function at the segment's center. The objective function is formulated by adding a binary variable for each segment and maintaining the original constraints. Optimal solutions of the modified problem are then found by non-linear solvers. The number of segments is determined by the pre-specified accuracy. For complete details see Aboolian et al. (2007b).

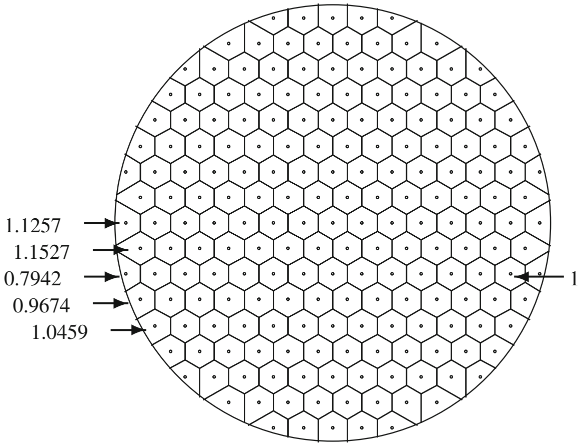
11.5.4 Integrating over a Circular Area

The solution method described below can be useful for many optimization problems that require integration over a circular area.

Drezner et al. (2018b) investigated the cover based competitive model (Drezner et al. 2011, 2012) when demand is generated continuously in an area. In order to evaluate the objective function for a particular disc, the area covered by the intersection of several discs with the particular disc is needed. Furthermore, if an intersection area is covered by k discs, the area is divided by k. Double integration over the particular disc is not simple. We could not find an explicit formula for the objective function which is discontinuous. However, calculating the value of the objective function at any point is fast and easy.

Two approaches are proposed for estimating the integral. One is generating a hexagonal pattern in the circle (for example, one of the options tested consists of 199 points, see Fig. 11.3). Each point covers some area. Points which are not near the circle's circumference cover an identical hexagonal area. The integral is estimated by a weighted summation. This relatively simple approach is recommended particularly when the demand in the circular area is not uniform.

Fig. 11.3 199 points in a circle



A second approach, which was found to be more accurate using the same computation time, is based on Gauss-Legendre quadrature (Abramowitz and Stegun 1972). K Gaussian-Legendre points are drawn on the x-axis through the center of the disc and the integral over each vertical segment can be explicitly evaluated and the one-dimensional integral calculated. In the website https://pomax.github.io/bezierinfo/legendre-gauss.html parameters for all $K \leq 64$ points are given. For complete details see Drezner et al. (2018b).

11.6 Conclusions

There are several topics that are worth further investigation. Distance and buying power at demand points are available. However, attractiveness and distance decay functions are not fully understood. In addition, solution methods can be further improved.

- In existing gravity models, all facilities have the same distance decay function. Drezner et al. (2018d) proposed to replace attractiveness multipliers by varying decay functions for different facilities. Demand attracted by more attractive facilities has a slower distance decay.
- Drezner and Drezner (2019) suggested a new solution algorithm. They assume that a budget is available for expansion of chain facilities. The part of the budget invested in improving an existing facility or constructing a new one is an integer multiplier of a basic value such as 0.1% of the available budget. The model is applied to solving the gravity model (Huff 1964, 1966) by a branch and bound algorithm.

- A retail facility to be located may be obnoxious to some neighborhoods. City zoning may disallow the location of commercial facilities in residential neighborhoods. A similar model with the Weber objective is proposed and solved in Drezner et al. (2018a).

References

Aboolian, R., Berman, O., & Krass, D. (2007a). Competitive facility location and design problem. *European Journal of Operations Research, 182*, 40–62.

Aboolian, R., Berman, O., & Krass, D. (2007b). Competitive facility location model with concave demand. *European Journal of Operations Research, 181*, 598–619.

Aboolian, R., Berman, O., & Krass, D. (2008). Optimizing pricing and location decisions for competitive service facilities charging uniform price. *Journal of the Operational Research Society, 59*, 1506–1519.

Aboolian, R., Berman, O., & Krass, D. (2009). Efficient solution approaches for discrete multi-facility competitive interaction model. *Annals of Operations Research, 167*, 297–306.

Abramowitz, M. & Stegun, I. (1972). *Handbook of mathematical functions*. New York: Dover Publications Inc.

Bell, D., Ho, T., & Tang, C. (1998). Determining where to shop: Fixed and variable costs of shopping. *Journal of Marketing Research, 35*(3), 352–369.

Berman, O., & Drezner, Z. (2007). The multiple server location problem. *Journal of the Operational Research Society, 58*, 91–99.

Berman, O., Drezner, T., Drezner, Z., & Krass, D. (2009). Modeling competitive facility location problems: New approaches and results. In M. Oskoorouchi (Ed.), *Tutorials in operations research* (pp. 156–181). San Diego: INFORMS.

Berman, O., & Krass, D. (2002). The generalized maximal covering location problem. *Computers & Operations Research, 29*, 563–591.

Berman, O., Krass, D., & Wang, J. (2006). Locating service facilities to reduce lost demand. *IIE Transactions, 38*, 933–946.

Black, W. (1984). Choice-set definition in patronage modeling. *Journal of Retailing, 60*, 63–85.

Buffa, E. S. (1976). *Operations management: The management of productive systems*. Santa Barbara: Wiley.

Christaller, W. (1966). *Central places in Southern Germany*. Englewood Cliffs: Prentice-Hall.

Clark, W. A. V. (1968). Consumer travel patterns and the concept of range. *Annals of the Association of American Geographers, 58*, 386–396.

Clark, W. A. V., & Rushton, G. (1970). Models of intra-urban behavior and their implications for central place theory. *Economic Geography, 46*, 486–497.

Daskin, M. S. (1995). *Network and discrete location: Models, algorithms, and applications*. New York: Wiley.

Drezner, T. (1994a). Locating a single new facility among existing unequally attractive facilities. *Journal of Regional Science, 34*, 237–252.

Drezner, T. (1994b). Optimal continuous location of a retail facility, facility attractiveness, and market share: An interactive model. *Journal of Retailing, 70*, 49–64.

Drezner, T. (1995). Competitive facility location in the plane. In Z. Drezner (Ed.), *Facility location: A survey of applications and methods* (pp. 285–300). New York: Springer.

Drezner, T. (1998). Location of multiple retail facilities with limited budget constraints – In continuous space. *Journal of Retailing and Consumer Services, 5*, 173–184.

Drezner, T. (2006). Derived attractiveness of shopping malls. *IMA Journal of Management Mathematics, 17*, 349–358.

Drezner, T. (2009a). Location of retail facilities under conditions of uncertainty. *Annals of Operations Research, 167*, 107–120.
Drezner, T. (2011). Cannibalization in a competitive environment. *International Regional Science Review, 34*, 306–322.
Drezner, T. (2014). A review of competitive facility location in the plane. *Logistics Research, 7*, 1–12.
Drezner, T., & Drezner, Z. (1996). Competitive facilities: Market share and location with random utility. *Journal of Regional Science, 36*, 1–15.
Drezner, T., & Drezner, Z. (1997). Replacing discrete demand with continuous demand in a competitive facility location problem. *Naval Research Logistics, 44*, 81–95.
Drezner, T., & Drezner, Z. (1998). Facility location in anticipation of future competition. *Location Science, 6*, 155–173.
Drezner, T., & Drezner, Z. (2000). Inferring attractiveness of shopping malls. *Communications of the Operations Research Society of Japan, 45*, 444–451.
Drezner, T., & Drezner, Z. (2001). A note on applying the gravity rule to the airline hub problem. *Journal of Regional Science, 41*, 67–73.
Drezner, T., & Drezner, Z. (2002a). Retail facility location under changing market conditions. *IMA Journal of Management Mathematics, 13*, 283–302.
Drezner, T., & Drezner, Z. (2002b). Validating the gravity-based competitive location model using inferred attractiveness. *Annals of Operations Research, 111*, 227–237.
Drezner, T., & Drezner, Z. (2004). Finding the optimal solution to the Huff competitive location model. *Computational Management Science, 1*, 193–208.
Drezner, T., & Drezner, Z. (2006). Multiple facilities location in the plane using the gravity model. *Geographical Analysis, 38*, 391–406.
Drezner, T., & Drezner, Z. (2007). The gravity p-median model. *European Journal of Operational Research, 179*, 1239–1251.
Drezner, T., & Drezner, Z. (2008). Lost demand in a competitive environment. *Journal of the Operational Research Society, 59*, 362–371.
Drezner, T., & Drezner, Z. (2011a). The gravity multiple server location problem. *Computers & Operations Research, 38*, 694–701.
Drezner, T., & Drezner, Z. (2011b). The Weber location problem: The threshold objective. *INFOR: Information Systems and Operational Research, 49*, 212–220.
Drezner, T., & Drezner, Z. (2012). Modelling lost demand in competitive facility location. *Journal of the Operational Research Society, 63*, 201–206.
Drezner, T., & Drezner, Z. (2016). Sequential location of two facilities: Comparing random to optimal location of the first facility. *Annals of Operations Research, 246*, 1–15.
Drezner, T., & Drezner, Z. (2017). Leader-follower models in facility location. In *Spatial interaction models* (pp. 73–104). Springer: New York.
Drezner, T., & Drezner, Z. (2019, in preparation). Discrete budget allocation in competitive facility location.
Drezner, T., Drezner, Z., & Eiselt, H. A. (1996). Consistent and inconsistent rules in competitive facility choice. *Journal of the Operational Research Society, 47*, 1494–1503.
Drezner, T., Drezner, Z., & Kalczynski, P. (2011). A cover-based competitive location model. *Journal of the Operational Research Society, 62*, 100–113.
Drezner, T., Drezner, Z., & Kalczynski, P. (2012). Strategic competitive location: Improving existing and establishing new facilities. *Journal of the Operational Research Society, 63*, 1720–1730.
Drezner, T., Drezner, Z., & Kalczynski, P. (2015). A leader-follower model for discrete competitive facility location. *Computers & Operations Research, 64*, 51–59.
Drezner, T., Drezner, Z., & Kalczynski, P. (2016). The multiple markets competitive location problem. *Kybernetes, 45*, 854 – 865.
Drezner, T., Drezner, Z., & Salhi, S. (2002a). Solving the multiple competitive facilities location problem. *European Journal of Operational Research, 142*, 138–151.

Drezner, T., Drezner, Z., & Schöbel, A. (2018a). The Weber obnoxious facility location model: A Big Arc Small Arc approach. *Computers and Operations Research, 98*, 240–250.
Drezner, T., Drezner, Z., & Shiode, S. (2002b). A threshold satisfying competitive location model. *Journal of Regional Science, 42*, 287–299.
Drezner, T., Drezner, Z., & Suzuki, A. (2018b, in review). A cover based competitive facility location model with continuous demand.
Drezner, T., Drezner, Z., & Zerom, D. (2018c). Competitive facility location with random attractiveness. *Operations Research Letters, 46*, 312–317.
Drezner, T., Drezner, Z., & Zerom, D. (2018d, under review). Facility dependent distance decay in competitive location.
Drezner, T., & Eiselt, H. A. (2002). Consumers in competitive location models. In Z. Drezner & H. W. Hamacher (Eds.), *Facility location: Applications and theory* (pp. 151–178). Berlin: Springer.
Drezner, T., Marcoulides, G. A., & Drezner, Z. (1998a). A procedure for estimating the attractiveness of shopping malls. In *Proceedings of 29th Annual DSI Meeting, Las Vegas, NV* (Vol. II, pp. 1090–1092).
Drezner, Z. (1982). Competitive location strategies for two facilities. *Regional Science and Urban Economics, 12*, 485–493.
Drezner, Z. (2009b). On the convergence of the generalized Weiszfeld algorithm. *Annals of Operations Research, 167*, 327–336.
Drezner, Z., & Suzuki, A. (2004). The big triangle small triangle method for the solution of nonconvex facility location problems. *Operations Research, 52*, 128–135.
Drezner, Z., & Wesolowsky, G. O. (1991). Facility location when demand is time dependent. *Naval Research Logistics, 38*, 763–777.
Drezner, Z., Wesolowsky, G. O., & Drezner, T. (1998b). On the logit approach to competitive facility location. *Journal of Regional Science, 38*, 313–327.
Drezner, Z., & Zerom, D. (2016). A simple and effective discretization of a continuous random variable. *Communications in Statistics - Simulation and Computation, 45*, 3798–3810.
Eiselt, H. A. (2011). Equilibria in competitive location models. In H. A. Eiselt & V. Marianov (Eds.), *Foundations of location analysis* (pp. 139–162). Berlin: Springer.
Eiselt, H. A., Marianov, V., & Drezner, T. (2015). Competitive location models. In G. Laporte, S. Nickel, & F. S. da Gama (Eds.), *Location science* (pp. 365–398). New York: Springer.
Farahani, R., Drezner, Z., & Asgari, N. (2009). Single facility location and relocation problem with time dependent weights and discrete planning horizon. *Annals of Operations Research, 167*, 353–368.
Farahani, R. Z., Rezapour, S., Drezner, T., & Fallah, S. (2014). Competitive supply chain network design: An overview of classifications, models, solution techniques and applications. *Omega, 45*, 92–118.
Farahani, R. Z., Rezapour, S., Drezner, T., Esfahani, A. M., & Amiri-Aref, M. (2015). Locating and capacity planning for retailers of a new supply chain to compete on the plane. *Journal of the Operational Research Society, 66*(7), 1182–1205.
Fernandez, J., Pelegrin, B., Plastria, F., & Toth, B. (2007). Solving a Huff-like competitive location and design model for profit maximization in the plane. *European Journal of Operational Research, 179*, 1274–1287.
Ghosh, A., & Craig, C. S. (1986). An approach to determining optimal location for new services. *Journal of Marketing Research, 23*, 354–362.
Ghosh, A., & Craig, C. S. (1991). FRANSYS: A franchise location model. *Journal of Retailing, 67*, 212–234.
Ghosh, A., & Rushton, G. (1987). *Spatial analysis and location-allocation models*. New York: Van Nostrand Reinhold Company.
Hakimi, S. L. (1981). On locating new facilities in a competitive environment. *Presented at the ISOLDE II conference, Skodsborg*.
Hakimi, S. L. (1983). On locating new facilities in a competitive environment. *European Journal of Operational Research, 12*, 29–35.

Hakimi, S. L. (1986). p-Median theorems for competitive location. *Annals of Operations Research, 6*, 77–98.
Hakimi, S. L. (1990). Locations with spatial interactions: Competitive locations and games. In P. B. Mirchandani & R. L. Francis (Eds.), *Discrete location theory* (pp. 439–478). New York: Wiley-Interscience.
Hillsman, E. L., & Rhoda, R. (1978). Errors in measuring distances from populations to service centers. *Annals of Regional Science, 12*, 74–88.
Hodgson, M. J. (1981). A location-allocation model maximizing consumers' welfare. *Regional Studies, 15*, 493–506.
Hodgson, M. J., & Neuman, S. (1993). A GIS approach to eliminating source C aggregation error in p-median models. *Location Science, 1*, 155–170.
Hotelling, H. (1929). Stability in competition. *Economic Journal, 39*, 41–57.
Huff, D. L. (1964). Defining and estimating a trade area. *Journal of Marketing, 28*, 34–38.
Huff, D. L. (1966). A programmed solution for approximating an optimum retail location. *Land Economics, 42*, 293–303.
Jain, A. K., & Mahajan, V. (1979). Evaluating the competitive environment in retailing using multiplicative competitive interactive models. In J. N. Sheth (Ed.), *Research in marketing* (Vol. 2, pp. 217–235). Greenwich: JAI.
Kirkpatrick, S., Gelat, C. D., & Vecchi, M. P. (1983). Optimization by simulated annealing. *Science, 220*, 671–680.
Küçükaydın, H., Aras, N., & Altınel, I. (2012). A leader–follower game in competitive facility location. *Computers & Operations Research, 39*, 437–448.
Leonardi, G., & Tadei, R. (1984). Random utility demand models and service location. *Regional Science and Urban Economics, 14*, 399–431.
Lösch, A. (1954). *The economics of location*. New Haven: Yale University.
Nakanishi, M., & Cooper, L. G. (1974). Parameter estimate for multiplicative interactive choice model: Least squares approach. *Journal of Marketing Research, 11*, 303–311.
O'Kelly, M. E. (1995). Inferred ideal weights for multiple facilities. In Z. Drezner (Ed.), *Facility location: A survey of applications and methods* (pp. 69–88). New York: Springer.
Plastria, F. (2005). Avoiding cannibalisation and/or competitor reaction in planar single facility location. *Journal of the Operational Research Society of Japan, 48*, 148–157.
Plastria, F., & Vanhaverbeke, L. (2008). Discrete models for competitive location with foresight. *Computers & Operations Research, 35*, 683–700.
Puerto, J., Rodríguez-Chía, A. M., & Tamir, A. (2009). Minimax regret single-facility ordered median location problems on networks. *INFORMS Journal on Computing, 21*, 77–87.
Redondo, J. L., Arrondo, A., Fernández, J., García, I., & Ortigosa, P. M. (2013). A two-level evolutionary algorithm for solving the facility location and design (1|1)-centroid problem on the plane with variable demand. *Journal of Global Optimization, 56*, 983–1005.
Redondo, J. L., Fernández, J., García, I., & Ortigosa, P. M. (2010). Heuristics for the facility location and design (1|1)-centroid problem on the plane. *Computational Optimization and Applications, 45*, 111–141.
Reilly, W. J. (1931). *The law of retail gravitation*. New York: Knickerbocker.
ReVelle, C. (1986). The maximum capture or sphere of influence problem: Hotelling revisited on a network. *Journal of Regional Science, 26*, 343–357.
Rezapour, S., Farahani, R. Z., & Drezner, T. (2011). Strategic design of competing supply chain networks for inelastic demand. *Journal of the Operational Research Society, 62*(10), 1784–1795.
Roberson, B. (2006). The colonel blotto game. *Economic Theory, 29*, 1–24.
Saidani, N., Chu, F., & Chen, H. (2012). Competitive facility location and design with reactions of competitors already in the market. *European Journal of Operational Research, 219*, 9–17.
Sáiz, M. E., Hendrix, E. M., Fernández, J., & Pelegrín, B. (2009). On a branch-and-bound approach for a Huff-like Stackelberg location problem. *OR spectrum, 31*, 679–705.
Serra, D., & ReVelle, C. (1995). Competitive location in discrete space. In Z. Drezner (Ed.), *Facility location: A survey of applications and methods* (pp. 367–386). New York: Springer.

Stackelberg, H. V. (1934). *Marktform und Gleichgewicht*. Vienna: Julius Springer.
Wilson, A. G. (1976). Retailers' profits and consumers' welfare in a spatial interaction shopping mode. In I. Masser (Ed.), *Theory and practice in regional science* (pp. 42–59). London: Pion.
Zeller, R. E., Achabal, D. D., & Brown, L. A. (1980). Market penetration and locational conflict in franchise systems. *Decision Sciences, 11*, 58–80.

Chapter 12
Cover-Based Competitive Location Models

Pawel Kalczynski

12.1 Introduction

In competitive location models a set of demand points, each with known buying power exist in a market area. Competing for the buying power in the area are several of one's chain facilities and competing facilities. If an area has only competitors' facilities and no chain facilities are present in the area, then the chain is considering an area. The competing facilities attract buying power from demand points, yielding market share (the proportion of total buying power in the area captured by one's chain). The objective common to all competitive location models is the maximization of market share. Usually, profit is assumed to be a monotonically increasing function of market share. Therefore, maximizing profit is associated with maximizing market share. Fernandez et al. (2007) and Redondo et al. (2009) deal explicitly with maximizing profit. If there is a cost differential between different locations, setup costs, as well as different pricing policies, which may vary by location, account for such cost differentials. For a review of competitive location models see Berman et al. (2009a).

Therefore, at the core of any competitive location model is the estimation of the market share attracted by each of the competing facilities. All models assume that the market share captured by a facility is dependent on (1) the distances between the demand points and the facilities and (2) the attractiveness of the facilities, and (3) the prices at the facilities. Estimating market share captured is typically done using one of the following approaches.

P. Kalczynski (✉)
Steven G. Mihaylo College of Business and Economics, California State University-Fullerton, Fullerton, CA, USA
e-mail: pkalczynski@fullerton.edu

H. A. Eiselt, V. Marianov (eds.), *Contributions to Location Analysis*, International Series in Operations Research & Management Science 281,
https://doi.org/10.1007/978-3-030-19111-5_12

Hotelling (1929) suggested a rule that each customer patronizes the facility that offers the lowest total cost (including the price of service and transportation cost). Hotelling's approach led to the "proximity" rule in which customers patronize the closest facility (when the prices are equal at all facilities). Competitors compete by setting different service price but, from customers' perspective (total cost), this rule implies that all competing facilities are equally attractive and that total buying power concentrated at a demand point is spent at the same facility. Drezner (1982) analyzed two problems on the plane: location of one competing facility and the leader–follower model (Stackelberg's equilibrium model, in which competitors react to leader's action). The proximity rule led to location-allocation models for the location of several facilities in a competitive environment (Hakimi 1983, 1986, 1990; ReVelle 1986; Ghosh and Rushton 1987; Serra and ReVelle 1995).

Different attractiveness levels of different facilities are incorporated in the proximity rule by defining a utility function (Drezner 1994a, 1995). Utility models were extended to random utility models (Leonardi and Tadei 1984; Drezner and Drezner 1996) or to the logit approach (Drezner et al. 1998).

Huff (1964, 1966) suggested applying the probabilistic gravity rule (Reilly 1931) for estimating market share. Drezner (1994b, 1995) suggested a multi-start approach to finding the best location for one new facility based on the gravity rule. Drezner and Drezner (2004) solved this problem optimally. The simultaneous location of multiple facilities according to the gravity rule was analyzed in Drezner et al. (2002a). Drezner (1998) formulated and solved the problem of locating several facilities, applying the gravity rule, when the attractiveness levels of new facilities are not given but they are variables and a given budget is available for constructing the facilities.

The abovementioned models assume that total demand is satisfied and divided among the competing facilities with no lost demand. Drezner and Drezner (2008) proposed a gravity-based model which considers lost demand. This happens when customers have no facility close enough to them; thus, their demand is unsatisfied. Such a model is realistic for non-essential services.

The problems discussed in this chapter are based on covering models. Covering problems have been researched for many years (for reviews see Schilling et al. 1993; Daskin 1995; Current et al. 2002; Plastria 2002). There are two types of covering problems: (1) covering all the points with the minimum number of facilities (the set covering problem, ReVelle et al. 1976) and (2) covering as many points (or total weight when each demand point has a different weight) with a given number of facilities (the max-covering problem). For network formulations see Church and ReVelle (1974), Megiddo et al. (1983), ReVelle (1986), and Berman (1994), and for planar problems see Drezner (1981), Watson-Gandy (1982), Drezner (1986), and Canovas and Pelegrin (1992).

Our competitive location model is based on equal division of buying power among facilities whose radius of influence captures that demand. A comprehensive discussion of this rule is presented in Section 2 of our original paper Drezner et al. (2011). Equal division may not be accurate for a single consumer but the aggregated market share is estimated reasonably well. This rule is much simpler to implement than gravity models or utility-based models. We only need to estimate the catchment area of competing facilities which yields their radius of influence. There are established methods for estimating the radius of influence of a facility (Beaumont 1991; Toppen and Wapenaar 1994). For example, license plates of cars in the parking lot are recorded and the addresses of the cars' owners obtained. Drezner (2006) conducted interviews with consumers patronizing different shopping malls asking them to provide the zip code of their residence and whether they came from home. Other approaches for estimating market share require numerous parameters for their implementation. Our approach requires only the establishment of the catchment area.

This chapter summarizes four papers on competitive location which are a result of my collaborative work with Dr. Zvi Drezner and Dr. Tammy Drezner:

1. A Cover-Based Competitive Location Model (Drezner et al. 2011)
2. Strategic Competitive Location: Improving Existing and Establishing New Facilities (Drezner et al. 2012)
3. A Leader-Follower Model for Discrete Competitive Location (Drezner et al. 2015)
4. The Multiple Markets Competitive Location Problem (Drezner et al. 2016).

Each of these papers is based on a new cover-based model for competitive location. This new model assumes that each facility has its own radius of influence (sphere of influence, catchment area) and that buying power of a demand point located within the radius of several facilities is equally divided among these facilities, while demand at demand points located outside of any facility's influence is lost.

Consider a demand point with buying power w. It is in the sphere of influence of F one's chain facilities and C competitors' facilities. Here, C contains facilities of all firms competing with one's chain. Suppose that the demand point is in the sphere of influence of q additional chain facilities. We calculate the additional market share gained by the chain's facilities from this demand point. Prior to the change in the chain's facilities, the buying power attracted by one's chain is $w\frac{F}{F+C}$. Note that if $F + C = 0$, i.e., the demand point is outside the sphere of influence of all competing facilities, no buying power is attracted. When the demand point is in the sphere of influence of q additional facilities, the market share attracted by the chain is $w\frac{F+q}{F+C+q}$. Simple algebraic manipulations lead to an increase in buying power attracted to one's chain of

$$w\frac{qC}{(F+C)(F+C+q)}. \tag{12.1}$$

Note that: If $q = 0$, there is no increase in the market share regardless of the values of F and C; if $F = C = 0$ and $q > 0$, then the gain in market share is w; if only one new facility is located, then q can be either 0 or 1.

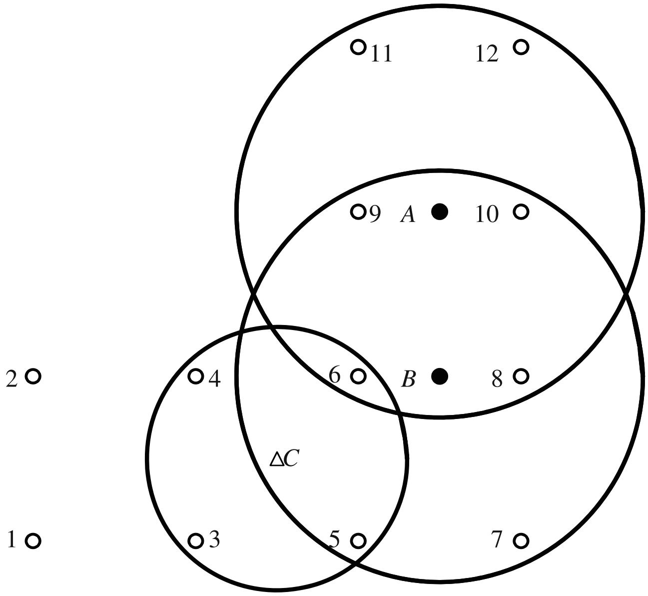


Fig. 12.1 The example problem. Open circle: demand point; filled circle: one's chain; open triangle: competitor

As an illustrative example, consider the problem in Fig. 12.1 originally presented in Drezner et al. (2011). There are 12 demand points located on a grid of size "1," one competitor C whose radius of influence is 0.8, and two more attractive (radius of influence of 1.25) one's chain facilities A and B. Assume that all demand points have a buying power of one unit. One's chain attracts demand points #7–#12, two-thirds of demand point #6, and one-half of demand point #5 for a total market share of $7\frac{1}{6}$ units.

Suppose that a new facility is to be located. If it has the same attractiveness as the competitor's (radius of 0.8), it can capture buying power from at most 4 demand points when it is located at a center of a square whose vertices are demand points. A quick inspection reveals that the best location for the new facility is at the center of the leftmost square capturing buying power from demand points #1–#4. The buying power of demand points #1–#2, which was lost before, is now fully captured. Half of the buying power at demand points #3–#4 is captured. The total market share is increased by 3 units leading to a total market share of $10\frac{1}{6}$ units. If the new facility has the same attractiveness as do the other two chain facilities (radius of 1.25), it can attract 6 demand points and its best location is in the middle between demand points #3 and #4. In that case, it will attract the same 3 units of buying power (all buying power from demand points #1, #2 and half the buying power from demand points #3, #4) and will increase the proportion of the buying power captured from demand points #5 and #6. The buying power captured from demand point #5 increases from one-half to two-thirds and the buying power captured from demand point #6 increases from two-thirds to three quarters for a gain of $3\frac{1}{4}$ units, capturing market share of $10\frac{5}{12}$ units. The competitor attracts $1\frac{7}{12}$ units and a more attractive facility cannot reduce this value. The maximum market share that can be attracted by one's chain by adding one chain facility is therefore $10\frac{5}{12}$ units.

In the remainder of this chapter, the summaries of the main contributions of the four papers are presented, each in a separate section. The chapter ends with a brief conclusion.

12.2 A Cover-Based Competitive Location Model

12.2.1 Locating New Facilities

The problem of locating one new facility in a competitive environment described in the introduction can be converted to the standard max-covering problem with one facility. Each demand point is evaluated and F, C are determined. The additional potential market share is evaluated using $q = 1$ in (12.3). If this value is 0, the demand point can be removed from the problem. The remaining demand points have assigned buying power determined by (12.3) and the maximum covering problem is solved by using, for example, the algorithm in Drezner (1981). This approach is difficult to extend to the location of more than one new facility. The buying power that needs to be assigned to each demand point depends on the number q of new facilities that cover the demand point in their sphere of influence. Therefore, we designed special algorithms for the location of multiple facilities.

Let S be the set of N potential sites either given as part of the problem definition or calculated as in Drezner et al. (2007) when all points on the plane are potential sites. Let a_{ij} for $i = 1, \ldots, n$ and $j \in S$ be an incident matrix. $a_{ij} = 1$ if demand point i is covered by potential location j, and $a_{ij} = 0$ otherwise. Let F_i be the number of one's chain facilities covering demand point i and C_i be the number of the competitor's facilities covering it. Let w_i be the buying power at demand point i. p new facilities are located at some candidate sites in S. Let x_j for $j \in S$ be a 0–1 variable. $x_j = 1$ if a new facility is located at candidate point j and $x_j = 0$ otherwise. The number of new facilities covering demand point i, q_i, is

$$q_i = \sum_{j=1}^{N} a_{ij} x_j. \qquad (12.2)$$

The increase in the market share, $\Delta M(q)$, by one's chain for a given vector $q = \{q_i\}$ is

$$\Delta M(q) = \sum_{i=1}^{n} w_i \frac{q_i C_i}{(F_i + C_i)(F_i + C_i + q_i)}. \qquad (12.3)$$

where w_i is the buying power at demand point i and q_i is calculated by (12.2). Our cover-based competitive location problem formulation is given by

$$\max\left\{\Delta M(X)=\sum_{i=1}^{n}\frac{w_i\frac{C_i}{F_i+C_i}\sum_{j=1}^{N}a_{ij}x_j}{F_i+C_i+\sum_{j=1}^{N}a_{ij}x_j}\right\} \tag{12.4}$$

subject to:

$$\sum_{j\in P}x_j=p$$

$$x_j\in\{0,1\}.$$

A special treatment is needed for the case $F_i+C_i=0$. In this case, if $q_i=0$ the increase in market share is 0, and if $q_i>0$, the increase in market share is w_i. We define two sets of demand points I_1 and I_2:

$$I_1=\{\,i\,|F_i+C_i>0\}\,;\quad I_2=\{\,i\,|F_i+C_i=0\} \tag{12.5}$$

and rewrite (12.4):

$$\max\left\{\Delta M(X)=\sum_{i\in I_1}\frac{w_i\frac{C_i}{F_i+C_i}\sum_{j=1}^{N}a_{ij}x_j}{F_i+C_i+\sum_{j=1}^{N}a_{ij}x_j}+\sum_{i\in I_2}w_i\min\left\{\sum_{j=1}^{N}a_{ij}x_j,\ 1\right\}\right\} \tag{12.6}$$

subject to:

$$\sum_{j\in P}x_j=p$$

$$x_j\in\{0,1\}.$$

Maximizing $\Delta M(X)$ as defined by (12.6) is a non-linear binary programming problem with one constraint. The objective function is a sum of fractional terms. The number of terms is equal to the number of demand points. Each term has linear functions both in the nominator and the denominator. There is only one linear constraint and all decision variables are binary. Therefore, the problem is a generalized binary linear fractional programming problem (Barros 1998, p. 98). Once the binary constraints are relaxed, the problem becomes the sum of linear fractional functions (SOLF) problem (Chen et al. 2005), which is a generalization of the classical linear fractional programming problem (Charnes and Cooper 1962). The SOLF problem is known to be NP-complete when more than one ratio is present in the objective function (Freund and Jarre 2001). The solution procedures for certain SOLF problems can be found in Chen et al. (2005), Nesterov and Nemirovskii (1995), and Falk and Palocsay (1992). Calculating one upper bound on the optimal solution to the generalized binary linear fractional programming problem requires the solution of a SOLF problem. However, solving such a relaxed problem requires significant computer time, especially when N constraints of the type $x_j \leq 1$ need to be added to the problem. In Drezner et al. (2011) we proposed an efficient upper bound, which exploits the special structure of our particular problem. The problem can also be solved heuristically by various metaheuristics such as tabu search, simulated annealing, genetic algorithms, or others.

12.2.2 Upper Bounds for the Cover-Based Competitive Location Problem

In order to apply a branch and bound algorithm, tight upper bounds need to be constructed. In Drezner et al. (2011) we suggested three upper bounds termed UB_1, UB_2, and UB_3. The second upper bound, UB_2, is based on UB_1. The third upper bound UB_3 is an improvement of UB_2 and also depends on UB_1. UB_3 is always tighter than the other two. The reader is referred to the original paper Drezner et al. (2011) for proofs.

12.2.2.1 First Upper Bound (UB_1)

Since $\sum_{j=1}^{N} a_{ij}x_j \geq 0$,

$$\Delta M(X) \leq \sum_{i\in I_1} \frac{w_i C_i \sum_{j=1}^{N} a_{ij}x_j}{(F_i + C_i)^2} + \sum_{i\in I_2} w_i \sum_{j=1}^{N} a_{ij}x_j$$
$$= \sum_{j=1}^{N} \left\{ \sum_{i\in I_1} \frac{w_i C_i a_{ij}}{(F_i + C_i)^2} + \sum_{i\in I_2} w_i a_{ij} \right\} x_j = \sum_{j=1}^{N} \gamma_j x_j, \qquad (12.7)$$

where $\gamma_j = \sum_{i=1}^{n} \frac{w_i C_i a_{ij}}{(F_i+C_i)^2}$ with the provision that if $F_i + C_i = 0$, substitute $C_i = 1$ (and $F_i = 0$).

The following knapsack problem yields an upper bound for the solution of (12.4) or (12.6):

$$\max\left\{\sum_{j=1}^{N}\gamma_j x_j\right\} \tag{12.8}$$

subject to:

$$\sum_{j=1}^{N} x_j = p$$

$$x_j \in \{0, 1\}.$$

The solution to this knapsack problem, UB_1, is the sum of the p largest values of γ_j.

12.2.2.2 Second Upper Bound (UB_2)

Since the arithmetic mean is greater than or equal to the geometric mean:

$$\Delta M(X) \le \sum_{i\in I_1}\frac{w_i C_i \sum_{j=1}^{N} a_{ij}x_j}{2(F_i+C_i)\sqrt{(F_i+C_i)\sum_{j=1}^{N} a_{ij}x_j}} + \sum_{i\in I_2} w_i\sqrt{\sum_{j=1}^{N} a_{ij}x_j}$$

$$= \sum_{i=1}^{n}\frac{w_i C_i\sqrt{\sum_{j=1}^{N} a_{ij}x_j}}{2(F_i+C_i)\sqrt{(F_i+C_i)}} \tag{12.9}$$

with the rule $\frac{C_i}{2(F_i+C_i)\sqrt{(F_i+C_i)}} = 1$ when $F_i + C_i = 0$.

Consider the following identity:

$$\sum_{i=1}^{n}\frac{w_i C_i\sqrt{\sum_{j=1}^{N} a_{ij}x_j}}{2(F_i+C_i)\sqrt{(F_i+C_i)}} = \sum_{i\in I_1}\frac{1}{2}\sqrt{\frac{w_i C_i}{F_i+C_i}} \times \frac{\sqrt{w_i C_i \sum_{j=1}^{N} a_{ij}x_j}}{(F_i+C_i)}$$

$$+ \sum_{i\in I_2}\sqrt{w_i} \times \sqrt{w_i\sum_{j=1}^{N} a_{ij}x_j}.$$

It can be written as $\sum_{i=1}^{n}\frac{w_iC_i\sqrt{\sum_{j=1}^{N}a_{ij}x_j}}{2(F_i+C_i)\sqrt{(F_i+C_i)}} = \sum_{i=1}^{n}\frac{1}{2}\sqrt{\frac{w_iC_i}{F_i+C_i}} \times \frac{\sqrt{w_iC_i\sum_{j=1}^{N}a_{ij}x_j}}{(F_i+C_i)}$ with the rule that when $F_i+C_i=0$, $\frac{C_i}{F_i+C_i}=4$ in the first term and $C_i=1$ in the second term.

By the Schwartz inequality (Hardy et al. 1952) ($\left\{\sum_{i=1}^{n}a_ib_i\right\}^2 \le \sum_{i=1}^{n}a_i^2\sum_{i=1}^{n}b_i^2$):

$$\sum_{i=1}^{n}\frac{w_iC_i\sqrt{\sum_{j=1}^{N}a_{ij}x_j}}{2(F_i+C_i)\sqrt{(F_i+C_i)}} \le \sqrt{\frac{1}{4}\sum_{i=1}^{n}\frac{w_iC_i}{F_i+C_i}\times\sum_{i=1}^{n}\frac{w_iC_i\sum_{j=1}^{N}a_{ij}x_j}{(F_i+C_i)^2}}$$
$$= \sqrt{\frac{1}{4}\sum_{i=1}^{n}\frac{w_iC_i}{F_i+C_i}\times\sum_{j=1}^{N}\gamma_jx_j}$$
$$\le \sqrt{\left\{\frac{1}{4}\sum_{i=1}^{n}\frac{w_iC_i}{F_i+C_i}\right\}UB_1} \tag{12.10}$$

with the rule that when $F_i+C_i=0$, $\frac{C_i}{F_i+C_i}=4$.

To implement UB_2 (12.10) in conjunction with UB_1, i.e., to use as an upper bound $\min\{UB_1, UB_2\}$, calculate: UB_1 and $K=\sum_{i=1}^{n}\frac{w_iC_i}{F_i+C_i}$. If $\frac{1}{4}K < UB_1$ use as upper bound $\frac{1}{2}\sqrt{K\times UB_1}$; otherwise, use UB_1 as the upper bound.

12.2.2.3 Third Upper Bound (UB_3)

In developing UB_2 we used inequalities based on a base of "2," i.e., the arithmetic mean is greater than or equal to the geometric mean and the Schwartz inequality. We can develop formulas based on a base of $\theta > 1$ (not necessarily integer) which reduces to UB_2 when $\theta = 2$ is used. Once the best value of θ is found, UB_3 must be better than UB_2 (or equal to UB_2 when the best θ is equal to 2).

To find UB_3, calculate UB_1, $\alpha = \frac{\sum_{i\in I_1}\frac{w_iC_i}{F_i+C_i}}{UB_1}$, and $\beta = \frac{\sum_{i\in I_2}w_i}{UB_1}$. When $0<\beta<1$ we need to find θ^* that satisfies the equation $\theta = 1+\frac{1}{\alpha}\mu^{\left\{\frac{\beta\mu}{(\theta-1)\alpha+\beta\mu}\right\}} - \frac{\beta\mu}{\alpha}$. Then, $UB_3 = \lambda(\theta^*)UB_1$ is calculated by

$$\lambda(\theta) = \frac{1}{\theta}\left\{(\theta-1)\left[\alpha + \frac{\theta^{\frac{\theta}{\theta-1}}}{\theta-1}\beta\right]\right\}^{\frac{\theta-1}{\theta}} = \left\{\frac{\theta-1}{\mu}\alpha + \beta\right\}^{1-\frac{1}{\theta}}. \tag{12.11}$$

Our original paper Drezner et al. (2011) describes an efficient technique for finding θ^* and the relationships among the three upper bounds.

12.2.3 Heuristic Algorithms

Four heuristic algorithms were constructed and tested: a greedy heuristic, an ascent heuristic, tabu search (Glover 1986), and simulated annealing (Kirkpatrick et al. 1983). We also tested an "improved greedy" approach, i.e., the ascent algorithm was applied to the solution of the greedy algorithm. Tabu search, simulated annealing, and genetic algorithms were constructed in Alp et al. (2003); Drezner et al. (2005); Berman and Drezner (2008) for the p-median and similar problems. The same principles can be adopted for the construction of such algorithms for the solution of our problem. The tabu search and simulated annealing algorithms tested were adopted from these papers.

12.2.3.1 The Greedy Heuristic

The problem is to select a set P of p sites out of N potential locations. We select the set P one facility at a time. The change in the value of the objective function ΔM is calculated by Eq. (12.3) when adding one facility at each of the N potential sites. The site with the largest increase in ΔM is selected for locating a new facility and remains in P for the rest of the algorithm. The process continues until sites for all p new facilities are selected.

12.2.3.2 The Ascent Algorithm

This algorithm is similar to the heuristic algorithm designed by Teitz and Bart (1968) for the solution of the p-median problem.

1. A set P of p sites out of the N available sites are randomly selected.
2. All $p(N-p)$ possible moves by removing one facility in P and adding one of the $N-p$ non-selected sites to P are evaluated.
3. If an improving move is found, the best improving move is executed.
4. If no improving move is found, the algorithm terminates with the last set P of p sites as the solution.

Note that evaluating all $p(N-p)$ moves can be done sequentially and thus streamlined and performed in a shorter run time. All p selected sites are removed in order and ΔM calculated for each. For each of these sites all $N - p$ possible additions are evaluated and the total ΔM is the sum of the changes.

12.2.3.3 Tabu Search

Tabu search (Glover 1986; Glover and Laguna 1997) proceeds from the solution found by the ascent algorithm in an attempt to escape local maxima and obtain better local maxima or the global maximum.

The following simple tabu scheme was used. A node is in the tabu list if it was recently removed from the selected set of nodes. It cannot re-enter the selected set while in the tabu list (unless its inclusion improves the best known solution). When the tabu list consists of $N - p$ members, no exchange is possible. Therefore, we opted to select the tabu tenure to be a fraction of $N - p$. Following extensive experiments we randomly selected the tabu tenure in each iteration in the range $[0.05(N-p), 0.45(N-p)]$. Since the run time of the ascent approach is relatively long, we experimented with relatively few iterations of the tabu search. If the number of the iterations of the ascent algorithm is h, then the number of extra tabu search iterations is set to $19h$, so the run time of the tabu search is about 20 times the run time of the ascent algorithm.

12.2.3.4 The Tabu Search for the Cover-Based Competitive Location Problem

1. A tenure vector consisting of an entry for each facility is maintained. The entry for a facility in the tenure vector is either the last iteration number at which it was removed from P or a large negative number when it was never removed from P.
2. Select the result of the ascent algorithm as a starting solution for the tabu search and as the best found solution. The number of iterations in the ascent algorithm is h.
3. Insert a large negative number (for example, $-N$) for every facility in the tenure vector.
4. Select the tabu tenure, T, in the range $[0.05(N-p), 0.45(N-p)]$.
5. Evaluate all moves (one node to be removed, $i_{out} \in P$, and one node to be added, $i_{in} \notin P$) and calculate the change in the value of the objective function by moving the facility from i_{out} to i_{in}.
6. If a move yields a solution better than the best found one, continue to evaluate all the moves and perform the best improving move. Update the best found solution and go to Step 3.
7. If no move yields a solution better than the best found solution, select the move which yields the best value of the objective function (whether improving or not) as long as the difference between the current iteration and the entry of i_{in} in the tenure vector does not exceed T.

8. Enter the current iteration number into entry i_{out} in the tenure vector. Go to Step 4.
9. Repeat Steps 4–8 for $19h$ iterations.

12.2.4 Computational Experiments

The branch and bound algorithm was coded in J♯. It was run on a machine with 6 CPUs (each clocked at 1.86 GHz) with each processor tackling a different problem. The greedy, ascent, tabu search, and simulated annealing algorithms were coded in Fortran, compiled by Intel 9.0 Fortran compiler and ran on a desktop computer with a 2.8 GHz Pentium IV processor and 256 MB RAM.

We experimented with the 40 Beasley (1990b) problems designed for testing p-median algorithms. The problems tested ranged between $100 \leq n \leq 900$ nodes and $5 \leq p \leq 200$ new facilities. Chain facilities are located at the first ten nodes and the competitors are located at the next 10 nodes. The remaining $n - 20$ nodes are candidate locations for the facilities in one's chain. The demand at node i is $1/i$. The same radius of influence was used for existing and new facilities.

Two sets of problems were run yielding a total of 80 problems. The two sets differ in their radius of influence.

For Set#1 the radius of influence was set to the smallest possible radius that ensures that every each node is attracted by at least one existing facility, whether one's chain facility or a competitor. This guarantees that there is no lost (unmet) demand, thus $\beta = 0$. The radius for each problem is calculated as follows: For each of the $n-20$ candidate nodes the distances to the first 20 nodes are calculated and the smallest distance determined. The maximum among these distances for all $n - 20$ candidate locations is the radius of influence. For Set#2 the radius of influence was set to $R = 20$. In this case there is lost demand prior to establishing new facilities.

12.2.4.1 Set#1

The branch and bound algorithm solved 20 of the 40 problems. The run of the remaining 20 problems was stopped after about 2 days unless we observed that the search is quite advanced after 2 days and was expected to finish in a reasonable additional time. They all reached the best known solution when terminated. A relative accuracy of 10^{-5} was used. For two problems the upper bound was within this relative accuracy from the best known solution so they were instantly solved at the root.

The greedy and ascent heuristic algorithms found the best known solution. The greedy solution was obtained in a few seconds. The ascent algorithm was repeated from 100 randomly generated starting solutions and found the best known solution in *all* 100 runs. Since the greedy and ascent algorithms performed well for Set#1 we saw no need to experiment with tabu search and simulated annealing for the problems in Set#1.

The branch and bound algorithm solved 20 of the 40 problems in Set#1 to optimality illustrating that the upper bound UB_3 is effective. The heuristic algorithms performed extremely well on this set of problems. The best known solution (for half of the problems it has been proven optimal) was found by all runs in a short run time.

12.2.4.2 Set#2

As to Set#2, the branch and bound algorithm found the optimal solution for 16 of the 40 problems. The heuristic algorithms were not as effective for Set #2 as they were for Set #1 and therefore we also experimented with tabu search and simulated annealing.

The ascent algorithm was run 100 times for each problem. It found the best known solution at least once for 37 out of the 40 problems. It found the best known solution in 54.6% of the runs and in all 100 runs for 12 problems.

The tabu search was run 10 times for each problem. It found the best known solution (including 16 known optimal solutions) for all 40 problems. For 29 problems it found the best known solution in all 10 runs. It found the best known solution in 88% of the runs. Run time, by design, was 20 times longer than that for the descent algorithm. Therefore, total run time for 10 tabu solutions was about double the time required for 100 runs of the ascent algorithm.

The results for simulated annealing were inferior to the other algorithms. Results for simulated annealing generally improve when more iterations are allowed in the algorithm. Therefore, longer run time was needed in order to obtain results comparable to those obtained by tabu search. We experimented with run times of more than six times those required for the ascent algorithm (average run time of about 8500 s for 10 runs of the simulated annealing) and obtained the best known results for only 27 of the problems, with the best result averaging 0.238% below the best known result. The best known result was obtained in 5.8 out of 10 runs, on the average. In 19 problems, though, simulated annealing obtained the best known results in all 10 runs.

The reader is referred to our original paper Drezner et al. (2011) for full results of the computational experiment.

12.3 Strategic Competitive Location

The new competitive location model originally proposed in Drezner et al. (2011) and described in Sect. 12.2 of this chapter inspired the follow-up project on strategic competitive location. In Drezner et al. (2011) the location of p new facilities with a given radius is investigated. In the strategic competitive location model proposed in our Drezner et al. (2012) paper, the radii of the facilities are variables, there is a budget constraint, and—in addition to constructing new facilities—we also consider an option to expand existing facilities.

There is a finite set of potential locations for new facilities. Each of the existing facilities has its radius of influence (sphere of influence, catchment area) which is monotonically increasing with its attractiveness. Upgrading existing facilities entails increasing their radius of influence, thereby increasing their catchment area. We do not consider downgrading or closing facilities. A fixed cost plus a variable cost, depending on the radius, is required for improvement. All potential locations for new facilities are defined with a radius of zero. Establishing a new facility requires a given fixed cost (usually greater than the fixed cost required for an improvement of an existing facility) plus a variable cost that is increasing with the radius. These three strategies were captured in a unified model presented in our paper Drezner et al. (2012) where existing facilities and potential new locations are defined as one set of locations, with corresponding radii and setup costs associated with each location.

Models existing prior to our Drezner et al. (2012) paper considered either improving existing facilities or constructing new ones. To the best of our knowledge only Küçükaydın et al. (2011) had analyzed the combination of both options before us, however, in a different context. In our paper an expansion of one's chain is achieved by one of the three strategies: (1) upgrading some or all of one's existing chain facilities, (2) constructing new chain facilities, (3) employing a combination of both (1) and (2). A given budget is available for such expansion. The objective of the chain is to attract the maximum market share (or to maximize additional market share captured following the expansion) within the given budget.

12.3.1 The Three Strategic Competitive Location Models

We consider three strategies, all encompassed in one unified model.

Improvement Strategy: (IMP) Only improvement of existing chain facilities is considered.

Construction Strategy: (NEW) Only construction of new facilities is considered.

Joint Strategy: (JNT) Both improvement of existing facilities and construction of new facilities are considered.

All strategies are treated in a unified model by assigning a radius of zero to potential locations for new facilities. Note that the NEW strategy is somewhat similar to the variable radius covering model (Berman et al. 2009b) where a covering model with no competition was proposed.

12.3.1.1 Notation

n The number of demand points

w_i The buying power at demand point i, $i = 1, \ldots, n$

F_i The number of facilities belonging to one's chain that attract demand point i

C_i The number of competitor facilities attracting demand point i

B	The budget for increasing the attractiveness of existing facilities or creating new ones
p	The number of chain facilities including potential locations for new facilities
r_j^o	The existing radius of facility j for $j = 1, \ldots, p$. For new facilities $r_j^o = 0$
$f(r)$	The cost of building a facility of radius r (a non-decreasing function of r)
r_j	The unknown assigned radius to facility j. $r_j > r_j^o$ for existing facilities and $r_j \geq 0$ for establishing new facilities
S_j	The fixed cost if facility j is improved or established

12.3.1.2 Calculating the Increase in Market Share

In this section we evaluate the increase in buying power captured by the chain as a result of an expansion. An expansion consists of increasing the attractiveness of some chain facilities and/or constructing new ones. The catchment area of a facility is a circle defined by its radius. Demand points in the facility's catchment area are attracted to the facility (are covered by the facility). If a demand point is in the catchment area of several facilities, its buying power is equally divided among these facilities (Drezner et al. 2011). There may exist extreme cases where such a rule can be improved. However this rule provides an estimate for the captured market share and such rare exceptions do not introduce a significant deviation to the estimate. As we explained in Drezner et al. (2011), this rule is simple and robust.

Demand point i is in the catchment area of F_i chain facilities and C_i competitors' facilities. Let q_i be the number of additional chain facilities attracting demand point i following an expansion of the chain. This means that following the expansion demand point i is in the catchment area of $F_i + q_i$ chain facilities. Prior to the expansion of chain facilities, the buying power attracted by one's chain is $\sum\limits_{i=1}^{n} w_i \frac{F_i}{F_i+C_i}$. Note that if $F_i + C_i = 0$, demand point i is outside the catchment area of all competing facilities, no buying power is captured, and the demand at demand point i is lost. Following the change in attractiveness, the market share attracted by the chain is $\sum\limits_{i=1}^{n} w_i \frac{F_i+q_i}{F_i+C_i+q_i}$. The increase in market share captured by one's chain for a given vector $q = \{q_i\}$, $\Delta M(q)$ is given by Eq. (12.3).

Define $\alpha_i(q_i)$ as the proportion of the demand from demand point i added to the chain's market share when demand point i is attracted to q_i additional chain facilities:

$$\alpha_i(q_i) = \frac{q_i C_i}{(F_i + C_i)(F_i + C_i + q_i)}. \tag{12.12}$$

It follows that

$$\Delta M(q) = \sum_{i=1}^{n} w_i \alpha_i(q_i). \tag{12.13}$$

It is easy to show that:

Property 12.1 $0 \le \alpha_i(q_i) \le 1$.

Property 12.2 If $q_i = 0$, there is no increase in the market share captured from demand point i regardless of the values of F_i and C_i, thus $\alpha_i(0) = 0$.

Property 12.3 If $F_i = C_i = 0$ and $q_i > 0$, then the gain in market share from demand point i is w_i, thus $\alpha_i(q_i) = 1$.

12.3.1.3 Preliminary Analysis

When a new facility is established, one can assign a radius of zero to it so that it attracts only the demand point at which it is located. However, the setup cost is added to the total cost. In order to simplify the presentation we assume that potential new locations have a radius of $-\epsilon$ for a very small $\epsilon > 0$, and $f(-\epsilon) = 0$.

If the radius of facility j is increased from r_j^o to $r_j > r_j^o$, the cost of the increase is $f(r_j) - f(r_j^o) + S_j$. Otherwise, the cost is zero. The objective is to maximize the market share attracted to one's chain by increasing some (or all) of the radii, subject to the budget constraint:

$$\sum_{j=1}^{p} \left\{ f(r_j) - f\left(r_j^o\right) \right\} + \sum_{r_j > r_j^o} S_j \le B.$$

The buying power at demand points that are attracted to one's chain facilities but are not attracted to any competitor is fully satisfied by the chain. Thus, the contribution of these demand points to the chain's market share cannot increase. Such demand points can be removed from consideration when calculating the increase in market share captured following the expansion.

Theorem 12.1 *For each facility there is a finite number of radii that should be considered for improvement. Consequently, there is a finite number of feasible candidate solutions.*

The reader is referred to our paper Drezner et al. (2012) for the proof.

Note that even though the number of feasible solutions is finite, it can be very large. If the total increase in market share were an additive function of the individual market share increases, a dynamic programming solution approach would be possible. However, since this condition does not hold, we propose branch and bound and heuristic algorithms rather than solving a non-linear program.

By Theorem 12.1 there is a list of improvement costs to be considered. To define this list let d_{ik} be the distance between demand point i and facility k (existing or new) and B_{ik} be the cost (extra budget required) of increasing the radius of facility k from r_k^o to d_{ik}

$$B_{ik} = \begin{cases} f(d_{ik}) - f(r_k^o) + S_k & \text{when } d_{ik} > r_k^o \\ 0 & \text{otherwise} \end{cases}. \tag{12.14}$$

The set $\{B_{ik}\}$ for a given k may have tied values (when $d_{ik} = d_{mk}$). The sorted list of costs considered for improvement of facility k, following the removal of tied values, consists of 0 (no improvement), and the remaining distinct values in the vector $\{B_{ik}\}$. It is defined as the sorted list of costs $B_K = \{b_{ik}\}$ of cardinality M_k so that $0 = b_{1k} < b_{2k} < \ldots < b_{M_k k}$.

By Theorem 12.1, all feasible candidate solutions are $b_{i_k k}$ for $0 \le i_k \le M_k$ such that $\sum_{k=1}^{p} b_{i_k k} \le B$. The additional market share captured by the chain by investing b_{ik} for improving facility k can be calculated for each feasible solution by obtaining $q_i, i = 1, \ldots, n$ and applying Eq. (12.1).

12.3.2 A Branch and Bound Algorithm

The total enumeration of all feasible candidate solutions can be performed by first evaluating $b_{i_1 1}$ for $i_1 = 1, \ldots M_1$, then for each $1 \le i_1 \le M_1$, evaluating all $b_{i_2 2}$ for $0 \le b_{i_2 2} \le B - b_{i_1 1}$, and so on. That means that for each $2 \le k \le p$ evaluating $0 \le b_{i_k k} \le B - \sum_{m=1}^{p-1} b_{i_m m}$.

Since the number of candidate feasible solutions can be prohibitively large, an upper bound on the possible increase in market share captured (once the first k radii are set) is required for solving moderately sized problems.

12.3.2.1 An Upper Bound

We construct an upper bound using a dynamic programming technique on upper bounds for each facility.

Lemma 12.1 $\alpha_i(q_i + 1) \le \alpha_i(q_i) + \alpha_i(1)$.

The proof of this lemma can be found in Drezner et al. (2012).

Let e_i be the market share added when demand point i is covered by a single additional chain facility. e_i is calculated by (12.1) or (12.12) using the present F_i and C_i for that demand point and $q_i = 1$:

$$e_i = \begin{cases} \dfrac{w_i C_i}{(F_i + C_i)(F_i + C_i + 1)} & \text{when } F_i + C_i > 0 \\ w_i & \text{when } F_i + C_i = 0 \end{cases}. \tag{12.15}$$

Theorem 12.2 *The market share added from one demand point when it is covered by q_i additional chain facilities is less than or equal to $q_i e_i$.*

Proof The theorem is trivially true for $q_i = 0$ and $q_i = 1$. The theorem follows for every $q_i \geq 0$ by mathematical induction on the value of q_i proven by applying Lemma 12.1.

By Theorem 12.2, if we add e_i for any instance of demand point i being covered by an additional chain facility, the sum of these values is an upper bound on the additional market share gained. This suggests an upper bound for the additional market share that can be gained by facilities $k, \ldots, p$ once the radii for the first $k-1$ facilities are known. Note that both e_i (12.15) and the lists B_k for $k = 1, \ldots, p$ can be calculated in the preamble to the branch and bound process.

To make the upper bound simpler to calculate and use, a parameter H (for example, $H = 10{,}000$) is selected. The budget B is then divided into H equal parts. The upper bound is calculated only for an available remaining budget of $h\frac{B}{H}$ for integer $0 \leq h \leq H$. If all values of the budget increase b_{ik} are integers, it is convenient to select $H = B$. When calculating the upper bound for any remaining budget, it is rounded up to the nearest h and the upper bound for this value is used. We create a matrix U of p columns and H rows. U_{hk} is the upper bound for a remaining budget of $h\frac{B}{H}$ available for improving facilities $k, \ldots, p$.

We calculate the upper bound matrix U backward by applying a dynamic programming approach starting from facility p. A matrix V of the same dimension as matrix U is calculated. The value of V_{hk} is the additional market share that can be obtained by using a budget $h\frac{B}{H}$ for $0 \leq h \leq H$ to improve facility k. To calculate column k (V_{hk} for $0 \leq h \leq H$):

1. Set $V_{hk} = 0$ for $h = 0, \ldots, H$.
2. Scan in order all demand points $i = 1, \ldots, n$.
3. The market share added when demand point i is covered by one extra chain facility, e_i, is calculated by Eq. (12.15).
4. The extra budget needed to cover demand point i, b_{ik} is calculated by Eq. (12.14).
5. e_i is added to all entries V_{hk} when the following two conditions hold:

   (a) $b_{ik} > 0$, and
   (b) $h\frac{B}{H} \geq b_{ik}$ which is the same condition as $h \geq \frac{H}{B} b_{ik}$.

   Note that if $b_{ik} = 0$ no action is taken regarding demand point i because $d_{ik} \leq r_k^o$ and the demand point is already covered by facility k.

The matrix U is calculated by using a dynamic programming approach on the matrix V. The last column p in U is identical to the last column p in V. The columns from $k = p - 1$ down are updated in reverse order of the column number by the following recursive formula:

$$U_{hk} = \max_{0 \leq s \leq h} \left\{ V_{sk} + U_{(h-s),(k+1)} \right\}.$$

U_{hk} is calculated starting with $h = H$ down to $h = 0$. Once U_{hk} is calculated, it can replace V_{hk} because only smaller values of h are needed for the calculation of the rest of the column. Therefore, the matrices V and U can occupy the same space in memory.

12.3.2.2 The Algorithm

Suppose that the costs for the first k facilities are assigned. This is represented by a vector $t(1), t(2), \ldots, t(k)$. The costs are $b_{t(j)j}$ for $j = 1, \ldots, k$. This represents a node in the tree. The budget used to expand the first k facilities is $B_0 = \sum_{j=1}^{k} b_{t(j)j}$ which leaves a budget of $B - B_0$ for expanding the remaining facilities. The upper bound on the additional market share captured by facilities $k + 1, \ldots, p$ is $U_{h,k+1}$, where $h = H\frac{B-B_0}{B}$ rounded up. The additional market share captured by the first k facilities is Δ_k and Δ^* is the best solution found so far.

1. Calculate U. Set $k = 1$, $t(1) = 1$, with a budget $B_0 = 0$ $(h = H)$, $\Delta_1 = 0$, and $\Delta^* = 0$. Note that $\Delta^* \geq 0$ can be obtained by a heuristic approach (see Sect. 12.3.3) such as a greedy approach.
2. If $\Delta_k + U_{h,k+1} \leq \Delta^* + \epsilon$, the rest of the tree from this node is fathomed. Go to Step 4.
3. Set $k = k + 1$.
   (a) If $k = p$, calculate the extra market share by using $B - B_0$ to expand facility p. Update Δ^* if necessary. Set $k = p - 1$ and go to Step 4.
   (b) If $k < p$, set $t(k) = 1$. B_0 is unchanged. No additional demand points are covered; thus, all F_i do not change. Go to Step 2.
4. Set $t(k) = t(k) + 1$. B_0 is changed to $B_0 + b_{t(k)k} - b_{t(k)-1,k}$.
   (a) If $B_0 > B$, the search moves back to facility $k - 1$. Set $k = k - 1$.

      If $k > 0$, go to Step 4.
      If $k = 0$, stop with Δ^* as the optimal solution within an accuracy of ϵ.

   (b) If $B_0 \leq B$, calculate $h = H\frac{B-B_0}{B}$ rounded up. The F_i and Δ_k are updated. Go to Step 2.

The reader is referred to our paper Drezner et al. (2012), which offers some interesting modifications to the branch and bound algorithm.

12.3.3 Heuristic Algorithms

We constructed a greedy heuristic, an ascent approach, and tabu search. We also constructed a simulated annealing algorithm but it did not perform well. Tabu search performed best, while the greedy approach was the fastest. We detail the greedy heuristic, the ascent approach (on which the tabu search is based), and tabu search.

For each existing facility or candidate location for a new facility, a list of all possible improvements and their associated cost is compiled. The solution consists of selecting one value from each list and finding the maximum increase in market share among all feasible selections.

12.3.3.1 The Greedy Heuristic

Following extensive experiments with various strategies we found the following approach the most effective:

1. Start with a cost of zero for each column.
2. Evaluate all feasible increases ΔB in the costs for each column and calculate the market share increase ΔM for each.
3. Select the column and cost that maximizes $\Delta M/\Delta B$.
4. Steps 2 and 3 are repeated until no ΔB is feasible.

12.3.3.2 The Ascent Algorithm

The starting solution for the ascent approach is the greedy solution using either the whole budget or a portion of it. Following extensive experiments, the following search neighborhood was found to be most effective: For all combinations of two columns, increases in cost are evaluated for one column combined with decreases in cost for the other column. The neighborhood consists of all feasible combinations based on pairs of columns. Only a partial set of feasible exchanges should be considered in the neighborhood. In the ascent approach, the largest increase in market share is selected until there is no improved combination in the neighborhood.

12.3.3.3 The Tabu Search

The tabu search (Glover 1986; Glover and Laguna 1997) extends the search once the ascent approach terminates. The moves of the ascent approach are continued, whether improving or not, for a pre-specified number of iterations (including the ascent iterations). A tabu list is defined. It consists of columns whose cost was decreased in recent moves. Increasing the cost of such columns is in the tabu list. Each iteration, one of the two possible moves is selected. If a solution better than the best found solution is found (regardless of the tabu list) it is selected. Otherwise, the solution obtained by a move not in the tabu list with the maximum market share (whether improving or not) is selected. The length of the tabu list is randomly generated between t_{min} and t_{max} for every iteration, where t_{min} and t_{max} are parameters of the algorithm.

Our original paper Drezner et al. (2012) offers several time-saving measures.

12.3.4 Computational Experiments

All solution methods were programmed in Fortran using double-precision arithmetic. The programs were compiled by the Intel 11.1 Fortran Compiler and seven different computers were used for running the programs. All computers used were multi-core computers and all cores, except one, were used on each computer with no parallel processing. A total of 23 cores were used.

To enable an easy replication of our test problems for future comparison, we experimented with the 40 Beasley (1990a) problems designed for testing p-median algorithms. The problems ranged between $100 \leq n \leq 900$ nodes. The number of new facilities for these problems was ignored. Chain facilities are located on the first ten nodes and the competitors are located on the next 10 nodes. The demand at node i is $1/i$ and the cost function is $f(r) = r^2$. Since all distances are integers, the cost for improving a facility is integer; thus, we set $H = B$. The same radius of influence was used for existing chain facilities and competing facilities. When new facilities can be added (strategies NEW and JNT), the last $n - 10$ nodes are candidate locations for the facilities in one's chain and are assigned a radius of 0. For each of the 40 problems, three sets of problems (strategies IMP, NEW, and JNT) were run, yielding a total of 120 instances. We experimented with various values of the coverage radius, a fixed cost for establishing a new facility, and an available budget.

The branch and bound algorithm starts with a lower bound of zero. No heuristic was run first to establish a tighter lower bound. An accuracy of $\epsilon = 10^{-5}$ was employed in the branch and bound approach. Following extensive experiments with the branch and bound and tabu algorithms we set the parameter L to 5 for IMP problems and 0 for NEW and JNT problems. The following parameters were used in the tabu search: The number of iterations was set to 1000, and the length of the tabu tenure was randomly generated every iteration between $t_{\min} = 5$ and $t_{\max} = 8$. The starting solutions for the tabu search are the results of the greedy algorithm using between 10% and 100% (randomly generated) of the available budget.

Optimal solutions were found by the branch and bound algorithm for all 40 IMP problems, 19 of the 40 NEW problems, and 15 of the 40 JNT problems. The average IMP solution was 5.87% below the JNT solution and the average NEW solution was 5.96% below the JNT one.

The tabu search for the IMP problems was replicated 10,000 times, while it was replicated 100 times for the other two strategies. The tabu search found the best known solution at least once for all 120 problems. It is found in 96.2% of the runs for the IMP strategy, 81.2% for the NEW strategy, and 75.9% of the time for the JNT strategy. The percent of the average tabu solution was only 0.002% below optimum for the IMP strategy, 0.320% below the best known NEW strategy solutions, and 0.116% below the best known JNT strategy solutions. The branch and bound results for the NEW strategy were 1.03% below the best known solution (obtained by tabu search) and the JNT strategy solution were 0.27% below the best known solution.

The reader is referred to Drezner et al. (2012) for full results.

12.4 A Leader–Follower Model for Discrete Competitive Location

Our next follow-up paper features a game-theoretical approach. The following two extensions of Drezner et al. (2011) were incorporated in Drezner et al. (2015):

Budget Constraints: Combining the location decision with facility design (treating the attractiveness level of the facility as a variable) was recently investigated in Aboolian et al. (2007), Drezner (1998), Drezner et al. (2011, 2012), Fernandez et al. (2007), Plastria and Carrizosa (2004), Redondo et al. (2010), and Toth et al. (2009). Drezner (1998) assumed that the facilities' attractiveness are variables. In that paper it is assumed that a budget is available for locating new facilities and for establishing their attractiveness levels. One needs to determine the facilities' attractiveness levels so that the available budget is not exceeded. Plastria and Vanhaverbeke (2008) combined the limited budget model with the leader–follower model. Aboolian et al. (2007) studied the problem of simultaneously finding the number of facilities, their respective locations, and attractiveness (design) levels.

Leader–Follower: The leader–follower model (Stackelberg 1934) considers a competitor's reaction to the leader's action. The leader decides to expand his chain. The follower is aware of the action taken by the leader and expands his facilities to maximize his own market share. The leader's objective becomes maximizing his market share *following* the follower's reaction. The leader–follower location model in a competitive environment was investigated in Drezner and Drezner (1998), Drezner (1982), Küçükaydın et al. (2012), Plastria and Vanhaverbeke (2008), Redondo et al. (2013, 2010), Saidani et al. (2012), and Sáiz et al. (2009).

With these extensions, we were able to analyze and solve the leader–follower model incorporating facilities' attractiveness (design), subject to limited budgets for both the leader and follower. We were also able to investigate what is the main source of extra market share for the leader and the follower.

12.4.1 Formulation of the Leader–Follower Model

We adapted the cover-based model Drezner et al. (2011, 2012). In our first paper on competitive location Drezner et al. (2011), the location of p new facilities with a given radius is sought so as to maximize the market share captured by one's chain.

In our second paper, Drezner et al. (2012), three strategies were investigated: In the improvement strategy (IMP) only the improvement of existing chain facilities is considered; in the construction strategy (NEW) only the construction of new facilities is considered; and in the joint strategy (JNT) both improvement of existing chain facilities and construction of new facilities are considered. All three strategies are treated in a unified model by assigning a radius of zero to potential locations of new facilities.

In this new formulation, the leader employs one of the three strategies and the follower also implements one of these three strategies. This setting gives rise to nine possible models. Each model is a combination of the strategy employed by the leader and the strategy employed by the follower. For example, the leader employs the JNT model, i.e., considers both improving existing facilities and establishing new ones, while the follower may employ the IMP model, i.e., only considers the improvement of his existing facilities. The most logical model is to employ for both the leader and the follower the JNT strategy which yields the highest market share. However, constructing new facilities or improving existing ones may not be a feasible option for the leader or the follower.

12.4.1.1 Notation

The set of potential locations for the facilities is discrete.

N	The set of demand points of cardinality n
w_i	The buying power at demand point i, $i = 1, \ldots, n$
L_i	The number of facilities that belong to the leader's chain that attract demand point i
F_i	The number of follower's facilities attracting demand point i
B_L	The budget available to the leader for increasing the attractiveness of existing facilities or constructing new ones
B_F	The budget available to the follower for increasing the attractiveness of existing facilities or constructing new ones
P_L	The set of the existing leader's facilities including potential locations for new facilities of cardinality p_L
P_F	The set of the existing follower's facilities including potential locations for new facilities of cardinality p_F
p	The total number of facilities. $p = p_L + p_F$
d_{ij}	The distance between demand point i and facility j
r_j^o	The present radius of facility j for $j = 1, \ldots, p$. For new facilities $r_j^o = -\epsilon$ for a very small ϵ to guarantee that new facilities do not attract demand at their potential location
$f(r)$	The cost of building a facility of radius r (a non-decreasing function of r)
r_j	The unknown radius assigned to facility j for $j = 1, \ldots, p$
R	The set of unknown radii $\{r_j\}$ for $j = 1, \ldots, p$

S_j The fixed cost if facility j is improved or established, i.e., $r_j > r_j^o$ for existing facilities and $r_j \geq 0$ for establishing new facilities

$C(r_j)$ The cost of improving facility j to a radius r_j. It is zero if $r_j = r_j^o$, and $f(r_j) - f(r_j^o) + S_j$, otherwise

Note that the radii r_j are continuous variables. However, it is sufficient to consider a finite number of radii in order to find the optimal solution. Consider the sorted vector of distances between facility j and all n demand points. A radius between two consecutive distances covers the same demand points as the radius equal to the shorter of the two distances yielding the same value of the objective function. Since the improvement cost is an increasing function of the radius, an optimal solution exists for radii that are equal to a distance to a demand point.

12.4.1.2 Calculating the Market Share

For demand point i, the numbers L_i and F_i can be calculated by counting the number of leader's facilities that cover demand point i, and the number of follower's facilities that cover it. Formally, let $\delta_{ij}(R) = \begin{cases} 1 & d_{ij} \leq r_j \\ 0 & \text{Otherwise} \end{cases}$, then L_i and F_i for a given strategy R are

$$L_i = \sum_{j \in P_L} \delta_{ij}(R); \quad F_i = \sum_{j \in P_F} \delta_{ij}(R). \tag{12.16}$$

The objective functions by the leader and the follower, before locating new facilities, are then calculated (Drezner et al. 2011, 2012):

$$MS_L = \sum_{i=1}^{n} w_i \frac{L_i}{L_i + F_i}. \tag{12.17}$$

$$MS_F = \sum_{i=1}^{n} w_i \frac{F_i}{L_i + F_i}. \tag{12.18}$$

Note that if $F_i = L_i = 0$, then in (12.17) $\frac{L_i}{L_i+F_i} = 0$ and in (12.18) $\frac{F_i}{L_i+F_i} = 0$ and the demand w_i associated with demand point i is lost.

Suppose the leader improves some of his facilities and establishes new ones. Note that F_i does not depend on the actions taken by the leader. The follower's problem is thus well-defined following the leader's action and can be optimally solved by the branch and bound algorithm detailed in Drezner et al. (2012).

Once the follower's optimal solution is known, the leader's objective function is well defined as his market share is calculated by (12.17) incorporating changes to the follower's facilities locations and radii.

12.4.1.3 Calculating the Increase in Market Share

To solve the follower's problem it is more efficient to maximize the increase in market share rather than the market share itself. Upper bounds developed for the increase in market share are tighter. Demand point i is in the catchment area of L_i leader's facilities and F_i follower's facilities. Suppose that a radius of some of the follower's facilities is increased and the number of follower's facilities that cover demand point i increased from F_i to $F_i + \Delta F_i$. Let $Q \subseteq N$ be the set of demand points for which $\Delta F_i > 0$. For $i \in Q$ the buying power attracted by the follower was $w_i \frac{F_i}{F_i+L_i}$ and after the change it is $w_i \frac{F_i+\Delta F_i}{F_i+L_i+\Delta F_i}$ leading to a market share increase of $w_i \frac{\Delta F_i L_i}{(F_i+L_i)(F_i+L_i+\Delta F_i)}$. The increase in market share is therefore:

$$\Delta M_F = \sum_{i \in Q} w_i \frac{\Delta F_i L_i}{(F_i + L_i)(F_i + L_i + \Delta F_i)}. \tag{12.19}$$

Note that when $F_i = L_i = 0$ for $i \in Q$, the ratio in (12.19) is equal to one and for such demand points the follower's market share is increased by w_i. The demand w_i was lost before the increase because no facility attracted it but, following the increase, the whole demand w_i is captured by the follower.

12.4.1.4 The Objective Functions

The follower wishes to maximize the increase in his market share ΔM_F calculated by Eq. (12.19). The follower "knows" the values of L_i for $i = 1, \ldots, n$ because his competitor (the leader) has already taken action. The follower can increase the radius of influence of his facilities subject to his available budget, thus increasing some of his radii of influence defining the set Q of demand points that are covered by at least one additional follower's facilities. The leader anticipates the follower's reaction. Therefore, once the follower's problem is solved, the values of L_i, F_i for $i = 1, \ldots n$ are all known and the leader's value of the objective function is calculated by (12.17).

12.4.1.5 The Constraints

The leader and follower cannot exceed their respective budgets. For a combined strategy $R = \{r_j\}$ by both competitors the constraints are

$$\sum_{j \in P_L} C(r_j) \leq B_L; \quad \sum_{j \in P_F} C(r_j) \leq B_F. \tag{12.20}$$

12.4.1.6 The Two Formulations

Once the strategy of the leader is known and thus L_i are defined, the follower's problem is

$$\max_{r_j,\ j\in P_F} \left\{ \sum_{i=1}^{n} w_i \frac{\sum\limits_{j\in P_F} \delta_{ij}(R)}{L_i + \sum\limits_{j\in P_F} \delta_{ij}(R)} \right\}$$

Subject to:

$$\sum_{j\in P_F} C(r_j) \le B_F. \tag{12.21}$$

The leader's problem needs to be formulated as a bi-level programming model (Gao et al. 2005; Sun et al. 2008):

$$\max_{r_j,\ j\in P_L} \left\{ \sum_{i=1}^{n} w_i \frac{\sum\limits_{j\in P_L} \delta_{ij}(R)}{\sum\limits_{j\in P_L} \delta_{ij}(R) + \sum\limits_{j\in P_F} \delta_{ij}(R)} \right\}$$

Subject to: (12.22)

$$\sum_{j\in P_L} C(r_j) \le B_L$$

$$r_j \text{ for } j \in P_F = \arg \left[\begin{array}{l} \max\limits_{r_j,\ j\in P_F} \left\{ \sum\limits_{i=1}^{n} w_i \frac{\sum\limits_{j\in P_F} \delta_{ij}(R)}{\sum\limits_{j\in P_L} \delta_{ij}(R) + \sum\limits_{j\in P_F} \delta_{ij}(R)} \right\} \\ \text{subject to: } \sum\limits_{j\in P_F} C(r_j) \le B_F. \end{array} \right]$$

Note that the follower problem may have several optimal solutions (each resulting in different leader's objective) and the leader does not know which one of these the follower will select. This issue exists in all leader–follower models.

12.4.2 Solution Algorithms

The follower's problems are identical to the three problems analyzed in Drezner et al. (2012) because market conditions are fully known to the follower. A branch and bound algorithm as well as a tabu search (Glover 1977, 1986; Glover and Laguna 1997) were proposed in Drezner et al. (2012) for the solution of each of these three strategies.

The branch and bound algorithm is based on an upper bound on the increase in market share which is calculated by dynamic programming. The set of possible locations of facilities is discrete. It is shown in Drezner et al. (2012) that, in order to find an optimal solution, only a finite list of radii needs to be considered at each location. Branching is performed on a tree whose nodes are combinations of facilities' locations and their possible radii. The same algorithms are used for solving the three strategies. The follower's problem should be solved by the branch and bound algorithm because it is essential to get his optimal solution.

It may be problematic to develop an effective upper bound for the leader's problem. Even if such an upper bound could be constructed, the number of nodes to be evaluated by the branch and bound procedure might be prohibitive. According to the computational experiments reported in Drezner et al. (2012), solving the simplest strategy problem (IMP) may require scanning more than 18,000 nodes for $n = 200$ problems and almost 170 million nodes for one $n = 300$ problem. Such a large number of nodes will severely restrict the size of the problems that can be solved. In Drezner et al. (2015) we proposed to solve the leader's problem by a tabu search algorithm. Note, however, that evaluating each move in the neighborhood requires finding an optimal solution to the follower's problem, and thus it affects the size of the neighborhood. For this reason, the number of iterations would be quite limited as well. We proposed a tabu search algorithm for the solution of the leader's problem similar to the algorithm proposed in Drezner et al. (2012).

12.4.2.1 The Greedy-Type Heuristic for Generating Starting Solutions

For each of the three strategies there is a list of existing facilities (either existing chain facilities with their given radii or potential locations for new facilities with a radius of zero). For each such facility a feasible list of radii is constructed. As explained in Drezner et al. (2012) this list consists of all the distances between the facility and the demand points. A solution is represented by a radius assigned to each facility (either no change in the radius or an increase in the radius with a setup cost added to the total cost).

A leader's starting solution satisfying the budget constraint is generated. No reaction by the follower is considered in the greedy-type algorithm. Following extensive experiments with various strategies, we found that the following approach generates the best starting solutions for heuristic algorithms (such as tabu search) applied for the solution of leader's problem. The follower's problem is solved by a rigorous algorithm and a starting solution is not needed. To insert a random component to the process (so that different starting solutions are generated when the process is repeated), only a random percent, in a given range, of the budget is used.

1. A budget of zero (i.e., the existing radius for the facility) is assigned to each facility.
2. Evaluate all feasible increases ΔB in the budgets for each facility and calculate the market share increase ΔM for each.
3. Select the facility that maximizes $\Delta M / \Delta B$.

4. Update the budget for the selected facility and the remaining available budget.
5. Steps 2 and 3 are repeated until no ΔB is feasible.

12.4.2.2 The Tabu Search Algorithm

We used the tabu search algorithm to solve the leader's problem. For each leader's solution the value of the leader's objective function is calculated by optimally solving the follower's problem and evaluating the extra market share attracted by the leader following the follower's reaction. A tabu list is created. At the start it is empty. A facility whose radius was recently reduced (note that it cannot be below the existing radius) is in the tabu list meaning that its radius is not considered for increase unless the best value of the objective function is improved by the move. A facility remains in the tabu list for tabu tenure iterations and the tabu tenure is randomly generated within a range every iteration. The process is continued for a pre-specified number of iterations and the best solution encountered during the search is the result of the tabu search.

12.4.2.3 The Neighborhood of the Tabu Search

We apply the neighborhood for solving the leader's problem, that was successfully employed in Drezner et al. (2012) for solving the follower's problem, with all the time-saving measures described there. For completeness we briefly describe the neighborhood construction.

- The individual budget currently used to expand facility k be b_k.
- In the current iteration a budget of $B_0 = \sum_{k=1}^{p} b_k$ is used.
- The maximum budget used by any facility is $B_{\max} = \max_{1 \le k \le p} \{b_k\}$.
- Cost increases are considered for all $k = 1, \ldots, p$ up to a budget of $B_L + B_{\max} - B_0$.
- For each such k and its cost increase, cost reductions are considered for all $j \neq k$ as long as $b_j > 0$. Let the budget B_0 following the increase in the budget of facility k be B'_L. It is required that $b_j \ge B'_L - B_L$. In addition, decreases in b_j are considered sequentially starting from the smallest decrease and moving up the line of decreases up to a decrease of b_j guaranteeing that the radius of facility j is not smaller than the existing radius.
- Once a decrease in b_j leads to a budget not exceeding B_L for the first time, the solution is considered for the move. Consequently, for each pair of facilities k, j, at most one radius for facility j is considered. Note that if the radius of a facility is decreased and some demand points that were covered by the facility are no longer covered, the same equation (12.19) can be used by defining Q as the set of demand points whose cover was reduced and using as F_i the number of facilities covering the demand point *following* the change.

It should be emphasized the follower's problem is optimally solved for *each move* in the neighborhood yielding the leader's objective function for that member of the neighborhood. This is required for only one reduction in the budget of some j for each increase in the budget for some k which limits the number of the follower's problems to be solved at each iteration of the tabu search to p_L.

12.4.2.4 The Algorithm

Let $F(X) = MS_L(X)$ be the value of the leader's objective function for a solution X.

1. Generate a feasible starting solution $X = X_0$, empty the tabu list. The best solution found so far is $X^* = X_0$ with the best value of the objective function found so far $F^* = F(X^*)$.
2. The tabu tenure is randomly generated in a pre-specified range $[t_{\min}, t_{\max}]$.
3. The value of the objective function is evaluated at all solutions in the neighborhood of X.
4. The best solution in the neighborhood is X'.
5. If $F(X') > F^*$, the next iterate is $X = X'$. The facility whose radius was reduced is entered into the tabu list and X^* and F^* are updated. Go to Step 7.
6. Otherwise, let X'' be the best solution in the neighborhood for which the facility whose radius is increased is not in the tabu list. The next iterate is $X = X''$. The facility whose radius was reduced is entered into the tabu list.
7. Increase the iteration number by one. Go to Step 2 unless the pre-specified number of iterations is reached.
8. The result of the tabu search is X^* with a value of the objective function F^*.

An efficient way to handle the tabu list (especially when the tabu tenure is randomly generated) is to maintain a tenure vector for all facilities. Initially, a large negative number is recorded for all facilities. When a facility is entered into the tabu list the iteration number is recorded for it. A facility is in the tabu list if the difference between the current iteration number and its recorded value in the tenure vector is less than or equal to the tabu tenure.

12.4.3 Computational Experiments

As in our previous papers, we experimented with the 40 Beasley (1990a) problem instances designed for testing p-median algorithms in order to enable an easy replication of our results. The problems ranged between $100 \leq n \leq 900$ nodes. The number of new facilities for these problems was ignored. The leader's facilities are located on the first ten nodes and the follower's facilities are located on the next 10 nodes. The problems are those tested in Drezner et al. (2012). The demand at node i is $1/i$ (for testing problems with no reaction by the competitor) and the cost function is $f(r) = r^2$. The same radius of influence was used for existing leader's and follower's facilities. When new facilities can be added (strategies NEW and JNT), $n - 10$ nodes are candidate locations for the new facilities (nodes that occupy one's facilities are not candidates for new facilities) and are assigned a radius of 0. We used $r_0^j = 20$, and $S_j = 0$ for expanding existing facilities and the same $S_j > 0$ for establishing any new facility.

The programs for finding the optimal solution, with no reaction by the competitor, were coded in Fortran and compiled by an Intel 11.1 Fortran Compiler with no parallel processing, and run on a desktop with the Intel 870/i7 2.93 GHz CPU Quad processor and 8 GB RAM. Only one thread was used. By the computational experiments in Drezner et al. (2012) a budget of 5000 leads to very long computational times. Therefore, in preparation for solving the leader–follower model, we first experimented with a budget of 1500. For larger budgets run times may be prohibitive and it may be necessary to replace the branch and bound procedure with the effective tabu search described in Drezner et al. (2012).

The branch and bound optimal algorithm is used to solve the follower's problem. Since it is used numerous times in the solution procedure for the leader's problem, we opted to apply a budget of 1500 and a setup cost of 500 for both the leader and the follower. Both the leader and the follower apply the JNT strategy. Tabu search, which does not guarantee optimality, is used to solve the leader's problem. We therefore repeated the solution of each problem instance for at least 20 times to assess the quality of the tabu search solutions. We were able to solve (in reasonable run times) problems with up to 400 demand points.

12.4.3.1 Computational Experiments with No Competitor's Reaction

In our "leader–follower" paper Drezner et al. (2015), we first reported experiments with solving the leader's problem with no reaction from the follower. This was needed to establish a baseline and was performed by the branch and bound rigorous algorithm. When solving the leader–follower problem, this algorithm is performed to find the follower's optimal solution and consequently the leader's objective function.

In Drezner et al. (2015), we also reported the analysis of the percent of market share captured: by the chain (leader), by the competitor (follower), and from lost demand as a function of the budget following the leader's optimal action. The follower takes no action. The setup cost is $S_j = 300$ and the JNT strategy is applied. As expected, one's chain market share increases and the competitor's market share declines. Some of the increase in the leader's market share comes at the expense of the competitor and some comes from capturing demand that is presently lost. It is interesting that the proportion of the additional market share from the competitor remains almost constant for all budgets tested. The average for all 40 problems is 44.2% for a budget of 1500, 44.2% for a budget of 2000, 44.9% for a budget of 2500, and 46.7% for a budget of 5000. A larger percentage of market share gained comes from lost demand. These percentages are the complements of the percentages gained from competitors or about 55%.

For the branch and bound algorithm for the JNT strategy solving the leader's problem when the follower does not react, problems with up to 400 demand points were solved in less than a quarter of a second. We then tested the three strategies for a setup cost of $S_j = 300$. The lower setup cost provides more options for the leader and, therefore, the run times (and number of nodes) are significantly higher, especially for large values of n.

12.4.3.2 Computational Experiments for Solving the Leader–Follower Problem

The tabu search procedure for finding the leader's best solution after the follower's reaction was programmed in C#. Its effectiveness was first tested on 160 JNT instances (40 instances for each budget) optimally solved, i.e., assuming no follower's reaction. Our tabu search was capable of finding optimal solutions to 148 out of 160 instances and sub-optimal solutions (avg. error: 0.11%, max error: 0.41%) to the remaining 12 instances. We also observed that, for the budget of 1500, only one instance (#31) was not solved optimally by tabu search and the error was 0.04%. In the subsequent experiments reported in Drezner et al. (2015), we considered a budget of 1500.

Next, we extended the leader's solution by adding the branch and bound procedure (the follower's solution) to the tabu search. The original parameters used for solving the leader's problem $\left(w_i = \frac{1}{i}\right)$ did not provide interesting results because the weights declined as the index of the demand point increased and thus both the leader and the follower concentrated their effort on attracting demand from demand points with a low index (high w_i) and "ignored" demand points with higher indices. We therefore assigned equal weights of "1" to all demand points. We used a budget of 1500 and a setup cost of 500 for both the leader and the follower.

As a result of extensive experiments, the following parameters were used in the tabu search for solving the leader's problem: The number of iterations was set to 1000, and the length of the tabu tenure was randomly generated every iteration between $t_{\min} = 5$ and $t_{\max} = 8$. The starting solutions for the tabu search are the results of the greedy algorithm described in Sect. 12.4.2.1 using between 10% and 100% (randomly generated) of the available budget.

The optimal solution for the follower was found by using the Fortran program that finds the optimal solution for the follower once the action taken by the leader is known. Run times were quite long so we solved the first 20 problems up to 400 demand points. We solved the first 10 problems 100 times each, the next five problems ($n = 300$) 50 times each, and the next 5 problems ($n = 400$) 20 times each. Recall that the follower's problem was solved optimally and thus these results are valid.

The reader is referred to our paper Drezner et al. (2015) for details and comprehensive discussion of the results.

12.5 The Multiple Markets Competitive Location Problem

In our third follow-up paper Drezner et al. (2016), we consider a competitive location model with a very large number of demand points and facilities. Applying existing solution methods may, at best, provide a good heuristic solution.

The basic problem the company faces is how to invest its available budget in order to expand chain facilities, either by improving the attractiveness of some existing ones, by building new facilities, or by a combination of both actions. Such problems cannot be optimally solved for large instances with currently available computational resources. In Drezner et al. (2016), we investigated a special case for which optimal solutions may be obtained for large problems, and illustrated this approach by optimally solving a problem with 5000 demand points and 400 existing facilities (200 chain facilities and 200 competing facilities).

It is quite common for large problems that a large market area consists of a union of mutually exclusive sub-markets. An international corporation (for example, McDonald's) has facilities in many markets that are mutually exclusive, i.e., customers in one market area do not patronize outlets in other markets or cross-patronizing between markets is negligible. This may well be the case even on a smaller scale when the market can be partitioned to "almost" mutually exclusive sub-markets when a large distance exists between clusters of demand points. For example, urban areas in Texas such as Dallas, Houston, San Antonio, Austin, etc. are mutually exclusive. Consumers residing in Dallas will rarely patronize a McDonald outlet in San Antonio.

The contribution of our Drezner et al. (2016) paper was twofold: (1) dealing with multiple mutually exclusive sub-markets, and (2) discretizing the budget so that its allocation to each sub-market is not a continuous variable.

Suppose that the market can be partitioned into m mutually exclusive sub-markets. If we know the budget allocated to each sub-market, we may be able to find the optimal solution (where to locate new facilities and which existing facilities to expand) for each sub-market separately. This simplifies the formulation. However, the resulting problem is intractable as well because m variables representing the budget allocated to each sub-market are added to the formulation (in addition to the decision variables in each sub-market). In addition, a constraint that the sum of these individual budgets is equal to the available budget is added. A Lagrangian approach (adding a Lagrange multiplier for the constraint on the total budget and finding its value) is not applicable to this particular problem. The formula for the profit obtained in a sub-market as a function of the budget allocated to that sub-market is not an explicit expression.

Three objectives are investigated: (1) Maximizing firm's profit, (2) maximizing firm's return on investment, and (3) maximizing profit subject to a minimum acceptable return on investment. The last objective is similar in many ways to the threshold concept where the objective is to minimize the probability of falling short of a profit threshold or a cost overrun (Drezner et al. 2002b; Drezner and Drezner 2011). The first paper to introduce the threshold concept was Kataoka (1963) in the context of transportation problems. Frank (1966, 1967) considered a model of minimizing the probability that the cost function in the Weber or minimax problems (Love et al. 1988) on a network exceeds a given threshold. The threshold concept has been employed in financial circles as a form of insurance on a portfolio, either to protect the portfolio or to protect firm's minimum profit (Jacobs and Levy 1996).

12.5.1 Multiple-Market Competitive Location Solution Approach

There are m mutually exclusive sub-markets, each with given data about chain facilities, competitors, and demand points. A budget B is available for an investment in all m sub-markets. In order to diversify the investment, we can impose a maximum budget of B_0 in each of the sub-markets. The maximum budget can be different for different sub-markets. Suppose that the budget B is divided into K units, each unit is $\frac{B}{K}$ dollars. For example, we can use $K = 1000$ so that each unit is 0.1% of the total budget. Since all m sub-markets are mutually exclusive we can find the maximum profit for each individual sub-market by investing in sub-market $j = 1, \ldots, m$ a budget of $b_j = i\frac{B}{K}$ for some $0 \le i \le K$. If the amount to be invested in a particular sub-market cannot exceed B_0 dollars, then $\frac{i}{K}B \le B_0$ leading to $0 \le i \le K\frac{B_0}{B} = i_{\max}$. We assume that the maximum profit for a given investment in a given sub-market can be found by an optimal algorithm or, if necessary, by a good heuristic algorithm. The result is a matrix P of $i_{\max}$ rows and m columns. The element p_{ij} for $1 \le i \le i_{\max}$ and $1 \le j \le m$ in the matrix is the maximum profit obtained by investing $i\frac{B}{K}$ in sub-market j. For $i = 0$ the profit is zero. The problem is solved in two phases:

12.5.1.1 Phase 1: Calculating the Maximum Profit of a Sub-market for All Possible Budgets

Since each sub-market is independent of the other sub-markets, the maximum profit obtained in a sub-market for a given budget can be found by any existing competitive location solution method. There are also heuristic approaches proposed for such problems when a sub-market leads to a large problem. A problem consisting of 5000 demand points is too big for most published approaches. However, as we illustrate below, if such a problem can be divided to 20 sub-markets consisting between 100 and 400 demand points each, it is tractable for most solution approaches. The following are examples of competitive models and solution approaches that can be applied to find the maximum profit for a sub-market for a given budget allocated to that sub-market:

- Aboolian et al. (2007) solved the multiple facility location problem with a limited budget in discrete space within a given $\alpha\%$ of optimality.

- Plastria and Vanhaverbeke (2008) solved the problem defined by Aboolian et al. (2007) in a leader–follower modification. The leader–follower model is also termed the Stackelberg's equilibrium model (Sáiz et al. 2009; Stackelberg 1934).
- Fernandez et al. (2007) and Toth et al. (2009) solved the same problem as Aboolian et al. (2007) in a planar environment.
- Drezner and Drezner (2004) solved optimally the single facility problem based on the gravity formulation for a given budget (attractiveness).
- Drezner et al. (2012) solved optimally the multiple facilities problem with a limited budget in discrete space. New facilities can be constructed and existing facilities improved.
- Drezner et al. (2015) solved the leader–follower version of the formulation in Drezner et al. (2012). The competitor (follower) is expected to improve his facilities or build new ones in response to the leader's action. The objective is to maximize the leader's market share following the follower's action.

For $K = 1000$ (a parameter), a matrix P of up to 1001 rows corresponding to the possible investments, and m columns corresponding to the m sub-markets can be calculated by solving $1000m$ sub-problems. Of course, an investment of zero yields zero profit and need not be solved.

12.5.1.2 Phase 2: Calculating the Total Profit for All Markets Combined

Once the matrix P is available, the distribution of B among the m sub-markets can be found in two ways. One way is solving a binary linear program and the other way is by dynamic programming.

12.5.1.3 Binary Linear Programming Formulation

Let x_{ij} for $1 \leq i \leq K$ and $1 \leq j \leq m$ be a binary variable that is equal to 1 if a budget of $i\frac{B}{K}$ is invested in sub-market j and zero otherwise. The total profit is $\sum\limits_{i=1}^{i_{\max}} \sum\limits_{j=1}^{m} p_{ij}x_{ij}$. The total investment is $\frac{B}{K}\sum\limits_{i=1}^{i_{\max}} \sum\limits_{j=1}^{m} ix_{ij}$

$$\max \left\{ \sum_{i=1}^{i_{\max}} \sum_{j=1}^{m} p_{ij}x_{ij} \right\} \tag{12.23}$$

Subject to:

$$\sum_{i=1}^{i_{\max}} x_{ij} \leq 1 \qquad \text{for } j = 1, \ldots, m \tag{12.24}$$

$$\sum_{i=1}^{i_{\max}} \sum_{j=1}^{m} i x_{ij} \le K \tag{12.25}$$

$$x_{ij} \in \{0, 1\} \tag{12.26}$$

which is binary linear program with $i_{\max} \times m$ variables and $m + 1$ constraints. The constraint (12.24) guarantees that only one budget value is selected for each sub-market and if all $x_{ij} = 0$ for sub-market j, then no budget is allocated to sub-market j.

12.5.1.4 Dynamic Programming

Row zero is added to matrix P with zero values. The stages in the dynamic programming are the maximum profit for a budget $i\frac{B}{K}$ by investing only in the first j sub-markets. Let the matrix $Q = [q_{ij}]$ be the maximum profit obtained by investing a budget of $i\frac{B}{K}$ in the first j sub-markets. By definition $q_{i1} = p_{i1}$. For $2 \le j \le m$ the following recursive relationship holds:

$$q_{ij} = \max_{0 \le r \le i} \left\{ q_{r,j-1} + p_{i-r,j} \right\}.$$

The values q_{im} are the maximum possible profit for spending a total budget $i\frac{B}{K}$ in all sub-markets. Some sub-markets may be assigned no investment. One advantage of dynamic programming over the binary linear programming approach is that the maximum profit is obtained for each partial budget in one application of the dynamic programming, while K solutions of the binary linear programming are required. In addition, the maximum return on investment (ROI) is obtained for any partial budget by one application of dynamic programming.

12.5.1.5 Maximizing Profit Subject to a Minimum ROI

Finding the maximum profit subject to a minimum ROI can be done using the results obtained for maximizing the profit for a given budget. The ROI is the ratio between the profit and the investment (budget). It can be calculated for each investment value yielding a vector of ROI values. The maximum profit for a ROI greater than a certain value is found by calculating the maximum profit for all investments whose ROI exceeds the given value.

It can also be done by solving binary linear programs similar to the formulation presented in Sect. 12.5.1.3. Only one additional constraint is added to the binary linear programming formulation (12.23)–(12.26). By definition, the ROI is the ratio between the profit and the investment. Therefore,

$$ROI = \frac{\sum_{i=1}^{i_{\max}} \sum_{j=1}^{m} p_{ij} x_{ij}}{\frac{B}{K} \sum_{i=1}^{i_{\max}} \sum_{j=1}^{m} i x_{ij}}.$$

Suppose that a minimum ROI of α is required. $ROI \geq \alpha$ is equivalent to:

$$\sum_{i=1}^{i_{\max}} \sum_{j=1}^{m} \left\{ p_{ij} - i\alpha \frac{B}{K} \right\} x_{ij} \geq 0 \ . \tag{12.27}$$

Constraint (12.27), which is linear, is added to the formulation (12.23)–(12.26) leading to a binary linear program with $i_{\max} \times m$ variables and $m + 2$ constraints.

12.5.2 An Illustrative Multiple Markets Example

Once the maximum profit for a given investment in an individual sub-market is found, our general framework can be implemented. All the formulations and solution procedures described in Sect. 12.5.1.1 can be used for this purpose. In Drezner et al. (2016), we opted to apply the optimal branch and bound algorithm proposed in Drezner et al. (2012) for finding the maximum profit by investing a given budget in a single sub-market.

The networks selected for our sub-markets are the first 20 Beasley (1990a) networks designed for the evaluation of algorithms for solving p-median problems. Beasley (1990a) did not consider competitive models. Demand points, existing facilities, and potential locations for new facilities are located at the nodes of the network. Distances along links are measured in tenths of miles. These networks are easily available for testing other models as well. They can be used for future comparisons.

- 5000 demand points are located in 20 sub-markets. Each sub-market consists of between 100 and 400 demand points.
- 200 chain facilities and 200 competing facilities presently operate in these sub-markets.
- Each demand point has an available buying power to be spent at one's facilities or the competitors' facilities.
- For simplicity of presentation, each sub-market has a total buying power of \$150 million for a total of \$3 billion.
- A budget of up to \$100 million is available for improvements of existing facilities and construction of new ones. No more than \$30 million can be allocated to each sub-market.

- Existing facilities can be expanded and new facilities can be constructed at any node of the network.
- Each facility has a "circle of influence" defined by a radius of influence inside which they attract customers.
- For simplicity of presentation we assume that each existing facility has a radius of influence of 2 miles.
- The cost of expanding a facility is proportional to the increase in the area of its circle of influence. Expanding a facility from the existing radius of influence of 2 miles to a radius of influence of r miles costs $r^2 - 4$ million.
- Building a new facility with radius of influence r entails a $5 million setup cost plus a cost of r^2 million.

We want to determine which, if any, of the 200 existing facilities should be expanded and at which of the 4800 potential locations should new facilities be constructed to maximize profit. Maximizing the return on investment (ROI) is also considered, as well as maximizing profit subject to a minimum ROI value. The radii of the expanded and new facilities are variables, for a total of 5000 variables.

The branch and bound optimal algorithm (Drezner et al. 2012) and the dynamic programming procedure were programmed in Fortran using double-precision arithmetic. The programs were compiled by the Intel 11.1 Fortran Compiler and run, with no parallel processing, on a desktop with the Intel 870/i7 2.93 GHz CPU Quad processor and 8 GB memory. Only one thread was used.

The matrix P contains 6020 values (301 rows for a budget of zero and between $0.1 and $30 million, and 20 columns, one for each sub-market), each being the maximum profit for a given budget invested in a given sub-market. Note that an investment of $0 yields a profit of $0. All 6020 optimal solutions that are needed for the construction of matrix P were obtained in about 103 min of computing time.

Once the matrix P is found, obtaining the maximum profit for all partial budgets by solving binary linear programs using CPLEX 12.A took about 3 s for solving each of the 300 problems. The 300 results using dynamic programming were obtained in less than 1 s. Finding the maximum profit subject to a minimum ROI requirement by solving the binary linear program required about 1.6 s. Once the 300 results found by dynamic programming are available, the solution to the maximum profit for a minimum ROI is found by constructing a simple excel file.

In Table 12.1 we summarize the maximum possible profit along with the maximum return on investment (ROI) and the corresponding investments leading to these profits and ROIs. In five of the 20 sub-markets no profit is possible and no investment should be made. If unlimited budget is available and the best investment strategy is selected for each sub-market, then the total investment is $298.5 million leading to a profit of $198.4 million and ROI of 0.665.

Sub-market #20 was selected for depiction of the profit and the ROI as a function of the investment in that sub-market. In Fig. 12.2, these graphs are depicted. As reported in Table 12.1, the maximum profit of $26.884 million dollars is obtained for an investment of $24.1 million and a maximum ROI of 1.51 is achieved for an investment of $14.5 million.

Table 12.1 Individual sub-markets results

Sub-market	Demand points	Maximizing profit		Maximizing ROI	
		Million $ to invest	Profit in million $	Million $ to invest	Max ROI
1	100	0	0	0	0
2	100	0	0	0	0
3	100	0	0	0	0
4	100	0	0	0	0
5	100	0.5	0.056	0.5	0.112
6	200	0	0	0	0
7	200	24.1	1.741	0.5	0.237
8	200	0.9	0.210	0.9	0.233
9	200	1.7	0.541	1.3	0.397
10	200	29.7	3.561	25.3	0.140
11	300	22.5	18.558	13.7	0.961
12	300	26.3	12.131	2.8	0.781
13	300	24.1	20.592	17.2	1.161
14	300	29.5	6.762	1.3	0.430
15	300	28.5	18.956	19.1	0.844
16	400	24.4	22.230	11.3	1.517
17	400	22.1	22.499	5.7	1.582
18	400	22.1	25.716	14.5	1.476
19	400	18.0	18.010	9.7	1.228
20	400	24.1	26.884	14.5	1.510

In Fig. 12.3, we depict the profit and ROI for the total investment in all 20 sub-markets. These values were obtained using dynamic programming. The profit increases as a function of total investment. However, ROI is quite erratic. ROI reaches the maximum when $5.7 million is invested in sub-market #17 and no investment made in other sub-markets.

In Fig. 12.4, the maximum profit for a minimum ROI value is plotted for an investment of up to $100 million. As expected, when higher minimum ROI is required the maximum profit declines.

12.6 Conclusions

This chapter summarized four papers on competitive location: Drezner et al. (2011, 2012, 2015, 2016). All four papers utilize the radius of influence and are based on an assumption of equal division of buying power among facilities whose radius of influence captures that power.

We presented efficient methods for locating multiple new, and expanding existing, facilities in such a competitive environment. We also presented a leader–

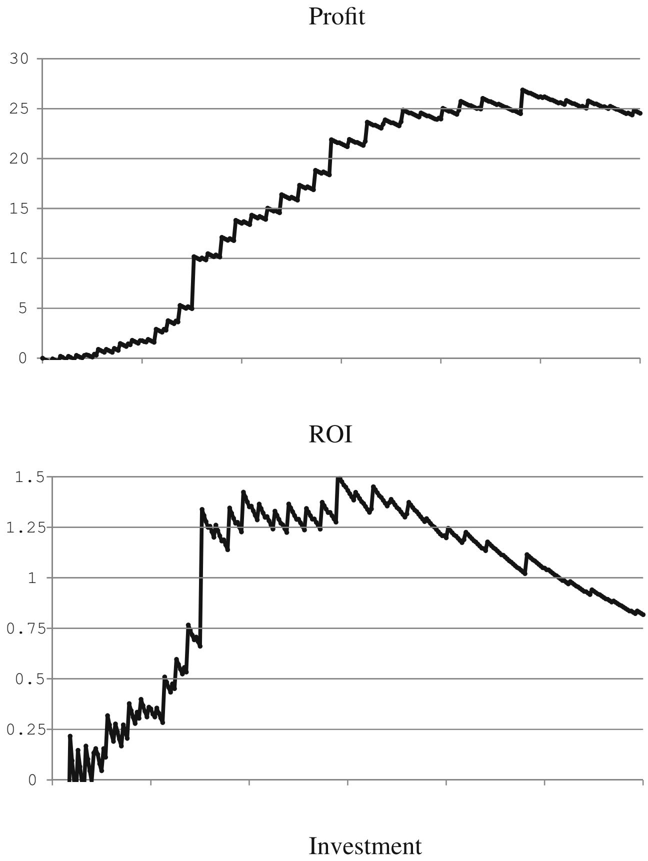


Fig. 12.2 Profit and ROI as a function of the investment in sub-market #20

follower model in which initial actions of the leader (locate new and/or improve existing facilities) are countered by follower's response, along with a solution method. Finally, we discussed a multiple disjoint markets problem with real-world objectives and presented efficient solution techniques.

All these methods have a common objective: the maximization of the market share. Since profit is (usually) a monotonically increasing function of market share captured, this objective is associated with maximizing profit.

The original idea of locating new facilities based on their radii of influence called for an extension allowing expanding existing facilities in addition to locating new ones. This concept, however, was considered in a static context in which a decision to locate new and expand existing facilities was based on the "ceteris paribus" assumption from the classical economics, i.e., no reaction from the competitors. This assumption is released in the leader–follower version of our model, in which the leader's decision takes into account the actions of competitors,

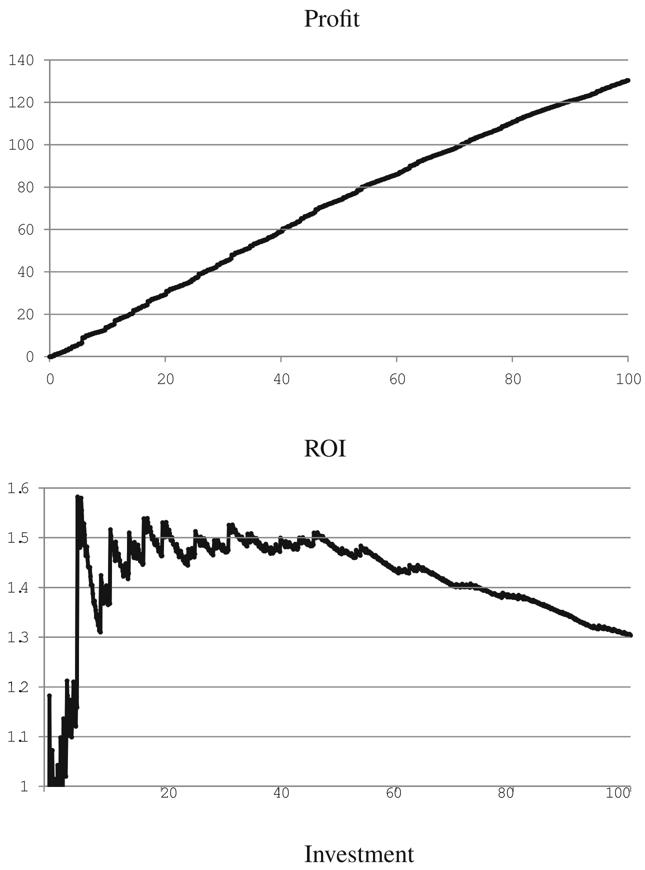


Fig. 12.3 Profit and ROI as a function of the total investment

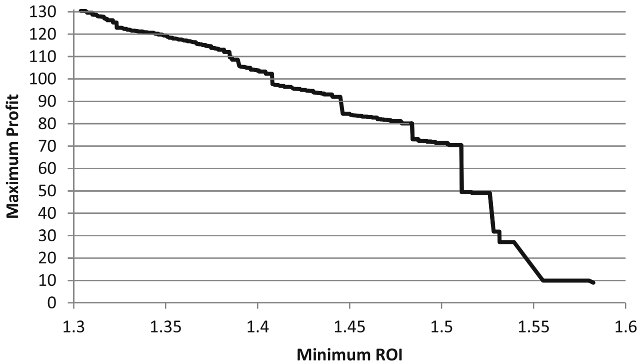


Fig. 12.4 Maximum profit subject to minimum ROI

directly following the leader's decision. Like in a game of chess, the leader must carefully consider each move (locating new and expanding existing facilities) in terms of the competitors' counter-moves. Adding this new dimension to the problem significantly increased the difficulty of finding a solution but it also made our model more realistic.

Even in its original formulation, the competitive location problem is combinatorially explosive as its complexity increases much faster than the number of facilities to be located or expanded. The practical considerations such as the concentration of demand points in urban areas and the limited sphere of influence for most types of facilities lead us to the model in which multiple disjoint markets are identified and considered as separate environments for locating new and expanding existing facilities. This new approach allows tackling large practical competitive location problems which would be too difficult or impossible to solve without splitting them into disjoint sub-markets.

Future research involving our competitive location model might consider downgrading and closing existing facilities (together with opening and upgrading) as a new strategy. Studying the effects of including this new strategy in the leader–follower model is another interesting research avenue to explore. Also, additional moves in the competitive game (e.g., leader's move, follower's response, leader's response) could be investigated, however, on a much smaller scale. Other game-theoretical approaches such as forming coalitions with some competitors could be investigated. Lastly, a gradual cover extension of our original model (with multiple radii of influence) could be considered.

References

Aboolian, R., Berman, O., & Krass, D. (2007). Competitive facility location and design problem. *European Journal of Operations Research, 182*, 40–62.

Alp, O., Drezner, Z., & Erkut, E. (2003). An efficient genetic algorithm for the *p*-median problem. *Annals of Operations Research, 122*, 21–42.

Barros, A. I. (1998). *Discrete and fractional programming techniques for location models*. Boston: Springer.

Beasley, J. E. (1990a). OR-library—Distributing test problems by electronic mail. *Journal of the Operational Research Society, 41*, 1069–1072. Also available at http://people.brunel.ac.uk/~mastjjb/jeb/orlib/pmedinfo.html

Beasley, J. E. (1990b). Or-library: Distributing test problems by electronic mail. *Journal of the Operational Research Society, 41*(11), 1069–1072.

Beaumont, J. R. (1991). GIS and market analysis. In D. J. Maguire, M. Goodchild, & D. Rhind (Eds.), *Geographical information systems: Principles and applications* (pp. 139–151). Harlow: Longman Scientific.

Berman, O. (1994). The *p*-maximal cover - *p*-Partial center problem on networks. *European Journal of Operational Research, 72*, 432–442.

Berman, O., & Drezner, Z. (2008). A new formulation for the conditional *p*-median and *p*-center problems. *Operations Research Letters, 36*, 481–483.

Berman, O., Drezner, T., Drezner, Z., & Krass, D. (2009a). Modeling competitive facility location problems: New approaches and results. In M. Oskoorouchi (Ed.), *Tutorials in operations research* (pp. 156–181). San Diego: INFORMS.

Berman, O., Drezner, Z., Krass, D., & Wesolowsky, G. O. (2009b). The variable radius covering problem. *European Journal of Operational Research, 196*, 516–525.

Canovas, L., & Pelegrin, B. (1992). Improving some heuristic algorithms for the rectangular *p*-cover problem. In J. A. Moreno-Perez (Ed.), *Proceedings of the VI Meeting of the EURO Working Group on Locational Analysis* (pp. 23–31). Tenerife: Universidad de La Laguna.

Charnes, A., & Cooper, W. W. (1962). Programming with linear fractional functionals. *Naval Research Logistics (NRL), 9*(3–4), 181–186.

Chen, D. Z., Daescu, O., Dai, Y., Katoh, N., Wu, X., & Xu, J. (2005). Efficient algorithms and implementations for optimizing the sum of linear fractional functions, with applications. *Journal of Combinatorial Optimization, 9*(1), 69–90.

Church, R. L., & ReVelle, C. S. (1974). The maximal covering location problem. *Papers of the Regional Science Association, 32*, 101–118.

Current, J., Daskin, M., & Schilling, D. (2002). Discrete network location models. In Z. Drezner & H. W. Hamacher (Eds.), *Facility location: Applications and theory* (pp. 81–118). Berlin: Springer.

Daskin, M. S. (1995). *Network and discrete location: Models, algorithms, and applications*. New York: Wiley.

Drezner, T. (1994a). Locating a single new facility among existing unequally attractive facilities. *Journal of Regional Science, 34*, 237–252.

Drezner, T. (1994b). Optimal continuous location of a retail facility, facility attractiveness, and market share: An interactive model. *Journal of Retailing, 70*, 49–64.

Drezner, T. (1995). Competitive facility location in the plane. In Z. Drezner (Ed.), *Facility location: A survey of applications and methods* (pp. 285–300). New York: Springer.

Drezner, T. (1998). Location of multiple retail facilities with limited budget constraints—In continuous space. *Journal of Retailing and Consumer Services, 5*, 173–184.

Drezner, T. (2006). Derived attractiveness of shopping malls. *IMA Journal of Management Mathematics, 17*, 349–358.

Drezner, T., & Drezner, Z. (1996). Competitive facilities: Market share and location with random utility. *Journal of Regional Science, 36*, 1–15.

Drezner, T., & Drezner, Z. (1998). Facility location in anticipation of future competition. *Location Science, 6*, 155–173.

Drezner, T., & Drezner, Z. (2004). Finding the optimal solution to the Huff competitive location model. *Computational Management Science, 1*, 193–208.

Drezner, T., & Drezner, Z. (2008). Lost demand in a competitive environment. *Journal of the Operational Research Society, 59*, 362–371.

Drezner, T., & Drezner, Z. (2011). The Weber location problem: The threshold objective. *INFOR: Information Systems and Operational Research, 49*, 212–220.

Drezner, T., Drezner, Z., & Kalczynski, P. (2011). A cover-based competitive location model. *Journal of the Operational Research Society, 62*, 100–113.

Drezner, T., Drezner, Z., & Kalczynski, P. (2012). Strategic competitive location: Improving existing and establishing new facilities. *Journal of the Operational Research Society, 63*, 1720–1730.

Drezner, T., Drezner, Z., & Kalczynski, P. (2015). A leader-follower model for discrete competitive facility location. *Computers & Operations Research, 64*, 51–59.

Drezner, T., Drezner, Z., & Kalczynski, P. (2016). The multiple markets competitive location problem. *Kybernetes, 45*, 854–865.

Drezner, T., Drezner, Z., & Salhi, S. (2002a). Solving the multiple competitive facilities location problem. *European Journal of Operational Research, 142*, 138–151.

Drezner, T., Drezner, Z., & Shiode, S. (2002b). A threshold satisfying competitive location model. *Journal of Regional Science, 42*, 287–299.

Drezner, Z. (1981). On a modified one-center problem. *Management Science, 27*, 848–851.

Drezner, Z. (1982). Competitive location strategies for two facilities. *Regional Science and Urban Economics, 12*, 485–493.

Drezner, Z. (1986). The p-cover problem. *European Journal of Operational Research, 26*, 312–313.

Drezner, Z., Hahn, P. M., & Taillard, É. D. (2005). Recent advances for the quadratic assignment problem with special emphasis on instances that are difficult for meta-heuristic methods. *Annals of Operations Research, 139*, 65–94.

Drezner, Z., Suzuki, A., & Drezner, T. (2007). Locating multiple facilities in a planar competitive environment. *Journal of the Operations Research Society of Japan, 50*, 1001–1014.

Drezner, Z., Wesolowsky, G. O., & Drezner, T. (1998). On the logit approach to competitive facility location. *Journal of Regional Science, 38*, 313–327.

Falk, J. E., & Palocsay, S. W. (1992). Optimizing the sum of linear fractional functions. In *Recent advances in global optimization* (pp. 221–258). Princeton: Princeton University Press.

Fernandez, J., Pelegrin, B., Plastria, F., & Toth, B. (2007). Solving a Huff-like competitive location and design model for profit maximization in the plane. *European Journal of Operational Research, 179*, 1274–1287.

Frank, H. (1966). Optimum location on a graph with probabilistic demands. *Operations Research, 14*, 409–421.

Frank, H. (1967). Optimum location on a graph with correlated normal demands. *Operations Research, 15*, 552–557.

Freund, R. W., & Jarre, F. (2001). Solving the sum-of-ratios problem by an interior-point method. *Journal of Global Optimization, 19*(1), 83–102.

Gao, Z., Wu, J., & Sun, H. (2005). Solution algorithm for the bi-level discrete network design problem. *Transportation Research Part B: Methodological, 39*, 479–495.

Ghosh, A., & Rushton, G. (1987). *Spatial analysis and location-allocation models*. New York: Van Nostrand Reinhold Company.

Glover, F. (1977). Heuristics for integer programming using surrogate constraints. *Decision Sciences, 8*, 156–166.

Glover, F. (1986). Future paths for integer programming and links to artificial intelligence. *Computers & Operations Research, 13*, 533–549.

Glover, F., & Laguna, M. (1997). *Tabu search*. Boston: Kluwer Academic Publishers.

Hakimi, S. L. (1983). On locating new facilities in a competitive environment. *European Journal of Operational Research, 12*, 29–35.

Hakimi, S. L. (1986). p-Median theorems for competitive location. *Annals of Operations Research, 6*, 77–98.

Hakimi, S. L. (1990). Locations with spatial interactions: Competitive locations and games. In P. B. Mirchandani & R. L. Francis (Eds.), *Discrete location theory* (pp. 439–478). New York: Wiley-Interscience.

Hardy, G., Littlewood, J. E., & Pólya, G. (1952). *Inequalities* (2nd ed.). New York: Cambridge University Press. First published in 1934.

Hotelling, H. (1929). Stability in competition. *Economic Journal, 39*, 41–57.

Huff, D. L. (1964). Defining and estimating a trade area. *Journal of Marketing, 28*, 34–38.

Huff, D. L. (1966). A programmed solution for approximating an optimum retail location. *Land Economics, 42*, 293–303.

Jacobs, B. I., & Levy, K. N. (1996). Residual risk: How much is too much? *Journal of Portfolio Management, 22*, 10–16.

Kataoka, S. (1963). A stochastic programming model. *Econometrica, 31*, 181–196.

Kirkpatrick, S., Gelat, C. D., & Vecchi, M. P. (1983). Optimization by simulated annealing. *Science, 220*, 671–680.

Küçükaydın, H., Aras, N., & Altınel, I. (2012). A leader–follower game in competitive facility location. *Computers & Operations Research, 39*, 437–448.

Küçükaydın, H., Aras, N., & Altınel, İ. K. (2011). A discrete competitive facility location model with variable attractiveness. *Journal of the Operational Research Society, 62*, 1726–1741.

Leonardi, G., & Tadei, R. (1984). Random utility demand models and service location. *Regional Science and Urban Economics, 14*, 399–431.

Love, R. F., Morris, J. G., & Wesolowsky, G. O. (1988). *Facilities location: Models & methods*. New York: North Holland.

Megiddo, N., Zemel, E., & Hakimi, S. L. (1983). The maximum coverage location problems. *SIAM Journal of Algebraic and Discrete Methods, 4*, 253–261.

Nesterov, Y. E., & Nemirovskii, A. (1995). An interior-point method for generalized linear-fractional programming. *Mathematical Programming, 69*(1–3), 177–204.

Plastria, F. (2002). Continuous covering location problems. In Z. Drezner & H. W. Hamacher (Eds.), *Facility location: Applications and theory* (pp. 37–79). Berlin: Springer.

Plastria, F., & Carrizosa, E. (2004). Optimal location and design of a competitive facility. *Mathematical Programming, 100*, 247–265.

Plastria, F., & Vanhaverbeke, L. (2008). Discrete models for competitive location with foresight. *Computers & Operations Research, 35*, 683–700.

Redondo, J. L., Arrondo, A., Fernández, J., García, I., & Ortigosa, P. M. (2013). A two-level evolutionary algorithm for solving the facility location and design (1|1)-centroid problem on the plane with variable demand. *Journal of Global Optimization, 56*, 983–1005.

Redondo, J. L., Fernandez, J., Garcia, I., & Ortigosa, P. M. (2009). Solving the multiple competitive facilities location and design problem on the plane. *Evolutionary Computation, 17*, 21–53.

Redondo, J. L., Fernández, J., García, I., & Ortigosa, P. M. (2010). Heuristics for the facility location and design (1|1)-centroid problem on the plane. *Computational Optimization and Applications, 45*, 111–141.

Reilly, W. J. (1931). *The law of retail gravitation*. New York: Knickerbocker Press.

ReVelle, C. (1986). The maximum capture or sphere of influence problem: Hotelling revisited on a network. *Journal of Regional Science, 26*, 343–357.

ReVelle, C., Toregas, C., & Falkson, L. (1976). Applications of the location set covering problem. *Geographical Analysis, 8*, 65–76.

Saidani, N., Chu, F., & Chen, H. (2012). Competitive facility location and design with reactions of competitors already in the market. *European Journal of Operational Research, 219*, 9–17.

Sáiz, M. E., Hendrix, E. M., Fernández, J., & Pelegrín, B. (2009). On a branch-and-bound approach for a Huff-like Stackelberg location problem. *OR Spectrum, 31*, 679–705.

Schilling, D. A., Vaidyanathan, J., & Barkhi, R. (1993). A review of covering problems in facility location. *Location Science, 1*, 25–55.

Serra, D., & ReVelle, C. (1995). Competitive location in discrete space. In Z. Drezner (Ed.), *Facility location: A survey of applications and methods* (pp. 367–386). New York: Springer.

Stackelberg, H. V. (1934). *Marktform und Gleichgewicht*. Vienna: Springer.

Sun, H., Gao, Z., & Wu, J. (2008). A bi-level programming model and solution algorithm for the location of logistics distribution centers. *Applied Mathematical Modelling, 32*, 610–616.

Teitz, M. B., & Bart, P. (1968). Heuristic methods for estimating the generalized vertex median of a weighted graph. *Operations Research, 16*, 955–961.

Toppen, F., & Wapenaar, H. (1994). *GIS in business: Tools for marketing analysis*. The Hague: EGIS Foundation.

Toth, B., Fernandez, J., Pelegrin, B., & Plastria, F. (2009). Sequential versus simultaneous approach in the location and design of two new facilities using planar Huff-like models. *Computers & Operations Research, 36*, 1393–1405.

Watson-Gandy, C. (1982). Heuristic procedures for the m-partial cover problem on a plane. *European Journal of Operational Research, 11*, 149–157.

Chapter 13
The Mean-Value-at-Risk Median Problem on a Network with Random Demand Weights

Chunlin Xin and Jiamin Wang

Dr. Zvi Drezner's research career has touched on many areas of location analysis. We devote the first part of this chapter to summarizing Zvi's vast contributions to the studies of the minimax and the maximum facility location problems. His relevant publications are grouped in terms of the characteristics of the problems investigated, including space, the number of facilities to locate, and completeness of information. In particular, we provide an overview of Zvi's work in the deterministic planar minimax problems. The second part of the chapter is our own paper on a network median problem when demand weights are independent random variables. The objective of the model proposed is to locate a single facility so as to minimize the expected total demand-weighted distance subject to a constraint on the value-at-risk (VaR). The study integrates the expectation criterion with the VaR measure and links different median models with random demand weights. Methods are suggested to identify dominant points for the optimal solution. An algorithm is developed for solving the problem.

C. Xin
School of Economics and Management, Beijing University of Chemical Technology, Beijing, China

J. Wang (✉)
College of Management, Long Island University Post, Brookville, NY, USA
e-mail: jiamin.wang@liu.edu

H. A. Eiselt, V. Marianov (eds.), *Contributions to Location Analysis*, International Series in Operations Research & Management Science 281,
https://doi.org/10.1007/978-3-030-19111-5_13

13.1 Zvi Drezner's Contributions to the Minimax Location Problems

Dr. Zvi Drezner is well known for his important contributions to various aspects of the location theory. In this section, we attempt to summarize his studies in the minimax and maximum facility location problems. The overview shows both the width and the depth of his pioneer work in the field. The remainder of the chapter is our own study on a median (minisum) location problem with random demand weights. The study integrates the expectation criterion with the value-at-risk (VaR) measure and links different probabilistic median problems. It is noted that our study does intersect with Zvi's interest in stochastic location analysis. We are deeply grateful to Drs. H. A. Eiselt and Vladimir Marianov, the editors of the present academic Festschrift, for giving us an opportunity to honor Zvi's great accomplishments in his research.

The minimax objective function seeks to improve the least payoff as much as possible. It is thus widely adopted to model "fairness" in facility location analysis (Eiselt and Laporte 1995), resulting in a large family of the minimax facility location problems. The classical minimax problem, also referred to as the center problem, is to locate facilities so that the maximum (weighted) distance between the farthest demand and its closest facility is minimized. This model can be applied to site public facilities such as fire stations and hospitals. The reader is referred to the excellent reviews of Drezner (2011) and Tansel (2011) on the minimax facility location problem in the plane and on a network, respectively.

The minimax criterion is appropriate for locating desirable facilities. But for undesirable facilities such as prisons, nuclear power plants, and airports, it makes sense to apply the maximin objective function so as to keep the demand away from the facilities as far as possible (Eiselt and Laporte 1995). An overview of the earlier work on the maximin facility location problem was given by Melachrinoudis (2011).

We now summarize Dr. Zvi Drezner's vast contributions to the studies of the minimax and the maximin problems. Tables 13.1 and 13.2 group a number of Zvi's relevant publications in terms of the characteristics of the problems investigated, including space, the number of facilities to locate, and completeness of information (a deterministic setting or a probabilistic setting). Note that a few publications appear in two groups, which means that they possess the characteristics of both groups. For instance, Drezner and Wesolowsky (2000) study both the single-facility minimax and maximin problems in the plane where the demand points of interest are aggregated into groups. The list of the publications in the tables is by no means complete. However, we feel that it is sufficient to evidence the width of his work in the field, which spans from continuous space to discrete space, from a single facility to multiple facilities, from the deterministic setting to the probabilistic setting, as well as from desirable facilities to undesirable ones.

We next present an overview of Zvi's studies of the deterministic minimax problems in the plane as this is the area that the majority of his publications focus on.

Table 13.1 Publications on the continuous minimax and maximin problems

	One facility	Multiple facilities
Planar	*Deterministic minimax*	*Deterministic minimax*
	Drezner et al. (1986)	Drezner and Wesolowsky (1978)
	Drezner and Shelah (1987)	Drezner (1984a)
	Drezner and Wesolowsky (1989)	Drezner (1984b)
	Drezner et al. (1991)	Drezner (1989)
	Drezner and Wesolowsky (1995)	Suzuki and Drezner (1996)
	Drezner and Wesolowsky (2000)	
	Berman et al. (2007)	
	Suzuki and Drezner (2009)	
	Probabilistic minimax	*Probabilistic minimax*
	Berman et al. (2003c)	Drezner (1987)
	Berman et al. (2003b)	
	Deterministic maximin	*Deterministic maximin*
	Drezner and Wesolowsky (1980)	Drezner and Wesolowsky (1985)
	Drezner and Wesolowsky (1996)	Welch et al. (2006)
	Drezner and Wesolowsky (2000)	
$R^d(d > 2)$	*Deterministic minimax*	*Deterministic minimax*
	Drezner and Gavish (1985)	Drezner (1991)

Table 13.2 Publications on the network minimax and maximin problems

One facility	Multiple facilities
Deterministic minimax	*Deterministic minimax*
Berman et al. (2001)	Berman et al. (2001)
Berman et al. (2003b)	Berman et al. (2005)
Berman et al. (2007)	Berman and Drezner (2008)
	Berman et al. (2009)
Probabilistic minimax	*Probabilistic minimax*
Berman and Drezner (2003)	Aboolian et al. (2009)
Berman et al. (2003a)	
Deterministic maximin	
Berman and Drezner (2000)	
Berman et al. (2003a)	

It appears that Zvi's earlier work was mainly concerned with the algorithms for solving the classical minimax problem. In Drezner and Shelah (1987), it was proven that the complexity of the Elzinga and Hearn (1972) algorithm for the Euclidean unweighted single-facility minimax problem is $O(n^2)$, where n is the number of demand points. Let $m > 1$ be the number of new facilities to locate. Drezner and Wesolowsky (1978) proposed that the optimal solution to the m-facility weighted minimax problem with the general l_p norm distances could be found by solving a set of ordinary differential equations via numerical integration.

Drezner (1984a,b) considered the Euclidean m-facility minimax problem and suggested a solution approach to partition the demand points into m distinctive subsets, and then find the optimal location of a facility for each subset. In Drezner (1984a), an exact solution method was developed that generates and evaluates all two partitions of the demand points when $m = 2$. In Drezner (1984b), two heuristics were suggested with an attempt to improve the demand point partitions iteratively for $m \geq 2$. A polynomial-time exact solution approach was also proposed that partitions all demand points into m disjoint subsets such that the single-facility minimax problem for each subset has an optimal objective function value no greater than the known upper bound to the original problem. The partitioning results are then used to update the upper bound iteratively, while the procedure terminates when the upper bound cannot be reduced further.

Zvi also studied problems that can be viewed as *direct extensions* of the classical minimax problem. Drezner et al. (1986) investigated the single-facility minimax-min location problem with the objective to minimize the range of the weighted distances. The proposed location analysis model can be applied in the public sector for which smaller variations in accessibility of the potential user population shall be preferred. Drezner (1989) considered the conditional multi-facility minimax problem in which m new facilities are to be sited in additional to the q existing facilities.

The distance function is not conventional in some other studies. Drezner and Wesolowsky (1989) and Drezner et al. (1991) considered the single-facility minimax problem with the rectilinear and Euclidean distances when the distance between any two points is assumed to be, respectively, asymmetric and a constant once it reaches some threshold. As the authors noted, the former model is valid in the rush hour traffic, or the traffic on an inclined surface, while the latter is applicable to exclude the customers who appear too far away from consideration. In Drezner and Wesolowsky (1995), the single-facility minimax problem on alternating one-way routes or streets was studied. The model presented in their study can be applied to achieve a more efficient traffic flow. The authors derived the rectilinear distance function between any two points on a one-way grid and developed an efficient algorithm for solving the problem.

The minimax problem with groups of demand points or continuous area demand as well as the transfer point location problem can be regarded as *indirect extensions* of the classical minimax problem. Aggregating the demand points into groups, Drezner and Wesolowsky (2000) generalized the single-facility minimax problem as well as the maximin problem by using the group-distance between the facility and a cluster of demand points instead of the traditional point-to-point distance. An evident application of the model is to deal with demand points that are widely dispersed in an area. Three group-distance measures were suggested and various model formulations were developed.

It is common in the literature to assume that demand originates from a finite number of demand points. But demand from a continuous area is more appropriate for mobile demand such as cell phone or wireless coverage. Suzuki and Drezner (1996) considered the m-minimax problem with area demand. It was suggested that the unweighted version of the problem is equivalent to covering the demand area by a pre-chosen number of circles with the smallest possible radius. The authors developed a heuristic solution procedure using the Voronoi diagram (Sugihara and Iri 1992, 1994). In the procedure, after a set of m centers is randomly generated in the area, the following two steps are repeated: a Voronoi diagram based on the centers is constructed, and then the centers are updated for the Voronoi diagram. Because the heuristic converges very slowly, it had to be terminated prematurely by the authors and was replaced by a non-linear convex programming model to improve the solution. In Suzuki and Drezner (2009), a solution method also based on the Voronoi diagram was developed, which iteratively changes the m centers by solving a linear programming model. It was reported that this procedure converges very quickly and therefore is more efficient.

Berman et al. (2007) discussed the transfer point location problem in the plane and on a network, where the trip from each demand point to the facility (such as a hospital) has to go through the transfer point (such as a helicopter pad). Assuming that the traveling speed is higher between the transfer point and the facility, they developed models to optimize the location of the transfer point for a given facility location. In the minimax model considered, the objective is to minimize the longest travel time it takes to access the facility from the demand points via the transfer point.

Dr. Zvi Drezner is a prolific and challenging intellectual with deep and wide-ranging interests. This short overview highlights his notable contributions to the studies of the minimax and the maximin facility location problems only. We note that his work has not only moved location analysis to an advanced level, but also had a transformative effect on the way scholars in the field conduct their studies. At the age of 75, Zvi remains active in his intellectual exploration. We are delighted to expect his continuing contributions to location analysis.

13.2 The Mean-Value-at-Risk Median Problem on a Network with Random Demand Weights

The median problem (Hakimi 1964, 1965) is classical in location theory. In the deterministic median problem, a weight is assigned to each demand point and the objective is to locate facilities so as to minimize the total demand-weighted distances between the facilities and the customers. The problem is equivalent to minimizing the average distance from a customer to a closest facility. The median model can be applied to site public and business facilities such as schools, libraries, hospitals, and warehouses (Christofides 1975).

Stochastic location analysis considers the impact of uncertainties in customer demand streams, travel times, and/or service times on location decisions. The reader is referred to the reviews written by Snyder (2006) and Correia and Saldanha da Gama (2015) on location analysis under uncertainty. In most of the stochastic location analysis literature the decision maker is assumed to be risk neutral. Take the median problem with random demand weights as an example. The expected total demand-weighted distance is minimized, but the risk, i.e., the possibility of the actual total demand-weighted distance exceeding the expected value is disregarded under the risk-neutral assumption.

There are studies in which risk was taken into consideration. For the median problem, a probability measure was adopted by Frank (1966, 1967) and Berman and Wang (2006) to evaluate the likelihood that the total demand-weighted distance exceeds some pre-selected level. Variance was used in Frank (1966) to minimize the volatility of the total demand-weighted distance. Wang (2007) and Wagner et al. (2009) presented models to minimize the value-at-risk (VaR) of the total demand-weighted distance. Daskin et al. (1997) and Chen et al. (2006), respectively, developed models to minimize the VaR and the conditional value-at-risk (CVaR) of the regret of the median problem decision-making due to random demand weights and travel time under different scenarios.

In the above studies the risk measure in use served as the objective function, while the expectation of the total demand-weighted distance, a long-term performance measure, was not considered. Though the mean-variance framework (Markowitz 1959) was applied to solve the median problem by Jucker and Hodder (1976), Hodder and Jucker (1985), and Berman et al. (2016), using variance or standard deviation to measure risk has been criticized as it takes into account both favorable and unfavorable deviations of a random distribution (Nawrocki 1999).

VaR is a quantile of a random loss distribution (Pflug 2000) for a given confidence level. It is commonly believed to be superior to variance or standard deviation as it focuses on the unfavorable tail of a distribution. In the current study, we develop a mean-VaR model for the median problem on a network to balance a short distance on average and the risk due to random demand weights. Note that VaR is not a coherent risk measure because it does not always obey subadditivity (Rockfella 2007). Lack of subadditivity violates the portfolio diversification theory, which states that an investor can reduce risk by including multiple assets in a portfolio (Wagner and Lau 1971), because the VaR of a portfolio can be higher than the sum of the individual assets' VaRs. But for the network median problem, this simply implies that combining two sub-networks into one would not reduce risk, which does not violate any principle. Therefore, VaR is an acceptable risk measure for the median problem.

In Sect. 13.3, we introduce the mean-VaR model and motivate the study by comparing it with the mean-variance model. In Sects. 13.4 and 13.5, we discuss solving the mean-VaR model for continuous and discrete random demand weights. In Sect. 13.6, an illustrative example is analyzed. Finally, we summarize the study and discuss future research directions.

13.3 Problem Statement

Let $G = (N, L)$ be an undirected network with a set of nodes N ($|N| = n$) and a set of links L ($|L| = l$). Denote by l_{ij} the length of link $(i, j) \in L$ and by $d(h, x)$ the shortest distance between node h and some point $x \in G$. When there is no ambiguity, the same notation $x \in [0, l_{ij}]$ represents both a point on link (i, j) and the distance of the point from node i (in order to uniquely locate a point, we specify the end-point of the link with the smaller node number to be node i and hence the other end-point to be node j).

We assume that demand generates from the nodes only. The demand weight $W_h d(h, x)$ represents the total demand-weighted distance from the demand nodes to point x. Denote by $\mathbf{E}$ the expectation operator. The *expected median* of the network, x_E, is a point such that $\mathbf{E}(D(x_E)) \leq \mathbf{E}(D(x))$ holds for any point $x \in G$. $R_E = \mathbf{E}(D(x_E))$ is referred to as the *expected median length* of the network.

It is natural to investigate the likelihood that the total demand-weighted distance between the facility and the demand points does not exceed R_E or some threshold value as it assesses the risk of a location decision. The *maximum probability median*, denoted by $x_{P,T}$, is a point such that $P(D(x_{P,T}) \leq T) \geq P(D(x) \leq T)$ is true for any point $x \in G$, where the pre-specified threshold value T can be regarded as an aspiration level that is desirable to achieve.

Given point x, we can define the value-at-risk (VaR) of the total demand-weighted distance at a given confidence level $\beta \geq 0.5$ as follows:

$$t_\beta(x) = \min\{T | P(D(x) \leq T) \geq \beta\}.$$

$t_\beta(x)$ can be interpreted as the lowest target level for the total demand-weighted distance to fall below with a confidence level of at least β. The *minimum VaR* median, denoted by $x_{\beta,\text{VaR}}$ is a point such that $t_\beta(x_{\beta,\text{VaR}}) \leq t_\beta(x)$ holds at any point $x \in G$ (Wang 2007).

In the expected median problem, the maximum probability median problem, and the minimum VaR median problems presented above, either the expected value or a risk measure is considered only. Here we propose a model, namely the mean-VaR median problem, which integrates both the expected value and the VaR for given T and $\beta \in [0, 1]$:

$$\begin{aligned} R_{\beta,T} &= \min_{x \in G} \mathbf{E}(D(x)) \\ \text{s.t.} & \\ P(D(x) &\leq T) \geq \beta. \end{aligned} \tag{13.1}$$

If the weight W_h is constant for every node h, then the model (13.1) reduces to

$$\min_{x \in G} D(x)$$
$$\text{s.t.}$$
$$D(x) \leq T.$$

It follows that the model above can be regarded as a constrained deterministic median problem, where the median of G is optimal if and only if the median length is no longer than T, and there is no feasible solution otherwise. We can thus interpret the model (13.1) as an immediate extension of the deterministic median problem.

Note that the model (13.1) is a constrained optimization problem. The constraint is equivalent to $t_\beta(x) \leq T$. Hence, the model is feasible for given β and T if the minimum VaR $t_\beta(x_{\beta,\text{VaR}}) \leq T$ holds. In the sequel, we assume that the model is feasible. This assumption is not so restrictive as parameters β and T can be adjusted to enforce feasibility.

Let $x_{\beta,T}$ be optimal to the model (13.1). We call $x_{\beta,T}$ the mean-VaR median and $R_{\beta,T}$ the mean-VaR median length of the network. By definition, $t_\beta(x)$ is non-decreasing in β. It follows that increasing β or decreasing T will make the constraint in the model (13.1) more stringent and therefore $R_{\beta,T}$ is non-decreasing in β but non-increasing in T.

It is evident that the mean-VaR median length is no shorter than the expected median length, i.e., $R_{\beta,T} \geq R_E$, while the equality holds, i.e., x_E is optimal if and only if $t_\beta(x_E) \leq T$ (if there are multiple solutions to the expected median problem, then we let x_E denote the expected median with the smallest VaR). $t_\beta(x_E)$ can be evaluated by definition (the reader may refer to the procedure developed in Wang (2007) when demand weights follow discrete probability distributions or the normal distribution). The lemma below presents a sufficient (but not necessary) condition for x_E to be optimal without calculating $t_\beta(x_E)$.

Lemma 13.1 *x_E is optimal to the model (13.1) (i) if $\beta = 0$ or $T \geq \sum_{h \in N} b_h d(h, x_E)$ when the demand weights follow independent continuous distributions; or (ii) if $\beta < P(W_1 = a_1, \cdots, W_n = a_n)$ or $T > \sum_{h \in N} b_h d(h, x_E)$ when the demand weights follow independent discrete distributions, where a_h and b_h are the lower and upper bounds (or the smallest and largest realizations) of the random weight W_h.*

Conceptually, we can conclude that: (i) given β, the mean-VaR model is infeasible if T is too small, or the expected median is also the mean-VaR median if T is sufficiently large; (ii) given T, the mean-VaR model is infeasible if β is too large, or the expected median is also the mean-VaR median if β is sufficiently small.

Next, we will connect the mean-VaR median $x_{\beta,T}$ with the minimum VaR median $x_{\beta,\text{VaR}}$ and the maximum probability median $x_{P,T}$. If there are ties, let $x_{\beta,\text{VaR}}$ and $x_{P,T}$ denote, respectively, the minimum VaR median and the maximum probability median with the lowest expected total demand-weighted distance. The next two lemmas are easy to prove.

Lemma 13.2 *$x_{\beta,VaR}$ is optimal to the model (13.1) if $T = t_\beta(x_{\beta,VaR})$.*

Table 13.3 Example 1: demand weight distributions

Weight	W_1		W_2	
Probability	0.5	0.5	0.4	0.6
Demand	8	10	2	12

Table 13.4 Total weighted distance distributions at point x

Realization	1	2	3	4
Probability	0.2	0.3	0.2	0.3
Value	$10 + 6x$	$60 - 4x$	$10 + 8x$	$60 - 2x$

Lemma 13.3 *$x_{P,T}$ is optimal to the model (13.1) if $\beta = P(D(x_{P,T}) \leq T)$.*

To motivate our study, consider a median problem on a segment of five units long with two nodes A and B. The demand weights of the two nodes are assumed to be independent, discrete random variables with the probability density functions presented in Table 13.3.

Let $\mathbf{S}(y)$ be the standard deviation of a random variable y. It is easy to verify $\mathbf{E}(W_1) = 9$, $\mathbf{E}(W_2) = 8$, $\mathbf{S}(W_1) = 1$, and $\mathbf{S}(W_2) = 4.9$. Suppose that a single facility is located at point x on the segment. The total weighted distance has mean $\mathbf{E}(D(x)) = 9x + 8(5 - x) = 40 + x$, and standard deviation

$$\mathbf{S}(D(x)) = \sqrt{\sum_{h \in N} [S(W_h)d(h,x)]^2} = \sqrt{x^2 + 4.9^2(5-x)^2} = \sqrt{25x^2 - 240x + 600}.$$

The probability density function of the weighted total distance over the segment is available in Table 13.4.

Figure 13.1 shows how the total demand-weighted distance's realizations (solid lines), expected value, and standard deviation (dash lines) evolve with x. Polyline $c_1c_2c_3$ represents the VaR function $t_\beta(x)$ when $\beta = 0.7$. We note that the risk-neutral solution node A ($x_E = 0$) that minimizes the expected total demand-weighted distance has the largest standard deviation. The probability that the total demand-weighted distance between the demand points and the expected median exceeds the expected median length $R_E = 40$, $P(D(x_E) > R_E)$, is as high as 0.6. In addition, the VaR value $t_\beta(x_E = 0)$ is 60, which is the highest value for $\beta = 0.7$. It appears that the risk-neutral solution is not desirable and that assessing the risk of a location decision is critical.

Let $T = 50$ and $\beta = 0.7$. Any point x at least 2.5 units away from node A is a maximum probability median as $P(D(x) \leq T)$ is 0.7. Though the maximum probability median model takes risk into consideration, our example seems to suggest that the model lacks the ability to discriminate satisfactory solutions. Examining the VaR function $t_\beta(x)$ in the figure, we observe that $x = 4.17$ is the minimum VaR median with an expected value of 44.17 and a VaR of 43.33. We note that any point between the minimum VaR median and node B is dominated as its expected value and VaR are both higher than the minimum VaR median. The mean-VaR median is the mid-point of the segment. The expected total demand-weighted distance and the VaR at this point are 42.50 and 50, respectively.

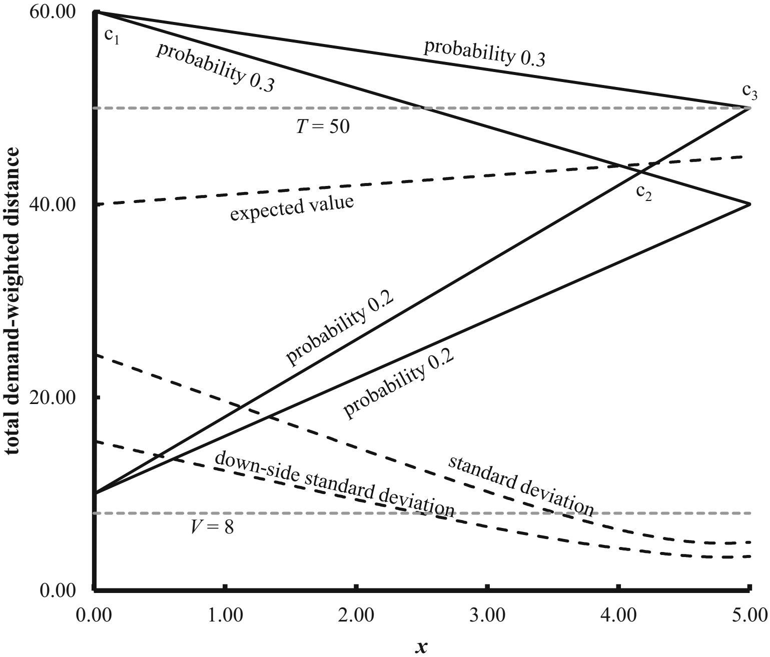


Fig. 13.1 A motivating example

Different from the mean-VaR model, the mean-variance model (Markowitz 1959) for the median problem presented below

$$\begin{aligned} &\min_{x \in G} \mathbf{E}(D(x)) \\ &\text{s.t.} \\ &\mathbf{S}(D(x)) \leq V, \end{aligned} \tag{13.2}$$

imposes a value V that bounds the standard deviation of the total demand-weighted distance from above. Figure 13.1 shows that $x = 3.54$ optimizes the mean-variance model with $V = 8$. We note that node optimality in general does not carry over to the mean-VaR model or the mean-variance model.

Standard deviation actually quantifies dispersions caused by both desirable and undesirable outcomes (Rockfella 2007). Therefore, good solutions may be inappropriately classified as infeasible to the mean-variance model. For instance, $x = 2.5$ would be deemed infeasible to the mean-variance median model with $V = 8$ because the standard deviation of the total demand-weighted total distance at that point is 12.5. In fact, the downside standard deviation (Nawrocki 1999) that measures the variability of realizations above the expected value only at $x = 2.5$ is 7.98, smaller than $V = 8$.

In the next two sections, solution approaches are developed for the model (13.1) with random demand weights of continuous and discrete probability distributions separately. Given $\beta \geq 0.5$, we consider a general problem where $t_\beta(x_{\beta,\mathrm{VaR}}) < T < t_\beta(x_E)$ and assume that the random demand weights are independent. The independence assumption will be relaxed later.

13.4 Continuous Random Weights

This section is devoted to solving the problem (13.1) when every demand weight w_h is a continuous random variable. It is sufficient to find an optimal point on every link of the network. In the sequel, we examine link (i, j). If $\mathbf{E}(D(x))$ is invariant with x on the link, then the problem reduces to finding feasible points only. We thus disregard this trivial case.

13.4.1 Dominant Points

Recall that $\mathbf{E}(D(x))$ is concave in terms of x over the link and hence one of the two nodes $x = 0$ and $x = l_{ij}$ is the expected median of the link. Denote by $x_E^{(i,j)}$ as follows:

$$x_E^{(i,j)} = \begin{cases} 0, & \text{if } \mathbf{E}(D(0)) < \mathbf{E}(D(l_{ij})) \\ l_{ij}, & \text{if } \mathbf{E}(D(0)) > \mathbf{E}(D(l_{ij})) \\ \arg\min\limits_{x=0,l_{ij}} t_\beta(x), & \text{if } \mathbf{E}(D(0)) = \mathbf{E}(D(l_{ij})) \end{cases}$$

Furthermore, denote by $\overline{x}_E^{(i,j)}$ the other nodal point.

Note that $x_E^{(i,j)}$ is optimal if $t_\beta(x_E^{(i,j)}) \leq T$. In the sequel, we assume $t_\beta(x_E^{(i,j)}) > T$. It is evident that $x_E^{(i,j)}$ is not feasible. Let $Z^{(i,j)} = \{x \in (i, j) | t_\beta(x) = T\}$. That is, $Z^{(i,j)}$ is the collection of any point x on the link such that $P(D(x) \leq T) = \beta$. As the VaR function $t_\beta(x)$ is continuous, the set $Z^{(i,j)}$ is empty when the smallest VaR on the link is greater than T, and non-empty otherwise. Denote the qth smallest element in the set $Z^{(i,j)}$ by $z_q^{(i,j)}$. For example, $z_1^{(i,j)}$ and $z_{|Z^{(i,j)}|}^{(i,j)}$ represent the smallest element and largest element, respectively, when $Z^{(i,j)}$ is not empty. The theorem below establishes a set of two dominant points, denoted by $Q^{(i,j)}$, in which the one with the smaller objective function value is optimal.

Theorem 13.1 *(i) The model (13.1) on link (i, j) is infeasible if $Z^{(i,j)} = \varnothing$. (ii) If $z_1^{(i,j)}$ is the only element in $Z^{(i,j)}$, then the dominant point set*

$Q^{(i,j)} = \{z_1^{(i,j)}, \overline{x}_E^{(i,j)}\}$. *(iii) If* $Z^{(i,j)}$ *has two or more elements,, then* $Q^{(i,j)} = \{z_1^{(i,j)}, z_{|Z^{(i,j)}|}^{(i,j)}\}$ *when* $t_\beta(\overline{x}_E^{(i,j)}) > T$*, and* $Q^{(i,j)} = \{z^*, \overline{x}_E^{(i,j)}\}$ *otherwise with* $z^* = \arg \max_{z=z_1^{(i,j)}, z_{|Z^{(i,j)}|}^{(i,j)}} |z - \overline{x}_E^{(i,j)}|$.

Proof (i) is straightforward. In case (ii), there is only one feasible interval on the link. In case (iii), there exit multiple non-overlapping feasible intervals. Note that the objective function is concave in terms of x. (ii) and (iii) follow because an optimal solution must be either the smallest feasible point or the largest feasible point.

On the basis of the above theorem, we develop the following solution procedure.

Algorithm 1 Finding an optimal point on link (i, j)

Step1 Determine $x_E^{(i,j)}$ and $\overline{x}_E^{(i,j)}$. Compute $t_\beta(x_E^{(i,j)})$. If $t_\beta(x_E^{(i,j)}) \le T$, stop and return $x_E^{(i,j)}$ as the optimal solution.

Step 2 Construct the set $Z^{(i,j)}$.

Step 3 If $Z^{(i,j)} = \varnothing$, the procedure terminates and the problem is infeasible on link (i, j).

Step 4 If $Z^{(i,j)}$ has only one element, then let $Q^{(i,j)} = \{z_1^{(i,j)}, \overline{x}_E^{(i,j)}\}$ and go to Step 6; if $Z^{(i,j)}$ has two or more elements, go to Step 5.

Step 5 If $t_\beta(\overline{x}_E^{(i,j)}) > T$, then $Q^{(i,j)} = \{z_1^{(i,j)}, z_{|Z^{(i,j)}|}^{(i,j)}\}$. Otherwise, let $z^* = \arg \max_{z=z_1^{(i,j)}, z_{|Z^{(i,j)}|}^{(i,j)}} \{|z - \overline{x}_E^{(i,j)}|\}$ and $Q^{(i,j)} = \{z^*, \overline{x}_E^{(i,j)}\}$.

Step 6 Return the point in $Q^{(i,j)}$ with the smaller objective function value as the optimal solution.

In the procedure presented above, it is essential to evaluate $t_\beta(x)$ at a given point x and construct the set $Z^{(i,j)}$. In the remainder of this section, we discuss these issues for random demand weights following the normal distribution and general probability distributions, respectively.

13.4.2 The Normal Distribution

Suppose that the demand weight W_h follows the normal distribution with a mean μ_h and a standard deviation σ_h. Given $x \in (i, j)$, $D(x)$ is a normal random variable with mean $\mathbf{E}(D(x)) = \sum_{h \in N} \mu_h d(h, x)$ and standard deviation $\mathbf{S}(D(x)) = \sqrt{\sum_{h \in N} \sigma_h^2 d^2(h, x)}$.

Given a threshold value T, $P(D(x) \le T)$ can be computed as

$$P(D(x) \le T) = \Phi(\frac{T - \mathbf{E}(D(x))}{\mathbf{S}(D(x))}),$$

where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal random variable. Denote by $\Phi^{-1}(\cdot)$ the inverse function of $\Phi(\cdot)$. It is easy to derive $t_\beta(x) = \mathbf{E}(D(x)) + \Phi^{-1}(\beta)\mathbf{S}(D(x))$.

Now we consider finding roots to the equation $t_\beta(x) = T$. The shortest distance from x to any node $h \in N$ can be evaluated as

$$d(h, x) = \min(d(h, i) + x, d(h, j) + l_{ij} - x).$$

An antipode c_h on link (i, j) with respect to node h is such that the shortest distances from it to h via nodes i or j are the same, i.e., $c_h = (d(h, j) + l_{ij} - d(h, i))/2$. We call a segment between two consecutive antipodes a *primary region*. As a node has at most one antipode on the link, the number of primary regions over a link does not exceed $n - 1$.

Note that the functional form of the distance function $d(h, \cdot)$ $\forall h \in N$ is identical within a primary region. We arbitrarily select a primary region $[c^{(m)}, c^{(m+1)}]$ with antipodes $c^{(m)}$ and $c^{(m+1)}$ as the end-points. For the primary region $[c^{(m)}, c^{(m+1)}]$, N is divided into two subsets N_L and N_R, where

$$N_L = \{h \in N | d(h, i) + c^{(m+1)} \le d(h, j) + l_{ij} - c^{(m+1)}\},$$

$$N_R = N - N_L.$$

Given $x \in [c^{(m)}, c^{(m+1)}]$, we have $d(h, x) = d(h, i)+x$ if $h \in N_L$ and $d(h, x) = d(h, j) + l_{ij} - x$ otherwise. $t_\beta(x)$ is thus expressed as

$$\begin{aligned} t_\beta(x) = &\sum_{h\in N_L} \mu_h [d(h, i) + x] + \sum_{h\in N_R} \mu_h \left[d(h, j) + l_{ij} - x\right] \\ &+\Phi^{-1}(\beta)\sqrt{\sum_{h\in N_L} \sigma_h^2 [d(h, i) + x]^2 + \sum_{h\in N_R} \sigma_h^2 \left[d(h, j) + l_{ij} - x\right]^2}. \end{aligned} \tag{13.3}$$

It can be shown that a necessary condition for a point x within the primary region to be a real root of the equation $t_\beta(x) = T$ is that x solves $F_1 x^2 + F_2 x + F_3 = 0$, where

$$\begin{aligned} F_1 &= \Phi^{-1}(\beta)^2(\sum_{h\in N_L} \sigma_h^2 + \sum_{h\in N_R} \sigma_h^2) - (\sum_{h\in N_L} \mu_h - \sum_{h\in N_R} \mu_h)^2, \\ F_2 &= 2\Phi^{-1}(\beta)^2[\sum_{h\in N_L} \sigma_h^2 d(h, i) - \sum_{h\in N_R} \sigma_h^2 (d(h, j) + l_{ij})] + \\ &\quad 2(\sum_{h\in N_L} \mu_h - \sum_{h\in N_R} \mu_h)[T - \sum_{h\in N_L} \mu_h d(h, i) - \sum_{h\in N_R} \mu_h (d(h, j) + l_{ij})], \end{aligned}$$

$$F_3 = \Phi^{-1}(\beta)^2[\sum_{h\in N_L}\sigma_h^2 d(h,i)^2 + \sum_{h\in N_R}\sigma_h^2(d(h,j)+l_{ij})^2] -$$
$$[T - \sum_{h\in N_L}\mu_h d(h,i) - \sum_{h\in N_R}\mu_h(d(h,j)+l_{ij})]^2.$$

Solving the quadratic equation $F_1x^2 + F_2x + F_3 = 0$, we may get zero, one, or two real roots within the primary region. If $F_2^2 - 4F_1F_3 < 0$, the quadratic equation has no real root and therefore there exists no point x within the primary region such that $t_\beta(x) = T$. Otherwise, the roots of the quadratic equation are computed as

$$x_1 = \frac{-F_2 + \sqrt{F_2^2 - 4F_1F_3}}{2F_1},$$
$$x_2 = \frac{-F_2 - \sqrt{F_2^2 - 4F_1F_3}}{2F_1}.$$

or

$$x_1 = -\frac{F_3}{F_2} \text{ if } F_1 = 0.$$

A root x_t solves $t_\beta(x) = T$ if (i) $c^{(m)} \le x_t \le c^{(m+1)}$; and (ii) $\sum_{h\in N}\mu_h d(h,x_t) \le T$ (because $\beta \ge 0.5$).

Given the matrix of distances between nodes, it takes time $O(n)$ to compute $t_\beta(x)$ at a given point x and solve the equation $t_\beta(x) = T$ within each primary region. There are at most n primary regions on the link. Hence the computational complexity of Algorithm 1 is $O(n^2)$ under the normality distribution.

13.4.3 General Distributions

If the demand weights follow arbitrary continuous distributions, Algorithm 1 is still applicable. If the closed-form expression of $P(D(x) \le T)$ can be developed or the probability $P(D(x) \le T)$ can be evaluated numerically for given T and x, e.g., when the demand weights follow the uniform distribution (see Sect. 13.6), we may use the line search approach to find the elements in the set $Z^{(i,j)}$, i.e., any point x on the link (i, j) such that $P(D(x) \le T) = \beta$. However, $P(D(x) \le T)$ may not have a closed form or may be rather time-consuming to evaluate for networks not sufficiently small. We thus suggest two approximation approaches when these difficulties are present.

According to the central limit theorem, when the number of nodes n is sufficiently large the total demand-weighted distance $\sum_{h \in N} W_h d(h, x)$ is approximately normal with mean $\mu(x) = \sum_{h=1}^{n} \mu_h d(x, h)$ and standard deviation $\sigma(x) = \sqrt{\sum_{h=1}^{n} \sigma_h^2 d^2(x, h)}$. By *normal approximation* we can treat the demand weight W_h of any node h as an independent normal random variable with mean μ_h and standard deviation σ_h and apply the results derived for the normally distributed demand weights to solve the problem.

Discretization, a univariate continuous density function, is a common method used in decision and risk analysis (Miller and Rice 1983; Keefer 1994). We here adopt the method suggested by Berman and Wang (2006) that approximates a continuous random demand weight W_h by a discrete probability distribution. By *discrete approximation*, each random weight becomes a discrete random variable and the model (13.1) can be solved using the procedure developed in the next section for random weights of discrete probability distributions.

We now consider approximating a continuous random demand weight W_h distributed over a closed interval $[a_h, b_h]$ by a discrete probability distribution represented by a set of values $w_h[k]$ and probabilities $p_h[k]$ $k = 1, 2, .., K$. If b_h is infinite, we shall use a sufficiently large but finite value $b_h^{'}$ to replace b_h as the upper limit. For example, $b_h^{'}$ can be determined by solving the equation $\varphi_h(b_h^{'}) = 0.99$, where $\varphi_h()$ is the cumulative distribution function of W_h.

A popular method is to choose Chebyshev points as the sample points. Given K,

$$w_h[k] = a_h + \left(\frac{b_h - a_h}{2}\right)\left[1 - \cos\frac{(2k-1)\pi}{2K}\right], \quad k = 1, 2, \ldots, K. \tag{13.4}$$

The probability mass $p_h[k]$ at $w_h[k]$ is calculated as

$$p_h[k] = \varphi_h(m_h[k]) - \varphi_h(m_h[k-1]), \quad k = 1, 2, \ldots, K \tag{13.5}$$

where

$$m_h[k] = \begin{cases} a_h & k = 0 \\ a_h + \left(\frac{b_h - a_h}{2}\right)\left[1 - \cos\frac{k\pi}{K}\right] & 1 \le k \le K-1 \\ b_h & k = K \end{cases}$$

Given z, $\varphi_h(z)$ is approximated by $\sum_{g=1}^{k} p_h(g)$ if there exists k $1 \le k < K$ such that $w_h[k] \le z < w_h[k+1]$. If $z < w_h[1]$ or $z \ge w_h[K]$, $\varphi_h(z)$ is respectively estimated as 0 or 1.

Table 13.5 Probability distribution for the example

k	1	2	3	4	5	6	7	8
$w_h[k]$	0.88	7.76	20.47	37.07	55.03	71.63	84.34	91.22
$p_h[k]$	0.1608	0.3298	0.2681	0.1414	0.0586	0.0218	0.0077	0.0119

As an example, consider a demand weight W_h following an exponential distribution with parameter $\lambda_h = 0.05$. The cumulative distribution function is $\varphi_h(z) = 1 - \exp(-0.05z)$. Note $a_h = 0$. Since $\varphi_h(92.10) = 0.99$, we have $b'_h = 92.10$. Let $K = 8$. Applying (13.4) and (13.5), we obtain the discrete distribution shown in Table 13.5.

13.5 Discrete Random Weights

In this section, the random weight W_h associated with every node h is assumed to follow a discrete probability distribution with possible realizations $w_h[k_h]$, $k_h = 1, 2, \cdots, K_h$, where K_h is the number of realizations. Without loss of generality, we assume $0 \le w_h[1] < w_h[2] < \cdots < w_h[K_h]$, $\forall h \in N$.

As in the previous section, we consider solving the problem (13.1) on link (i, j). At a given point $x \in G$, the total demand-weighted distance $D(x)$ is a discrete random variable with $\sum\limits_{h\in N} w_h[1]d(h,x)$ and $\sum\limits_{h\in N} w_h[K_h]d(h,x)$ being the smallest and largest realizations, respectively.

Let $\mathbf{W} = (W_1, W_2, \ldots, W_n)$ be the random vector of demand weights and S be the set of all possible realizations of $\widehat{\mathbf{w}} = (\widehat{w}_1, \widehat{w}_2, \ldots, \widehat{w}_n)$ of the random vector $\mathbf{W}$. It is easy to see that the cardinality of S is $|S| = \prod\limits_{h\in N} K_h$.

Any vector $\widehat{\mathbf{w}} \in S$ corresponds to a total demand-weighted distance function $f_{\widehat{\mathbf{w}}}(\cdot) = \sum\limits_{h\in N} \widehat{w}_h d(h, \cdot)$. Given point x on the link, $T > 0$ and $\widehat{\mathbf{w}} \in S$, define

$$Y_{\widehat{\mathbf{w}}}(x, T) = \begin{cases} 1 & \text{if } f_{\widehat{\mathbf{w}}}(x) \le T \\ 0 & \text{otherwise.} \end{cases}$$

The probability $P(D(x) \le T)$ is calculated as

$$P(D(x)) = \sum_{\widehat{\mathbf{w}}\in S} Y_{\widehat{\mathbf{w}}}(x, T) P_{\widehat{\mathbf{w}}}, \tag{13.6}$$

where $P_{\widehat{\mathbf{w}}} = \prod\limits_{h\in N} P(W_h = \widehat{w}_h)$.

Wang (2007) proved that determining whether the inequality $P(D(x) \le T) \ge \beta$ or, equivalently, $t_\beta(x) \le T$, holds for given $x \in G$, $T > 0$ and $\beta > 0$ is NP-complete. It is natural to conclude that the model (13.1) is NP-hard.

Given a point x, $t_\beta(x)$ can be evaluated using the procedure developed by Wang (2007). But for random weights of discrete probability distributions, we find it more convenient to examine the function $P(D(x) \leq T)$, instead of the function $t_\beta(x)$. Note that $P(D(x) \leq T)$ changes its value only at a jump point x where there exists at least a vector $\widehat{\mathbf{w}} \in S$ such that $\sum_{h \in N} \widehat{w}_h d(h, x) = T$ is true. Recall the definitions of the primary regions and the sets N_L and N_R introduced in Sect. 13.4. Consider a primary region $[c^{(m)}, c^{(m+1)}]$ on link (i, j). Given a vector $\widehat{\mathbf{w}} \in S$, the jump point can be computed as $x = \frac{T - \sum_{h \in N_L} \widehat{w}_h d(h,i) - \sum_{h \in N_R} \widehat{w}_h (d(h,j) + l_{ij})}{\sum_{h \in L} \widehat{w}_h - \sum_{h \in R} \widehat{w}_h}$ if $\sum_{h \in L} \widehat{w}_h \neq \sum_{h \in R} \widehat{w}_h$ and $c^{(m)} \leq x \leq c^{(m+1)}$. It is obvious that the maximum number of jump points within a primary region is $|S|$.

Redefine the set $Z^{(i,j)}$ as the collection of any feasible jump points of the function $P(D(x) \leq T)$ on the link. Again denote the qth smallest element in the set $Z^{(i,j)}$ by $z_q^{(i,j)}$. It is easy to see that the theorem presented in Sect. 13.4 is still valid.

A natural conclusion of the above analysis is that we can apply Algorithm 1 to search for an exact optimal point on link (i, j). However, the computational complexity of the algorithm is no longer polynomial time. The normal approximation approach is thus recommended when the network becomes too large.

13.6 An Illustrative Example

To illustrate Algorithm 1 developed above, let's solve the mean-VaR model (13.1) on a line of four demand nodes that are one unit away from the neighboring nodes. These nodes are labeled as 1–4 in sequence along the line. Assume that each demand weight W_h follows a uniform distribution over interval (a_h, b_h). Table 13.6 presents the parameters a_h and b_h.

Let $\beta = 0.75$ and $T = 31.6$. It is easy to verify that the expected median is node 3 with an expected median length of 17.00 and a VaR of 32.62. The minimum VaR median is node 2 with an expected total demand-weighted distance of 18.5 and a VaR of 27.62.

Berman and Wang (2006) derived $P(D(x) \leq T)$ at point x as follows:

Table 13.6 Demand weight distributions

h	1	2	3	4
a_h	5	5	0	5
b_h	10	10	5	10

$$P(D(x) \leq T) = \begin{cases} \dfrac{\sum_{\mathbf{v}\in\mathbf{C}} (sgn\ \mathbf{v})(T - \mathbf{d}'\mathbf{a} - \boldsymbol{\alpha}'\mathbf{v})_+^n}{n! \prod_{h=1}^{n} \alpha'_h}, & \sum_{h=1}^{n} \alpha_h a_h < T < \sum_{h=1}^{n} \alpha_h b_h \\ 0, & T \leq \sum_{h=1}^{n} \alpha_h a_h \\ 1, & T \geq \sum_{h=1}^{n} \alpha_h b_h \end{cases} \quad (13.7)$$

where $\mathbf{a} = (a_1, a_2, \ldots, a_n)'$, $\mathbf{d} = (d(x,1), d(x,2), \ldots, d(x,n))'$, $\alpha = (d(x,1)(b_1 - a_1), d(x,2)(b_2 - a_2), \ldots, d(x,n)(b_n - a_n))'$, $(y)_+^n = y^n$ if $y \geq 0$ and $(y)_+^n = 0$ if $y < 0$, and the set $\mathbf{C}$ is a collection of all $n-$tuple Cartesian products of the set $\{0, 1\}$. Hence, any vector $\mathbf{v} = (v_1, v_2, \ldots, v_n)'$ in the set $\mathbf{C}$ consists of n elements that are 0's or 1's.

We note that some of the elements in vector $\boldsymbol{\alpha}$ may be zero at point x. Define $B = \{h; \alpha'_h = 0\}$. Denote by $\bar{B}$ the complement of B and by $|\bar{B}|$ the number of elements in $\bar{B}$. Let $\mathbf{C}'$ be the set that contains all $|\bar{B}|$-tuple Cartesian products of the set $\{0, 1\}$. $P(D(x) \leq T)$ can be re-written as

$$P(D(x) \leq T) = \begin{cases} \dfrac{\sum_{\mathbf{v}\in\mathbf{C}'} (sgn\ \mathbf{v})(T - \boldsymbol{\alpha}\mathbf{a} - \boldsymbol{\alpha}'\mathbf{v})_+^{|\bar{B}|}}{|\bar{B}|! \prod_{h\in\bar{B}} \alpha'_h}, & \sum_{h=1}^{n} \alpha_h a_h < T < \sum_{h=1}^{n} \alpha_h b_h \\ 0, & T \leq \sum_{h=1}^{n} \alpha_h a_h \\ 1, & T \geq \sum_{h=1}^{n} \alpha_h b_h \end{cases}$$

As an illustration, let's compute $P(D(x) \leq T)$ at point $x = 0.83$ on the segment $(2, 3)$. We have $n = 4$, $\mathbf{a} = (5, 5, 0, 5)'$, $\mathbf{d} = (1.83, 0.83, 0.17, 1.17)'$, $\boldsymbol{\alpha} = (14.15, 9.15, 4.15, 5.85)'$. Recall $T = 31.6$. Applying Eq. (13.7), we obtain $P(D(x) \leq T) = \beta = 0.75$. Consequently, $x = 0.83$ belongs to the set $Z^{(2,3)}$. In fact, it is the only element in the set. On the segment (2, 3), node 3 is the median, but it is not feasible to the mean-VaR model. Therefore, $\overline{x}_E^{(i,j)} = 0$, i.e., node 2. By Algorithm 1, the dominant set $Q^{(2,3)} = \{0, 0.83\}$. As $x = 0.83$ has a smaller expected total demand-weighted distance of 18.36, it is the mean-VaR median on the segment.

13.7 Concluding Remarks

To the best of our knowledge, this is the first study to integrate both the expected value criteria with the VaR measure in a location decision problem. The objective of the model proposed in this study is to find the best location to site a facility, namely the mean-VaR median, to minimize the expected total demand-weighted distance subject to a constraint that bounds the VaR of the total demand-weighted distance from above. It is shown that the deterministic median problem, the maximum probability median problem, and the minimum VaR median problem are all special cases of the mean-VaR median problem. A solution approach is developed that identifies the dominant points for the optimal solution. We also present the methods to calculate the dominant points when the demand weights follow, respectively, discrete and continuous probability distributions.

It is assumed in the study that all random demand weights are independent. When the demand weights are not independent, scenario planning (Snyder et al. 2007) can be applied where a number of scenarios of the demand weights are identified. Let's assume that there are M scenarios, w'_{hr} denotes the demand weight at node h under scenario r, and p'_r be the scenario probability $r = 1, 2, \cdots M$. It is easy to see that $P(D(x) \leq T) = \sum_{h \in N} p'_r s_r$, where $s_r = 1$ if $\sum_{h \in N} w'_{hr} d(h, x) \leq T$ and $s_r = 0$ otherwise.

In the current study, we consider the single-facility mean-VaR median problem where the facility can be located anywhere in the network. We are investigating its multi-facility version. However, because the dominant point set appears difficult to construct if multiple facilities are located, we may have to limit the potential sites to be the nodal points only.

References

Aboolian, R., Berman, O., & Drezner, Z. (2009). The multiple server center location problem. *Annals of Operations Research, 167*, 337–352.

Berman, O., & Drezner, Z. (2000). A note on the location of an obnoxious facility on a network. *European Journal of Operational Research, 120*, 215–217.

Berman, O., & Drezner, Z. (2003). A probabilistic one-center location problem on a network. *The Journal of the Operational Research Society, 54*, 871–877.

Berman, O., & Drezner, Z. (2008). A new formulation for the conditional p-median and p-center problems. *Operations Research Letters, 36*, 481–483.

Berman, O., Drezner, Z., Tamir, A., & Wesolowsky, G. O. (2009). Optimal location with equitable loads. *Annals of Operations Research, 167*, 307–325.

Berman, O., Drezner, Z., Wang, J., & Wesolowsky, G. (2003a). The minimax and maximin location problems on a network with uniform distributed weights. *IIE Transactions, 35*, 1017–1025.

Berman, O., Drezner, Z., & Wesolowsky, G. (2001). Location of facilities on a network with groups of demand points. *IIE Transactions, 33*, 637–648.

Berman, O., Drezner, Z., & Wesolowsky, G. (2003b). The expected maximum distance objective in facility location. *Journal of Regional Science, 43*, 735–748.

Berman, O., Drezner, Z., & Wesolowsky, G. (2005). The facility and transfer points location problem. *International Transactions in Operations Research, 12*, 387–402.

Berman, O., Drezner, Z., & Wesolowsky, G. (2007). The transfer point location problem. *European Journal of Operational Research, 179*, 978–989.

Berman, O., Sanajian, N., & Wang, J. (2016). Location choice and risk attitude of a decision maker. *Omega, 66*, 170–181.

Berman, O., & Wang, J. (2006). The 1-median and 1-antimedian problems with continuous probabilistic demand weights. *Information Systems and Operational Research, 44*, 267–283.

Berman, O., Wang, J., Drezner, Z., & Wesolowsky, G. (2003c). A probabilistic minimax location problem on the plane. *Annals of Operations Research, 122*, 59–70.

Chen, G., Daskin, M., Shen, Z., & Uryasev, S. (2006). The α-reliable mean-excess regret model for stochastic facility location modeling. *Naval Research Logistics, 53*, 617–626.

Christofides, N. (1975). *Graph theory - An algorithmic approach*. London: Academic.

Correia, I., & Saldanha da Gama, F. (2015). Facility location under uncertainty. In G. Laporte, S. Nickel, & F. Saldanha da Gama (Eds.), *Location science* (pp. 177–203). New York: Springer.

Daskin, M., Hesse, S., & ReVelle, C. (1997). α-Reliable p-minimax regret: A new model for strategic facility location modeling. *Location Science, 5*, 227–246.

Drezner, Z. (1984a). The planar two-center and two-median problems. *Transportation Science, 18*, 351–361.

Drezner, Z. (1984b). The p-center problem – Heuristic and optimal algorithms. *The Journal of the Operational Research Society, 35*, 741–748.

Drezner, Z. (1987). Heuristic solution methods for two location problems with unreliable facilities. *The Journal of the Operational Research Society, 38*, 509–514.

Drezner, Z. (1989). Conditional p-center problem. *Transportation Science, 23*, 51–53.

Drezner, Z. (1991). The weighted minimax location problem with set-up costs and extensions. *RAIRO - Operations Research - Recherche Opérationnelle, 25*, 55–64.

Drezner, Z. (2011). Continuous center problems. In H.A. Eiselt & V. Marianov (Eds.), *Foundations of location analysis* (pp. 63–78). New York: Springer.

Drezner, Z., & Gavish, B. (1985). ε-Approximations for multidimensional weighted location problems. *Operations Research, 33*, 772–783.

Drezner, Z., Mehrez, A., & Wesolowsky, G. (1991). The facility location problems with limited distances. *Transportation Science, 25*, 182–187.

Drezner, Z., & Shelah, S. (1987). On the complexity of the Elzinga-Hearn algorithm for the 1-center problem. *Mathematics of Operations Research, 12*, 255–261.

Drezner, Z., Thisse, J., & Wesolowsky, G. (1986). The minmax-min location problem. *Journal of Regional Science, 26*, 87– 101.

Drezner, Z., & Wesolowsky, G. (1978). A new method for the multifacility minimax location problem. *The Journal of the Operational Research Society, 29*, 1095–1101.

Drezner, Z., & Wesolowsky, G. (1980). A maximin location problem with maximum distance constraints. *AIIE Transactions, 12*, 249–252.

Drezner, Z., & Wesolowsky, G. (1985). Location of multiple obnoxious facilities. *Transportation Science, 19*, 193–202.

Drezner, Z., & Wesolowsky, G. (1989). The asymmetric distance location problem. *Transportation Science, 23*, 1–8.

Drezner, Z., & Wesolowsky, G. (1995). Location on a one-way rectilinear grid. *The Journal of the Operational Research Society, 46*, 735–746.

Drezner, Z., & Wesolowsky, G. (1996). Obnoxious facility location in the interior of a planar network. *Journal of Regional Science, 35*, 675–688.

Drezner, Z., & Wesolowsky, G. (2000). Location models with groups of demand points. *Information Systems and Operational Research, 38*, 359– 372.

Eiselt, H.A., & Laporte, G. (1995). Objectives in location problems. In Z. Drezner (Ed.), *Facility location* (pp. 151–180). New York: Springer.

Elzinga, J., & Hearn, D. (1972). Geometrical solutions for some minimax location. *Transportation Science, 6*, 379–394.

Frank, H. (1966). Optimum location on a graph with probabilistic demands. *Operations Research, 14*, 404–421.

Frank, H. (1967). Optimum location on a graph with correlated normal demands. *Operations Research, 14*, 552–557.

Hakimi, S. (1964). Optimal location of switching centers and the absolute centers and medians of a graph. *Operations Research, 12*, 450–459.

Hakimi, S. (1965). Optimum distribution of switching centers in a communication network and some related graph theoretic problems. *Operations Research, 13*, 462–475.

Hodder, J., & Jucker, J. (1985). A simple plant location model for quantity setting firms subject to price uncertainty. *European Journal of Operational Research, 21*, 39–46.
Jucker, J., & Hodder, J. (1976). The simple plant-location problem under uncertainty. *Operations Research, 24*, 1045–1055.
Keefer, D. (1994). Certainty equivalents for three-point discrete-distribution approximations. *Management Science, 40*, 760–773.
Markowitz, H. (1959). *Portfolio selection: Efficient diversification of investment*. New York: Wiley.
Melachrinoudis, E. (2011). The location of undesirable facilities. In H.A. Eiselt & V. Marianov (Eds.), *Foundations of location analysis* (pp. 207–239). New York: Springer.
Miller, A., & Rice, T. (1983). Discrete approximations of probability distributions. *Management Science, 29*, 352–362.
Nawrocki, D. (1999). A brief history of downside risk measures. *The Journal of Investing, 8*, 9–25.
Pflug, G. (2000). Some remarks on the value-at-risk and the conditional value-at-risk. In S. Urayasev (Ed.), *Probabilistic constrained optimization: Methodology and applications* (pp. 272–281). Boston: Kluwer Academic Publishers.
Rockfella, R. (2007). Coherent approaches to risk in optimization under uncertainty. *INFORMS Tutorials in Operations Research, 8*, 38–61.
Snyder, L. (2006). Facility location under uncertainty: A review. *IIE Transactions, 38*, 537–554.
Snyder, L., Daskin, M., & Teo, C. (2007). The stochastic location model with risk pooling. *European Journal of Operational Research, 179*, 1221–1238.
Sugihara, K., & Iri, M. (1992). Construction of the Voronoi diagram for "one million" generators in single-precision arithmetic. *Proceedings of the IEEE, 80*, 1471–1484.
Sugihara, K., & Iri, M. (1994). A robust topology-oriented incremental algorithm for Voronoi diagram. *International Journal of Computational Geometry and Applications, 4*, 179–228.
Suzuki, A., & Drezner, Z. (1996). The *p*-center location problem in an area. *Location Science, 4*, 69–82.
Suzuki, A., & Drezner, Z. (2009). The minimum equitable radius location problem with continuous demand. *European Journal of Operational Research, 195*, 17–30.
Tansel, B. (2011). Discrete center problems. In H.A. Eiselt & V. Marianov (Eds.), *Foundations of location analysis* (pp. 79–106). New York: Springer.
Wagner, M., Bhadury, J., & Peng, S. (2009). Risk management in uncapacitated facility location models with random demands. *Computers & Operations Research, 36*, 1002–1011.
Wagner, M., & Lau, S. (1971). The effect of diversification on risk. *Financial Analysis Journal, 27*, 48–53.
Wang, J. (2007). The β-reliable median on a network with discrete probabilistic demand weights. *Operations Research, 55*, 966–975.
Welch, S., Salhi, S., & Drezner, Z. (2006). The multifacility maximin planar location problem with facility interaction. *IMA Journal of Management Mathematics, 17*, 397–412.

Chapter 14
A Bivariate Exponential Distribution

Dawit Zerom and Zvi Drezner

We introduce a new bivariate exponential distribution that is analytically tractable and easily implementable. Some of its properties are discussed. Explicit expressions of the expected value of the larger and of the smaller of a pair of correlated exponentials are also provided. An application to tandem queues offers several interesting implications.

We first review three of our papers that provide background to the main result of this chapter—a new bivariate exponential distribution.

Drezner et al. (2010) improved the Kolmogorov–Smirnov test for normality (Massey 1951). In the current implementation of the Kolmogorov–Smirnov test, given data are compared with a normal distribution that uses the sample mean and the sample variance. Drezner et al. (2010) proposed to select the mean and variance of the normal distribution that provides the closest fit to the data. This is like shifting and stretching the reference normal distribution so that it fits the data in the best possible way.

Drezner and Zerom (2016) proposed a generally applicable discretization method to approximate a continuous distribution on a real line with a discrete one, supported by a finite set. The method adopts a criterion which is shown to be flexible in approximating higher order features of the underlying continuous distribution while preserving its mean and variance. In Table 14.1 the abscissas and probabilities of discretizations of standardized normal, uniform, and exponential distributions using $K = 5$ and $K = 10$ discrete points are given.

Drezner et al. (2018) proposed a competitive location model using the gravity model for estimating market share assuming that attractiveness levels are random.

D. Zerom (✉) · Z. Drezner
Steven G. Mihaylo College of Business and Economics, California State University-Fullerton, Fullerton, CA, USA
e-mail: dzerom@fullerton.edu; zdrezner@fullerton.edu

H. A. Eiselt, V. Marianov (eds.), *Contributions to Location Analysis*, International Series in Operations Research & Management Science 281,
https://doi.org/10.1007/978-3-030-19111-5_14

Table 14.1 Discretizations for normal, uniform, and exponential distributions

Normal				Uniform				Exponential			
$K = 5$		$K = 10$		$K = 5$		$K = 10$		$K = 5$		$K = 10$	
x_i	p_i	x_i	p_i	x_i	p_i	x_i	p_i	x_i	p_i	x_i	p_i
1.7522	0.1120	2.4001	0.0245	0.1006	0.2180	0.0548	0.1144	0.2099	0.4425	0.1372	0.2833
0.7812	0.2556	1.6287	0.0681	0.3096	0.1904	0.1644	0.1034	0.9174	0.2797	0.4782	0.1928
0.0000	0.2647	1.0827	0.1095	0.5000	0.1834	0.2652	0.0970	1.8047	0.1872	0.8166	0.1547
−0.7812	0.2556	0.6242	0.1407	0.6904	0.1904	0.3609	0.0934	3.0866	0.0748	1.2092	0.1295
−1.7522	0.1120	0.2043	0.1572	0.8994	0.2180	0.4539	0.0918	5.1668	0.0158	1.6914	0.1001
		−0.2043	0.1572			0.5461	0.0918			2.2854	0.0683
		−0.6242	0.1407			0.6391	0.0934			3.0137	0.0402
		−1.0827	0.1095			0.7348	0.0970			3.9232	0.0205
		−1.6287	0.0681			0.8356	0.1034			5.1505	0.0085
		−2.4001	0.0245			0.9452	0.1144			7.1674	0.0021

One of the solution approaches is to discretize the normal distribution of the attractiveness level using the approach in Drezner and Zerom (2016). In order to perform the calculations, a bivariate normal distribution needs to be discretized.

When X_1 and X_2 are uncorrelated standardized normal distributions, Y_1 and Y_2 are standardized bivariate normal with a correlation ρ.

$$Y_1 = \sqrt{\frac{1+\sqrt{1-\rho^2}}{2}}X_1 \pm \sqrt{\frac{1-\sqrt{1-\rho^2}}{2}}X_2$$
$$Y_2 = \sqrt{\frac{1+\sqrt{1-\rho^2}}{2}}X_2 \pm \sqrt{\frac{1-\sqrt{1-\rho^2}}{2}}X_1 \qquad (14.1)$$

The sign between the terms is "+" for positive ρ and "−" for negative ρ. It is easy to show that Y_1 and Y_2 have an average of 0, variance of 1, and a correlation ρ between them. If we need non-standardized Y_1 and Y_2, we can use $\mu_1 + \sigma_1 Y_1$ and similarly for Y_2.

When X_1 and X_2 are discretized using K points, then using Eq. (14.1) yields

$$Z_{ij} = \left(\sqrt{\frac{1+\sqrt{1-\rho^2}}{2}}x_i \pm \sqrt{\frac{1-\sqrt{1-\rho^2}}{2}}x_j\ ,\ \sqrt{\frac{1+\sqrt{1-\rho^2}}{2}}x_j \pm \sqrt{\frac{1-\sqrt{1-\rho^2}}{2}}x_i\right) \qquad (14.2)$$

with probability $p_i p_j$. The sign between the terms is "+" for positive ρ and "−" for negative ρ. Note that if the marginal distributions are not standardized, each coordinate can be adjusted by multiplying the corresponding σ and adding the corresponding μ.

For example, a twenty-five point (K^2 where $K = 5$) discretization of the standardized bivariate normal distribution is plotted in Fig. 14.1 for several correlation values. Note that the discretization accurately captures the dependence structure of the underlying bivariate normal random variables.

14.1 Introduction

The need for a bivariate exponential distribution may arise in many practical settings. Consider a queueing system in which two servers process the same arriving customer in tandem (or in sequence), i.e., each customer is serviced at the first server and then proceeds to the second server before leaving the system. Since each

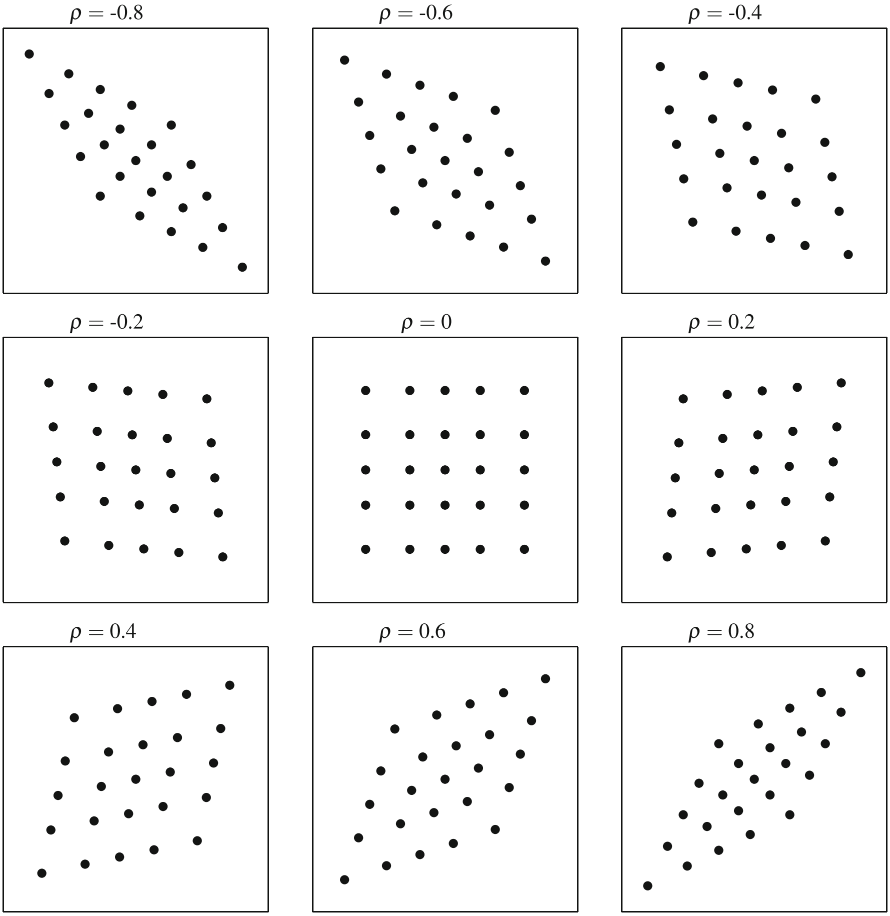


Fig. 14.1 Distribution of points for bivariate normal

customer must be processed by both servers, service times at the two servers for a given customer are likely to be correlated. For example, in the case of vehicles passing through border crossing facilities, Cetin and List (2004) noted that if processing a vehicle takes a long time at the primary inspection (first server), say, because of a paperwork, it is reasonable to assume that the service times at the secondary inspection (second server) will also be long, inducing positive correlation between the two service times. A bivariate exponential distribution can be used to model such correlation between service times.

Unlike the normal distribution, the exponential distribution does not have a unique natural extension to the bivariate or the multivariate case. Thus, a variety of bivariate exponential distributions have been introduced in the literature (for example, Gumbel (1960), Downton (1970), and Arnold and Strauss (1988)). Recently, Kim and Kim (2018) use the bivariate exponential distribution of Downton (1970) to model the correlation between inter-arrival and service times in an application to queueing systems. Other applications of bivariate exponential distributions to queueing include, among others, Mitchell et al. (1977) and Cetin and List (2004).

In this paper we introduce a new bivariate exponential distribution. For $\lambda_1 > 0$, $\lambda_2 > 0$ and correlation coefficient $0 \le \rho \le 1$, the joint probability density function of the bivariate random vector (X, Y) is given by

$$f_{X,Y}(x, y) = \lambda_1 \lambda_2 \sqrt{\frac{1+\rho}{1-\rho}} \exp\left(-\frac{\lambda_x x + \lambda_y y}{\sqrt{1-\rho}}\right) \tag{14.3}$$

where λ_x and λ_y are simple functional of λ_1, λ_2, and ρ; see (14.3). Let $\mu_1 = E(X)$, $\mu_2 = E(Y)$, $\sigma_1^2 = Var(X_1)$, and $\sigma_2^2 = Var(Y)$. The parameters λ_1 and λ_2 are shown to have exact functional relationship with μ_1, μ_2, and ρ. Therefore, when sample data on (X, Y) is available, the parameters λ_1 and λ_2 can computed quite easily. Let the random vector $(\widetilde{X}, \widetilde{Y})$ be a special case of (X, Y) when $\rho = 0$. In this case, the marginals are exponential and their corresponding means are $\widetilde{\mu}_1 = 1/\lambda_1$ and $\widetilde{\mu}_2 = 1/\lambda_2$. Let $\delta = \widetilde{\mu}_2/\widetilde{\mu}_1$. Further, let $-\frac{\pi}{2} \le \theta \le \frac{\pi}{2}$ such that $\sin\theta = \rho$. The proposed bivariate exponential distribution has the following interesting properties that can be useful in some applications.

1. $E(X + Y) = \widetilde{\mu}_1 + \widetilde{\mu}_2$.
2. $\frac{\mu_1}{\widetilde{\mu}_1} = 1 + \frac{\sin\theta/2}{\sqrt{1+\rho}}(\delta - 1)$ and $\frac{\mu_2}{\widetilde{\mu}_2} = 1 + \frac{\sin\theta/2}{\sqrt{1+\rho}}(\delta^{-1} - 1)$.
   Note that when $\delta \ge 1$, $\frac{\mu_1}{\widetilde{\mu}_1} \ge 1$ and $\frac{\mu_2}{\widetilde{\mu}_2} \le 1$. When $\delta \le 1$, $\frac{\mu_1}{\widetilde{\mu}_1} \le 1$ and $\frac{\mu_2}{\widetilde{\mu}_2} \ge 1$. Therefore, the role of positive correlation is to redistribute the means while maintaining the total sum.
3. $Var(X + Y) = \widetilde{\mu}_1^2 + \widetilde{\mu}_2^2$.
4. $\frac{\sigma_1^2}{\widetilde{\mu}_1^2} = \frac{1}{1+\rho}\left[1 + (\sin\theta/2)^2\ (\delta^2 - 1)\right]$ and $\frac{\sigma_2^2}{\widetilde{\mu}_2^2} = \frac{1}{1+\rho}\left[1 + (\sin\theta/2)^2\ (\delta^{-2} - 1)\right]$.

Note that when $\delta \ge 1$, $\frac{\sigma_1^2}{\widetilde{\mu}_1^2} \ge \frac{1}{1+\rho}$ and $\frac{\sigma_2^2}{\widetilde{\mu}_2^2} \le \frac{1}{1+\rho}$. When $\delta \le 1$, $\frac{\sigma_1^2}{\widetilde{\mu}_1^2} \le \frac{1}{1+\rho}$ and $\frac{\sigma_2^2}{\widetilde{\mu}_2^2} \ge \frac{1}{1+\rho}$. Again, the role of positive correlation is to redistribute the variances while maintaining the total sum. In the case of $\delta = 1$, both variances σ_1^2 and σ_2^2 are only $1/(1+\rho)$ of their corresponding variances under zero correlation. Hence, positive correlation reduces both. When correlation is non-zero, $\sigma_1^2 + \sigma_2^2 \le \widetilde{\mu}_1^2 + \widetilde{\mu}_2^2$.

The rest of the paper is organized as follows. In Sect. 14.2, we describe a general framework for creating correlated bivariate random variables. In Sect. 14.3, we provide the case of correlated exponentials. As a part of Sect. 14.3, we also provide easy to use analytical expressions of the expected values of the maximum and the minimum of the correlated exponentials. In Sect. 14.4, a brief application to tandem queues is discussed.

14.2 General Correlated Bivariate Random Variables

Consider a bivariate random vector $(\widetilde{X}, \widetilde{Y})$ where their correlation $Cor(\widetilde{X}, \widetilde{Y}) = 0$ and both $\widetilde{X}$ and $\widetilde{Y}$ are standardized to have mean 0 and variance 1. The goal is to create an associated correlated bivariate random vector (X, Y). To this end, let

$$X = \alpha\widetilde{X} + \beta\widetilde{Y} \quad \text{and} \quad Y = \alpha\widetilde{Y} + \beta\widetilde{X} \tag{14.4}$$

where α and β are unknown parameters. It can be seen that $E(X) = E(Y) = 0$. Further,

$$Var(X) = Var(Y) = \alpha^2 + \beta^2 \quad \text{and} \quad Cor(X, Y) = \frac{2\alpha\beta}{\alpha^2 + \beta^2}. \tag{14.5}$$

Suppose that (X, Y) must satisfy the following properties:

$$Var(X) = Var(Y) = 1 \quad \text{and} \quad Cor(X, Y) = \rho \tag{14.6}$$

where $\rho \geq 0$. To simplify the presentation we define $-\frac{\pi}{2} \leq \theta \leq \frac{\pi}{2}$ such that

$$\sin\theta = \rho.$$

In the subsequent derivations, we exploit the following useful trigonometric identities:

$\rho = 2\sin\frac{\theta}{2}\cos\frac{\theta}{2}$.
$\cos^2\frac{\theta}{2} = \frac{1+\sqrt{1-\rho^2}}{2}$; $\sin^2\frac{\theta}{2} = \frac{1-\sqrt{1-\rho^2}}{2}$.
$\tan\frac{\theta}{2} = \frac{\rho}{1+\sqrt{1-\rho^2}}$; $\cot\frac{\theta}{2} = \frac{1+\sqrt{1-\rho^2}}{\rho}$.

Solving for α and β in (14.5) that satisfy (14.6), (X, Y) becomes

$$X = \cos\frac{\theta}{2}\widetilde{X} + \sin\frac{\theta}{2}\widetilde{Y} \quad \text{and} \quad Y = \cos\frac{\theta}{2}\widetilde{Y} + \sin\frac{\theta}{2}\widetilde{X}. \tag{14.7}$$

The reverse transformation of (14.7) is

$$\widetilde{X} = \frac{\cos\frac{\theta}{2}X - \sin\frac{\theta}{2}Y}{\sqrt{1-\rho^2}} \quad \text{and} \quad \widetilde{Y} = \frac{\cos\frac{\theta}{2}Y - \sin\frac{\theta}{2}X}{\sqrt{1-\rho^2}} \tag{14.8}$$

Equations (14.7) and (14.8) imply that for any given pair $(\widetilde{X} = \widetilde{x}, \widetilde{Y} = \widetilde{y})$ there is only one corresponding pair $(X = x, Y = y)$ and vice versa. Let $f_{X,Y}(x, y)$ denote the joint probability density function (pdf) of (X, Y). Then,

$$f_{X,Y}(x,y) = K\ f_{\widetilde{X}}\left(\frac{\cos\frac{\theta}{2}x - \sin\frac{\theta}{2}y}{\sqrt{1-\rho^2}}\right) \times f_{\widetilde{Y}}\left(\frac{\cos\frac{\theta}{2}y - \sin\frac{\theta}{2}x}{\sqrt{1-\rho^2}}\right) \tag{14.9}$$

where K is the normalizing constant and $f_{\widetilde{X}}(\cdot)$ and $f_{\widetilde{Y}}(\cdot)$ are the marginal pdf of $\widetilde{X}$ and $\widetilde{Y}$, respectively. Let $f_X(x)$ and $f_Y(y)$ denote the marginal pdf of X and Y, respectively. It follows that

$$f_X(x) \propto \int_{-\infty}^{\infty} f_{\widetilde{X}}\left(\frac{\cos\frac{\theta}{2}x - \sin\frac{\theta}{2}Y}{\sqrt{1-\rho^2}}\right) \times f_{\widetilde{Y}}\left(\frac{\cos\frac{\theta}{2}Y - \sin\frac{\theta}{2}x}{\sqrt{1-\rho^2}}\right) dY. \tag{14.10}$$

The pdf $f_Y(y)$ is defined analogously.

14.3 Correlated Exponentials

Using the general framework outlined in Sect. 14.2, we introduce a new correlated bivariate exponential distribution focusing on the case of $\rho \geq 0$. The following standardized exponential distribution leads to the desired result:

$$f_{\widetilde{X}}(\tilde{x}) = \sqrt{1+\rho}e^{-\sqrt{1+\rho}\tilde{x}}, \quad \tilde{x} \geq 0 \tag{14.11}$$

and similarly for $\widetilde{Y}$. Using (14.10), such choice of marginals for the uncorrelated $\widetilde{X}$ and $\widetilde{Y}$ leads to

$$f_X(x) \propto e^{-x}$$

and similarly for Y. Let U and V be two uncorrelated exponential random variables with a $\lambda = 1$. Set $\widetilde{X} = \frac{U}{\lambda_1\sqrt{1+\rho}}$ and $\widetilde{Y} = \frac{V}{\lambda_2\sqrt{1+\rho}}$. Using (14.7),

$$X = \cos\frac{\theta}{2}\widetilde{X} + \sin\frac{\theta}{2}\widetilde{Y} = \frac{U}{\lambda_1\sqrt{1+\rho}}\sqrt{\frac{1+\sqrt{1-\rho^2}}{2}} + \frac{V}{\lambda_2\sqrt{1+\rho}}\sqrt{\frac{1-\sqrt{1-\rho^2}}{2}}$$

$$Y = \cos\frac{\theta}{2}\widetilde{Y} + \sin\frac{\theta}{2}\widetilde{X} = \frac{U}{\lambda_1\sqrt{1+\rho}}\sqrt{\frac{1-\sqrt{1-\rho^2}}{2}} + \frac{V}{\lambda_2\sqrt{1+\rho}}\sqrt{\frac{1+\sqrt{1-\rho^2}}{2}}$$

The reverse transformation is

$$X\cos\frac{\theta}{2} - Y\sin\frac{\theta}{2} = (\cos^2\frac{\theta}{2} - \sin^2\frac{\theta}{2})\widetilde{X} = \sqrt{1-\rho^2}\widetilde{X}$$

$$= \frac{U}{\lambda_1\sqrt{1+\rho}}\sqrt{1-\rho^2} = \frac{U\sqrt{1-\rho}}{\lambda_1}$$

leading to:

$$U = \lambda_1 \frac{X\cos\frac{\theta}{2} - Y\sin\frac{\theta}{2}}{\sqrt{1-\rho}}; \quad V = \lambda_2 \frac{Y\cos\frac{\theta}{2} - X\sin\frac{\theta}{2}}{\sqrt{1-\rho}}$$

Since the density function of U is e^{-u} and similarly for V,

$$\begin{aligned} f_{X,Y}(x,y) &= K\, e^{-\lambda_1 \frac{x\cos\frac{\theta}{2} - y\sin\frac{\theta}{2}}{\sqrt{1-\rho}}} \times e^{-\lambda_2 \frac{y\cos\frac{\theta}{2} - x\sin\frac{\theta}{2}}{\sqrt{1-\rho}}} \\ &= K\, e^{-\frac{x\left\{\lambda_1\cos\frac{\theta}{2} - \lambda_2\sin\frac{\theta}{2}\right\} + y\left\{\lambda_2\cos\frac{\theta}{2} - \lambda_1\sin\frac{\theta}{2}\right\}}{\sqrt{1-\rho}}} \end{aligned}$$

where K is the normalizing constant. To obtain K, note that

$$\begin{aligned} \frac{1}{K} &= \int_0^\infty e^{-\frac{x\left\{\lambda_1\cos\frac{\theta}{2} - \lambda_2\sin\frac{\theta}{2}\right\}}{\sqrt{1-\rho}}} \int_{x\tan\frac{\theta}{2}}^{x\cot\frac{\theta}{2}} e^{-\frac{y\left\{\lambda_2\cos\frac{\theta}{2} - \lambda_1\sin\frac{\theta}{2}\right\}}{\sqrt{1-\rho}}}\, dy\, dx \\ &= \frac{\sqrt{1-\rho}}{\lambda_2\cos\frac{\theta}{2} - \lambda_1\sin\frac{\theta}{2}} \int_0^\infty e^{-\frac{x\left\{\lambda_1\cos\frac{\theta}{2} - \lambda_2\sin\frac{\theta}{2}\right\}}{\sqrt{1-\rho}}} \\ &\quad \left\{ e^{-\frac{x\tan\frac{\theta}{2}\left\{\lambda_2\cos\frac{\theta}{2} - \lambda_1\sin\frac{\theta}{2}\right\}}{\sqrt{1-\rho}}} - e^{-\frac{x\cot\frac{\theta}{2}\left\{\lambda_2\cos\frac{\theta}{2} - \lambda_1\sin\frac{\theta}{2}\right\}}{\sqrt{1-\rho}}} \right\} dx \\ &= \frac{1-\rho}{\lambda_2\cos\frac{\theta}{2} - \lambda_1\sin\frac{\theta}{2}} \left\{ \frac{1}{\lambda_1\cos\frac{\theta}{2} - \lambda_2\sin\frac{\theta}{2} + \tan\frac{\theta}{2}\left(\lambda_2\cos\frac{\theta}{2} - \lambda_1\sin\frac{\theta}{2}\right)} \right. \\ &\quad \left. - \frac{1}{\lambda_1\cos\frac{\theta}{2} - \lambda_2\sin\frac{\theta}{2} + \cot\frac{\theta}{2}\left(\lambda_2\cos\frac{\theta}{2} - \lambda_1\sin\frac{\theta}{2}\right)} \right\} \\ &= \frac{1-\rho}{\lambda_2\cos\frac{\theta}{2} - \lambda_1\sin\frac{\theta}{2}} \left\{ \frac{\cos\frac{\theta}{2}}{\lambda_1\left(\cos^2\frac{\theta}{2} - \sin^2\frac{\theta}{2}\right)} - \frac{\sin\frac{\theta}{2}}{\lambda_2\left(\cos^2\frac{\theta}{2} - \sin^2\frac{\theta}{2}\right)} \right\} \\ &= \frac{1-\rho}{\lambda_1\lambda_2\sqrt{1-\rho^2}}. \end{aligned}$$

Furthermore, because

$$x\cos\frac{\theta}{2} - y\sin\frac{\theta}{2} \geq 0 \quad \text{and} \quad y\cos\frac{\theta}{2} - x\sin\frac{\theta}{2} \geq 0,$$

we require that

$$x\tan\frac{\theta}{2} \leq \ y \ \leq x\cot\frac{\theta}{2}.$$

The joint pdf of (X, Y) is given by

$$f_{X,Y}(x, y) = \lambda_1\lambda_2\sqrt{\frac{1+\rho}{1-\rho}}e^{-\frac{x\left\{\lambda_1\cos\frac{\theta}{2}-\lambda_2\sin\frac{\theta}{2}\right\}+y\left\{\lambda_2\cos\frac{\theta}{2}-\lambda_1\sin\frac{\theta}{2}\right\}}{\sqrt{1-\rho}}} \tag{14.12}$$

for

$$x\tan\frac{\theta}{2} \leq \ y \ \leq x\cot\frac{\theta}{2}$$

and zero otherwise.

Theorem 14.1 $E(X) = \left[\frac{\cos\frac{\theta}{2}}{\sqrt{1+\rho}}\right]\frac{1}{\lambda_1} + \left[\frac{\sin\frac{\theta}{2}}{\sqrt{1+\rho}}\right]\frac{1}{\lambda_2}$; $E(Y) = \left[\frac{\cos\frac{\theta}{2}}{\sqrt{1+\rho}}\right]\frac{1}{\lambda_2} + \left[\frac{\sin\frac{\theta}{2}}{\sqrt{1+\rho}}\right]\frac{1}{\lambda_1}$.

Proof

$$E(X) = \lambda_1\lambda_2\sqrt{\frac{1+\rho}{1-\rho}}\int_0^\infty xe^{-\frac{x\left\{\lambda_1\cos\frac{\theta}{2}-\lambda_2\sin\frac{\theta}{2}\right\}}{\sqrt{1-\rho}}}\int_{x\tan\frac{\theta}{2}}^{x\cot\frac{\theta}{2}} e^{-\frac{y\left\{\lambda_2\cos\frac{\theta}{2}-\lambda_1\sin\frac{\theta}{2}\right\}}{\sqrt{1-\rho}}}\,dy\,dx$$

$$= \frac{\lambda_1\lambda_2\sqrt{1+\rho}}{\lambda_2\cos\frac{\theta}{2}-\lambda_1\sin\frac{\theta}{2}}\int_0^\infty xe^{-\frac{x\left\{\lambda_1\cos\frac{\theta}{2}-\lambda_2\sin\frac{\theta}{2}\right\}}{\sqrt{1-\rho}}}$$

$$\left\{e^{-\frac{x\tan\frac{\theta}{2}\left\{\lambda_2\cos\frac{\theta}{2}-\lambda_1\sin\frac{\theta}{2}\right\}}{\sqrt{1-\rho}}} - e^{-\frac{x\cot\frac{\theta}{2}\left\{\lambda_2\cos\frac{\theta}{2}-\lambda_1\sin\frac{\theta}{2}\right\}}{\sqrt{1-\rho}}}\right\}dx$$

$$= \frac{\lambda_1\lambda_2\sqrt{1+\rho}}{\lambda_2\cos\frac{\theta}{2}-\lambda_1\sin\frac{\theta}{2}}\left\{\frac{1-\rho}{\left[\lambda_1\cos\frac{\theta}{2}-\lambda_2\sin\frac{\theta}{2}+\tan\frac{\theta}{2}\left(\lambda_2\cos\frac{\theta}{2}-\lambda_1\sin\frac{\theta}{2}\right)\right]^2}\right.$$

$$\left.-\frac{1-\rho}{\left[\lambda_1\cos\frac{\theta}{2}-\lambda_2\sin\frac{\theta}{2}+\cot\frac{\theta}{2}\left(\lambda_2\cos\frac{\theta}{2}-\lambda_1\sin\frac{\theta}{2}\right)\right]^2}\right\}$$

$$= \frac{\lambda_1\lambda_2\sqrt{1+\rho}(1-\rho)}{\lambda_2\cos\frac{\theta}{2}-\lambda_1\sin\frac{\theta}{2}}\left\{\frac{\cos\frac{\theta}{2}}{\lambda_1\cos\theta}-\frac{\sin\frac{\theta}{2}}{\lambda_2\cos\theta}\right\}\left\{\frac{\cos\frac{\theta}{2}}{\lambda_1\cos\theta}+\frac{\sin\frac{\theta}{2}}{\lambda_2\cos\theta}\right\}$$

$$= \frac{\sqrt{1+\rho}(1-\rho)}{\sqrt{1-\rho^2}}\left\{\frac{\cos\frac{\theta}{2}}{\lambda_1\cos\theta}+\frac{\sin\frac{\theta}{2}}{\lambda_2\cos\theta}\right\}$$

$$= \frac{\lambda_2\cos\frac{\theta}{2}+\lambda_1\sin\frac{\theta}{2}}{\lambda_1\lambda_2\sqrt{1+\rho}} = \left[\frac{\cos\frac{\theta}{2}}{\sqrt{1+\rho}}\right]\frac{1}{\lambda_1}+\left[\frac{\sin\frac{\theta}{2}}{\sqrt{1+\rho}}\right]\frac{1}{\lambda_2}$$

Similarly,

$$E(Y) = \left[\frac{\cos\frac{\theta}{2}}{\sqrt{1+\rho}}\right]\frac{1}{\lambda_2}+\left[\frac{\sin\frac{\theta}{2}}{\sqrt{1+\rho}}\right]\frac{1}{\lambda_1}$$

Theorem 14.2 $Var(X) = [E(X)]^2 - \frac{\rho}{(1+\rho)\lambda_1\lambda_2}$; $Var(Y) = [E(Y)]^2 - \frac{\rho}{(1+\rho)\lambda_1\lambda_2}$.

Proof

$$E(X^2) = \lambda_1\lambda_2\sqrt{\frac{1+\rho}{1-\rho}}\int_0^\infty x^2 e^{-\frac{x\left\{\lambda_1\cos\frac{\theta}{2}-\lambda_2\sin\frac{\theta}{2}\right\}}{\sqrt{1-\rho}}}\int_{x\tan\frac{\theta}{2}}^{x\cot\frac{\theta}{2}} e^{-\frac{y\left\{\lambda_2\cos\frac{\theta}{2}-\lambda_1\sin\frac{\theta}{2}\right\}}{\sqrt{1-\rho}}}\,dy\,dx$$

$$= \frac{\lambda_1\lambda_2\sqrt{1+\rho}}{\lambda_2\cos\frac{\theta}{2}-\lambda_1\sin\frac{\theta}{2}}\int_0^\infty x^2 e^{-\frac{x\left\{\lambda_1\cos\frac{\theta}{2}-\lambda_2\sin\frac{\theta}{2}\right\}}{\sqrt{1-\rho}}}$$

$$\left\{e^{-\frac{x\tan\frac{\theta}{2}\left\{\lambda_2\cos\frac{\theta}{2}-\lambda_1\sin\frac{\theta}{2}\right\}}{\sqrt{1-\rho}}} - e^{-\frac{x\cot\frac{\theta}{2}\left\{\lambda_2\cos\frac{\theta}{2}-\lambda_1\sin\frac{\theta}{2}\right\}}{\sqrt{1-\rho}}}\right\}dx$$

$$= \frac{\lambda_1\lambda_2\sqrt{1+\rho}}{\lambda_2\cos\frac{\theta}{2}-\lambda_1\sin\frac{\theta}{2}}\left\{\frac{2(1-\rho)\sqrt{1-\rho}}{\left[\lambda_1\cos\frac{\theta}{2}-\lambda_2\sin\frac{\theta}{2}+\tan\frac{\theta}{2}\left(\lambda_2\cos\frac{\theta}{2}-\lambda_1\sin\frac{\theta}{2}\right)\right]^3}\right.$$

$$\left.-\frac{2(1-\rho)\sqrt{1-\rho}}{\left[\lambda_1\cos\frac{\theta}{2}-\lambda_2\sin\frac{\theta}{2}+\cot\frac{\theta}{2}\left(\lambda_2\cos\frac{\theta}{2}-\lambda_1\sin\frac{\theta}{2}\right)\right]^3}\right\}$$

$$= \frac{2\lambda_1\lambda_2\sqrt{1-\rho^2}(1-\rho)}{\lambda_2\cos\frac{\theta}{2}-\lambda_1\sin\frac{\theta}{2}}\left\{\frac{\cos\frac{\theta}{2}}{\lambda_1\cos\theta}-\frac{\sin\frac{\theta}{2}}{\lambda_2\cos\theta}\right\}$$

$$\left\{\frac{\cos^2\frac{\theta}{2}}{\lambda_1^2\cos^2\theta}+\frac{\sin\frac{\theta}{2}\cos\frac{\theta}{2}}{\lambda_1\lambda_2\cos^2\theta}+\frac{\sin^2\frac{\theta}{2}}{\lambda_2^2\cos^2\theta}\right\}$$

$$= 2(1-\rho)\left\{\frac{\cos^2\frac{\theta}{2}}{\lambda_1^2\cos^2\theta} + \frac{\sin\frac{\theta}{2}\cos\frac{\theta}{2}}{\lambda_1\lambda_2\cos^2\theta} + \frac{\sin^2\frac{\theta}{2}}{\lambda_2^2\cos^2\theta}\right\}$$
$$= 2\frac{\lambda_2^2\cos^2\frac{\theta}{2} + \lambda_1\lambda_2\sin\frac{\theta}{2}\cos\frac{\theta}{2} + \lambda_1^2\sin^2\frac{\theta}{2}}{\lambda_1^2\lambda_2^2(1+\rho)}$$

$$Var(X) = \frac{2\lambda_2^2\cos^2\frac{\theta}{2} + 2\lambda_1\lambda_2\sin\frac{\theta}{2}\cos\frac{\theta}{2} + 2\lambda_1^2\sin^2\frac{\theta}{2}}{\lambda_1^2\lambda_2^2(1+\rho)} - \frac{\lambda_2^2\cos^2\frac{\theta}{2} + 2\lambda_1\lambda_2\sin\frac{\theta}{2}\cos\frac{\theta}{2} + \lambda_1^2\sin^2\frac{\theta}{2}}{\lambda_1^2\lambda_2^2(1+\rho)}$$

$$Var(X) = \frac{\lambda_2^2\cos^2\frac{\theta}{2} + \lambda_1^2\sin^2\frac{\theta}{2}}{\lambda_1^2\lambda_2^2(1+\rho)} = \left[\frac{\cos\frac{\theta}{2}}{\sqrt{1+\rho}}\right]^2\frac{1}{\lambda_1^2} + \left[\frac{\sin\frac{\theta}{2}}{\sqrt{1+\rho}}\right]^2\frac{1}{\lambda_2^2}$$
$$= [E(X)]^2 - \frac{\rho}{(1+\rho)\lambda_1\lambda_2}.$$

Similarly,

$$Var(Y) = [E(Y)]^2 - \frac{\rho}{(1+\rho)\lambda_1\lambda_2}.$$

Useful Reformulations Let μ_1 and μ_2 be the means of X and Y, respectively. Using Theorem 14.1,

$$\left[\frac{\cos\frac{\theta}{2}}{\sqrt{1+\rho}}\right]\frac{1}{\lambda_1} + \left[\frac{\sin\frac{\theta}{2}}{\sqrt{1+\rho}}\right]\frac{1}{\lambda_2} = \mu_1; \quad \left[\frac{\sin\frac{\theta}{2}}{\sqrt{1+\rho}}\right]\frac{1}{\lambda_1} + \left[\frac{\cos\frac{\theta}{2}}{\sqrt{1+\rho}}\right]\frac{1}{\lambda_2} = \mu_2.$$

$$\sqrt{1-\rho^2}\frac{1}{\lambda_1} = \sqrt{1+\rho}(\mu_1\cos\frac{\theta}{2} - \mu_2\sin\frac{\theta}{2})$$

$$\lambda_1 = \frac{\sqrt{1-\rho}}{\mu_1\cos\frac{\theta}{2} - \mu_2\sin\frac{\theta}{2}}; \quad \lambda_2 = \frac{\sqrt{1-\rho}}{\mu_2\cos\frac{\theta}{2} - \mu_1\sin\frac{\theta}{2}}.$$

There is a limit to the ratio of the two means to yield a positive λ_1 and λ_2: $\tan\frac{\theta}{2} < \frac{\mu_1}{\mu_2} < \cot\frac{\theta}{2}$. Therefore, the unknown parameters λ_1 and λ_2 can be easily computed when real data on (X, Y) is available. Further, note that

$$\lambda_1\lambda_2 = \frac{1-\rho}{\mu_1\mu_2 - \frac{\rho}{2}(\mu_1^2 + \mu_2^2)}$$

and applying Theorem 14.2,

$$Var(X) = \mu_1^2 - \frac{\rho}{(1+\rho)\lambda_1\lambda_2} = \mu_1^2 - \frac{\rho}{1-\rho^2}\left[\mu_1\mu_2 - \frac{\rho}{2}(\mu_1^2+\mu_2^2)\right]$$
$$Var(Y) = \mu_2^2 - \frac{\rho}{(1+\rho)\lambda_1\lambda_2} = \mu_2^2 - \frac{\rho}{1-\rho^2}\left[\mu_1\mu_2 - \frac{\rho}{2}(\mu_1^2+\mu_2^2)\right]. \quad (14.13)$$

14.3.1 Expected Maximum and Expected Minimum

We provide analytical expressions for the expected maximum and expected minimum for (X, Y) that are generated from $f_{X,Y}(x, y)$ in (14.12). First, we prove that for any two random variables (that may be correlated) the sum of the expected minimum and expected maximum is equal to the sum of the means. This result allows us to just develop a formula for the expected maximum and consequently also obtain the formula for the expected minimum.

Theorem 14.3 $E[\min\{X, Y\}] + E[\max\{X, Y\}] = E(X) + E(Y)$.

Proof Consider two discrete probability distributions with n points each (x_i, p_i^x) and (y_i, p_i^y) which may be correlated so that the probability of (x_i, y_j) is p_{ij} which is not necessarily equal to $p_i^x p_j^y$. However, $\sum_{i=1}^{n} p_{ij} = p_j^y\ \forall j$ and $\sum_{j=1}^{n} p_{ij} = p_i^x\ \forall i$. Consider all pairs (x_i, y_j). For each pair one of them is the minimum and the other one is the maximum. Therefore, $\min\{x_i, y_j\} + \max\{x_i, y_j\} = x_i + y_j$. We get that

$$\begin{aligned}
E[\min\{X, Y\}] + E[\max\{X, Y\}] &= E[\min\{X, Y\} + \max\{X, Y\}] \\
&= \sum_{i,j=1}^{n} \left[\min\{x_i, y_j\} + \max\{x_i, y_j\}\right] p_{ij} \\
&= \sum_{i,j=1}^{n} (x_i + y_j) p_{ij} \\
&= \sum_{i=1}^{n}\left[\sum_{j=1}^{n} p_{ij}\right] x_i + \sum_{j=1}^{n}\left[\sum_{i=1}^{n} p_{ij}\right] y_j \\
&= \sum_{i=1}^{n} p_i^x x_i + \sum_{j=1}^{n} p_j^y y_j = E(X) + E(Y)
\end{aligned}$$

The result is true for any continuous distribution as well. A continuous distribution is a limit of a discrete probability distribution when $n \to \infty$. For example, Nadarajah and Kotz (2008) give exact formulas for the expected maximum value and minimum value of two correlated normal random variables. Let (X, Y) have means of (μ_1, μ_2) with standard deviations (σ_1, σ_2), and correlation coefficient ρ. Define $\sigma = \sqrt{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2}$. Then,

$$E(\max\{X, Y\}) = \mu_1 \Phi\left(\frac{\mu_1 - \mu_2}{\sigma}\right) + \mu_2 \Phi\left(\frac{\mu_2 - \mu_1}{\sigma}\right) + \sigma\phi\left(\frac{\mu_1 - \mu_2}{\sigma}\right) \tag{14.14}$$

$$E(\min\{X, Y\}) = \mu_1 \Phi\left(\frac{\mu_2 - \mu_2}{\sigma}\right) + \mu_2 \Phi\left(\frac{\mu_1 - \mu_2}{\sigma}\right) - \sigma\phi\left(\frac{\mu_1 - \mu_2}{\sigma}\right) \tag{14.15}$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ are the density function and the cumulative distribution of the standardized normal distribution. For example, if $\mu_1 = \mu_2 = \mu$, then $E(\max\{X, Y\}) = \mu + \frac{\sigma}{\sqrt{2\pi}}$ and $E(\min\{X, Y\}) = \mu - \frac{\sigma}{\sqrt{2\pi}}$.

For a bivariate normal distribution $E[\min\{X, Y\}] + E[\max\{X, Y\}] = \mu_1 + \mu_2$ because $\Phi(a) + \Phi(-a) = 1$, confirming Theorem 14.3 in this case.

Now, we consider the case of correlated exponentials. The probability that the maximum value $\max(X, Y) \leq t$ is equal to the probability that $X \leq t$ and $Y \leq t$. This probability is the integral of the density function $f_{X,Y}(x, y)$ over the area depicted in Fig. 14.2.

Let $F(t)$ be $Pr(\max(X, Y) \leq t)$.

$$F(t) = \int_0^{t \tan\frac{\theta}{2}} \int_{x \tan\frac{\theta}{2}}^{x \cot\frac{\theta}{2}} \phi(x, y)\, dy\, dx + \int_{t \tan\frac{\theta}{2}}^{t} \int_{x \tan\frac{\theta}{2}}^{t} \phi(x, y)\, dy\, dx$$

Substituting the density function (14.12):

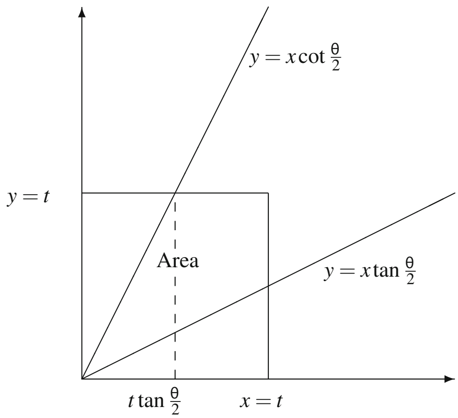


Fig. 14.2 The integration area

$$F(t) = \int_0^{t\tan\frac{\theta}{2}} \int_{x\tan\frac{\theta}{2}}^{x\cot\frac{\theta}{2}} \lambda_1\lambda_2\sqrt{\frac{1+\rho}{1-\rho}} e^{-\frac{X\left\{\lambda_1\cos\frac{\theta}{2}-\lambda_2\sin\frac{\theta}{2}\right\}+Y\left\{\lambda_2\cos\frac{\theta}{2}-\lambda_1\sin\frac{\theta}{2}\right\}}{\sqrt{1-\rho}}}\,dy\,dx$$

$$+\int_{t\tan\frac{\theta}{2}}^{t} \int_{x\tan\frac{\theta}{2}}^{t} \lambda_1\lambda_2\sqrt{\frac{1+\rho}{1-\rho}} e^{-\frac{X\left\{\lambda_1\cos\frac{\theta}{2}-\lambda_2\sin\frac{\theta}{2}\right\}+Y\left\{\lambda_2\cos\frac{\theta}{2}-\lambda_1\sin\frac{\theta}{2}\right\}}{\sqrt{1-\rho}}}\,dy\,dx$$

Once $F(t)$ is found, then $E[\max(X,Y)] = \int_0^\infty t\frac{dF(t)}{dt}\,dt$ because $\int_0^\infty \frac{dF(t)}{dt}\,dt = 1$. In the Appendix we show that

$$E[\max(X,Y)] = \left(\frac{1}{\lambda_1}+\frac{1}{\lambda_2}\right)\frac{1+\sqrt{1-\rho^2}}{2} - \frac{1}{\lambda_1+\lambda_2}\left\{\frac{\sqrt{1-\rho^2}}{1-\rho\frac{\lambda_1^2+\lambda_2^2}{2\lambda_1\lambda_2}}\right\}$$
$$+\left(\frac{1}{\lambda_1}+\frac{1}{\lambda_2}\right)\rho^2\frac{(\lambda_1^2+\lambda_2^2)\sqrt{1-\rho}}{4\sqrt{1+\rho}(\lambda_1\lambda_2-\frac{\rho}{2}(\lambda_1^2+\lambda_2^2))} \tag{14.16}$$

$$E[\min(X,Y)] = \left(\frac{1}{\lambda_1}+\frac{1}{\lambda_2}\right)\frac{1-\sqrt{1-\rho^2}}{2} + \frac{1}{\lambda_1+\lambda_2}\left\{\frac{\sqrt{1-\rho^2}}{1-\rho\frac{\lambda_1^2+\lambda_2^2}{2\lambda_1\lambda_2}}\right\}$$
$$-\left(\frac{1}{\lambda_1}+\frac{1}{\lambda_2}\right)\rho^2\frac{(\lambda_1^2+\lambda_2^2)\sqrt{1-\rho}}{4\sqrt{1+\rho}(\lambda_1\lambda_2-\frac{\rho}{2}(\lambda_1^2+\lambda_2^2))} \tag{14.17}$$

The formulas are valid for $\lambda_1\lambda_2-\frac{\rho}{2}(\lambda_1^2+\lambda_2^2)>0$ or $\rho<\frac{2\lambda_1\lambda_2}{\lambda_1^2+\lambda_2^2}=1-\frac{(\lambda_1-\lambda_2)^2}{\lambda_1^2+\lambda_2^2}$. For example, when $\lambda_2=2\lambda_1$, ρ should be lower than 0.8.

14.3.1.1 Special Cases

When $\rho=0$,

$$E[\max(X,Y)] = \frac{1}{\lambda_1}+\frac{1}{\lambda_2}-\frac{1}{\lambda_1+\lambda_2};\quad E[\min(X,Y)] = \frac{1}{\lambda_1+\lambda_2}$$

When $\lambda_1=\lambda_2=\lambda$:

$$E[\max(X,Y)] = \frac{1}{\lambda}\left\{1+\sqrt{1-\rho^2}-\frac{\sqrt{1-\rho^2}}{2(1-\rho)}+\frac{\rho^2\sqrt{1-\rho}}{(1-\rho)\sqrt{1+\rho}}\right\}$$

$$= \frac{1}{\lambda}\left\{1 - \frac{1}{2}\sqrt{\frac{1+\rho}{1-\rho}} + \sqrt{1-\rho^2} + \frac{\rho^2}{\sqrt{1-\rho^2}}\right\} = \frac{1}{\lambda}\left\{1 - \frac{1}{2}\sqrt{\frac{1+\rho}{1-\rho}} + \frac{1}{\sqrt{1-\rho^2}}\right\}$$

$$= \frac{1}{\lambda}\left\{1 + \frac{1-\frac{1}{2}(1+\rho)}{\sqrt{1-\rho^2}}\right\} = \frac{1}{\lambda}\left\{1 + \frac{1}{2}\frac{1-\rho}{\sqrt{1-\rho^2}}\right\} = \frac{1}{\lambda}\left\{1 + \frac{1}{2}\sqrt{\frac{1-\rho}{1+\rho}}\right\}$$

and

$$E[\min(X,Y)] = \frac{1}{\lambda}\left\{1 - \frac{1}{2}\sqrt{\frac{1-\rho}{1+\rho}}\right\}$$

14.4 Correlated Service Times in Tandem Queues

To illustrate the possible application of the proposed bivariate exponential distribution, we consider a queueing system in which two servers process the same arriving customer in tandem (or in sequence), i.e., each customer is serviced at the first server and then proceeds to the second server before leaving the system. Since each customer must be processed by both servers, service times at the two servers for a given customer may become correlated. When service times follow (14.12), we provide analytical expression for the total waiting time in the system.

An M/G/1 queue has an arrival rate λ, average service time μ with a variance σ^2. Note that in most literature μ is defined as average service rate (1/service time). The average time a customer spends in the system (time in queue plus service time) is by Pollaczek's formula Gelenbe et al. (1998):

$$W = \mu + \frac{\lambda\sigma^2 + \lambda\mu^2}{2(1-\lambda\mu)} = \frac{\mu}{1-\lambda\mu} + \mu - \frac{\mu}{1-\lambda\mu} + \frac{\lambda\sigma^2 + \lambda\mu^2}{2(1-\lambda\mu)}$$
$$= \frac{\mu}{1-\lambda\mu} + \lambda\frac{\sigma^2 - \mu^2}{2(1-\lambda\mu)}. \tag{14.18}$$

Now consider two sequential servers. The arrival rate must be lower than the slower service rate between the two servers. The arrival rate is the same for both servers even when, for example, the first server is faster. Let the average service time for the first server be μ_1 with variance σ_1^2 and the second server has a mean service time μ_2 with variance σ_2^2, by (14.13):

$$\sigma_1^2 = \mu_1^2 - \frac{\rho}{1-\rho^2}\left[\mu_1\mu_2 - \frac{\rho}{2}(\mu_1^2+\mu_2^2)\right]$$
$$= \mu_1^2 - \frac{\rho}{1-\rho^2}\left[(1-\rho)\mu_1\mu_2 - \frac{\rho}{2}(\mu_1-\mu_2)^2\right]$$
$$= \mu_1^2 - \frac{\rho}{1+\rho}\mu_1\mu_2 + \frac{\rho^2}{2(1-\rho^2)}(\mu_1-\mu_2)^2$$
$$= \sigma_1^2 - \mu_1^2 = \frac{\rho}{1+\rho}\left\{\frac{\rho}{2(1-\rho)}(\mu_1-\mu_2)^2 - \mu_1\mu_2\right\}$$

and thus the total time W_1 is by (14.18):

$$W_1 = \frac{\mu_1}{1-\lambda\mu_1} + \frac{\rho\lambda}{2(1+\rho)(1-\lambda\mu_1)}\left\{\frac{\rho}{2(1-\rho)}(\mu_1-\mu_2)^2 - \mu_1\mu_2\right\}$$
$$= \frac{\mu_1}{1-\lambda\mu_1} + \frac{\rho\lambda\mu_1\mu_2}{2(1-\rho^2)(1-\lambda\mu_1)}\left\{\frac{\rho}{2}\left(\frac{\mu_1}{\mu_2}+\frac{\mu_2}{\mu_1}\right) - 1\right\} \tag{14.19}$$

and $W_1 + W_2$ is

$$W_1 + W_2 = \frac{\mu_1}{1-\lambda\mu_1} + \frac{\mu_2}{1-\lambda\mu_2} + \frac{\rho\lambda\mu_1\mu_2}{2(1-\rho^2)}$$
$$\left\{\frac{\rho}{2}\left(\frac{\mu_1}{\mu_2}+\frac{\mu_2}{\mu_1}\right) - 1\right\}\left\{\frac{1}{1-\lambda\mu_1} + \frac{1}{1-\lambda\mu_2}\right\} \tag{14.20}$$

The waiting time in line for an uncorrelated server (1 or 2) is $W_q = \frac{\lambda\mu^2}{1-\lambda\mu}$. The reduction in total time can be written as

$$\frac{\rho\lambda\mu_1\mu_2}{2(1-\rho^2)}\left\{1 - \frac{\rho}{2}\left(\frac{\mu_1}{\mu_2}+\frac{\mu_2}{\mu_1}\right)\right\}\left\{\frac{1}{1-\lambda\mu_1} + \frac{1}{1-\lambda\mu_2}\right\}$$
$$= \frac{\rho\left\{1 - \frac{\rho}{2}\left(\frac{\mu_1}{\mu_2}+\frac{\mu_2}{\mu_1}\right)\right\}}{2(1-\rho^2)}\left\{\frac{\mu_1}{\mu_2}W_{q1} + \frac{\mu_2}{\mu_1}W_{q2}\right\}$$

leading to:

$$W_1 + W_2 = \frac{\mu_1}{1-\lambda\mu_1} + \frac{\mu_2}{1-\lambda\mu_2} - \frac{\rho\left\{1 - \frac{\rho}{2}\left(\frac{\mu_1}{\mu_2}+\frac{\mu_2}{\mu_1}\right)\right\}}{2(1-\rho^2)}\left\{\frac{\mu_1}{\mu_2}W_{q1} + \frac{\mu_2}{\mu_1}W_{q2}\right\} \tag{14.21}$$

If there are two sequential servers with means μ_1 and μ_2 and correlation ρ between them satisfying $\tan\frac{\theta}{2} < \frac{\mu_1}{\mu_2} < \cot\frac{\theta}{2}$, then the total service time is given by (14.20) or (14.21).

Theorem 14.4 *For* $\tan\frac{\theta}{2} < \frac{\mu_1}{\mu_2} < \cot\frac{\theta}{2}$, $\frac{\rho}{2}\left(\frac{\mu_1}{\mu_2}+\frac{\mu_2}{\mu_1}\right) < 1$ *and thus total service time declines for* $\rho > 0$.

Proof Consider $\mu_1 \geq \mu_2$. Define $\phi = \frac{\mu_1}{\mu_2} \geq 1$. We get $\frac{\rho}{2}\left(\frac{\mu_1}{\mu_2}+\frac{\mu_2}{\mu_1}\right) = \frac{\rho}{2}\left(\phi+\frac{1}{\phi}\right)$. The function $\phi+\frac{1}{\phi}$ increases when $\phi \geq 1$ increases and obtains its maximum value at the largest possible ϕ which is $\phi = \cot\frac{\theta}{2}$. Therefore, for $1 \leq \phi < \cot\frac{\theta}{2}$

$$\frac{\rho}{2}\left(\phi+\frac{1}{\phi}\right) < \frac{\rho}{2}\left(\cot\frac{\theta}{2}+\tan\frac{\theta}{2}\right)$$
$$= \sin\frac{\theta}{2}\cos\frac{\theta}{2}\left(\cot\frac{\theta}{2}+\tan\frac{\theta}{2}\right) = \cos^2\frac{\theta}{2}+\sin^2\frac{\theta}{2} = 1,$$

which proves the theorem.

Lemma 14.1 $\rho > \tan\frac{\theta}{2}$.

Proof $\rho = 2\sin\frac{\theta}{2}\cos\frac{\theta}{2} = 2\tan\frac{\theta}{2}\cos^2\frac{\theta}{2}$ and $\cos\frac{\theta}{2} \geq \frac{\sqrt{2}}{2}$ because $\theta \leq \frac{\pi}{2}$.

Theorem 14.5 *The maximum reduction in waiting time is obtained for* $\rho = \frac{\min\{\mu_1,\mu_2\}}{\max\{\mu_1,\mu_2\}}$.

Proof Define $\phi = \frac{\mu_1}{\mu_2}$. For a given ϕ the maximum reduction is obtained by:

$$\frac{d}{d\rho}\frac{\rho\left\{1-\frac{\rho}{2}\left(\phi+\frac{1}{\phi}\right)\right\}}{2(1-\rho^2)}$$
$$= \frac{2\left[1-\rho\left(\phi+\frac{1}{\phi}\right)\right](1-\rho^2)+4\rho^2\left\{1-\frac{\rho}{2}\left(\phi+\frac{1}{\phi}\right)\right\}}{4(1-\rho^2)^2} = 0.$$

Leading to:

$$1-\rho\left(\phi+\frac{1}{\phi}\right)-\rho^2+\rho^3\left(\phi+\frac{1}{\phi}\right)+2\rho^2-\rho^3\left(\phi+\frac{1}{\phi}\right) = 0.$$

$$1-\rho\left(\phi+\frac{1}{\phi}\right)+\rho^2 = 0.$$

$$\rho = \frac{\left(\phi+\frac{1}{\phi}\right)\pm\sqrt{\left(\phi+\frac{1}{\phi}\right)^2-4}}{2} = \phi \text{ or } \frac{1}{\phi} \text{ depending on whether } \phi \geq 1 \text{ or } \phi \leq 1.$$

By Lemma 14.1, the maximum reduction is obtained for $\rho = \phi$ or $\frac{1}{\phi}$. Since $\rho \leq 1$, $\rho = \frac{\min\{\mu_1,\mu_2\}}{\max\{\mu_1,\mu_2\}}$.

The maximum reduction in total waiting time for $\mu_2 \geq \mu_1$ is

$$\frac{\rho\left\{1-\frac{\rho}{2}\left(\rho+\frac{1}{\rho}\right)\right\}}{2(1-\rho^2)}\left\{\rho W_{q1}+\frac{1}{\rho}W_{q2}\right\}=\frac{\rho}{4}\left\{\rho W_{q1}+\frac{1}{\rho}W_{q2}\right\}$$
$$=\frac{1}{4}\left\{\rho^2 W_{q1}+W_{q2}\right\} \quad (14.22)$$

and in general for any μ_1, μ_2:

$$\text{Maximum reduction is:} \quad \frac{\mu_1^2 W_{q1}+\mu_2^2 W_{q2}}{4\max\{\mu_1,\mu_2\}^2} \quad (14.23)$$

obtained for $\rho = \frac{\min\{\mu_1,\mu_2\}}{\max\{\mu_1,\mu_2\}}$. The maximum reduction in waiting time in line is greater for the server with the longer average service time.

In the special case of $\mu_1 = \mu_2 = \mu$ the waiting time in line for one server for independent servers is $\frac{\lambda\mu^2}{1-\lambda\mu}$. Therefore, total time in the system is reduced by a fraction $\frac{0.5\rho}{1+\rho}$ of the waiting time in line for both servers combined. This fraction is a monotonically increasing function of ρ with a maximum of 0.25 for $\rho = 1$. Note that when $\mu_1 = \mu_2$ then the condition $\tan\frac{\theta}{2} < \frac{\mu_1}{\mu_2} < \cot\frac{\theta}{2}$ is satisfied for every $\rho \geq 0$.

Appendix: Proofs of (14.16) and (14.17)

The first integral is

$$I_1(t)=\lambda_1\lambda_2\sqrt{\frac{1+\rho}{1-\rho}}\int_0^{t\tan\frac{\theta}{2}} e^{-\frac{X\left\{\lambda_1\cos\frac{\theta}{2}-\lambda_2\sin\frac{\theta}{2}\right\}}{\sqrt{1-\rho}}}\int_{x\tan\frac{\theta}{2}}^{x\cot\frac{\theta}{2}} e^{-\frac{Y\left\{\lambda_2\cos\frac{\theta}{2}-\lambda_1\sin\frac{\theta}{2}\right\}}{\sqrt{1-\rho}}}\,dy\,dx$$

$$=\frac{\lambda_1\lambda_2\sqrt{1+\rho}}{\lambda_2\cos\frac{\theta}{2}-\lambda_1\sin\frac{\theta}{2}}\int_0^{t\tan\frac{\theta}{2}} e^{-\frac{X\left\{\lambda_1\cos\frac{\theta}{2}-\lambda_2\sin\frac{\theta}{2}\right\}}{\sqrt{1-\rho}}}$$
$$\left\{e^{-\frac{x\tan\frac{\theta}{2}\left\{\lambda_2\cos\frac{\theta}{2}-\lambda_1\sin\frac{\theta}{2}\right\}}{\sqrt{1-\rho}}}-e^{-\frac{x\cot\frac{\theta}{2}\left\{\lambda_2\cos\frac{\theta}{2}-\lambda_1\sin\frac{\theta}{2}\right\}}{\sqrt{1-\rho}}}\right\}dx$$

$$=\frac{\lambda_1\lambda_2\sqrt{1+\rho}}{\lambda_2\cos\frac{\theta}{2}-\lambda_1\sin\frac{\theta}{2}}\int_0^{t\tan\frac{\theta}{2}}\left\{e^{-\frac{x\lambda_1\left\{\cos^2\frac{\theta}{2}-\sin^2\frac{\theta}{2}\right\}}{\sqrt{1-\rho}\cos\frac{\theta}{2}}}-e^{-\frac{x\lambda_2\left\{\cos^2\frac{\theta}{2}-\sin^2\frac{\theta}{2}\right\}}{\sqrt{1-\rho}\sin\frac{\theta}{2}}}\right\}dx$$

$$= \frac{\lambda_1\lambda_2\sqrt{1+\rho}}{\lambda_2\cos\frac{\theta}{2} - \lambda_1\sin\frac{\theta}{2}} \int\limits_0^{t\tan\frac{\theta}{2}} \left\{ e^{-\frac{x\lambda_1\sqrt{1+\rho}}{\cos\frac{\theta}{2}}} - e^{-\frac{x\lambda_2\sqrt{1+\rho}}{\sin\frac{\theta}{2}}} \right\} dx$$

$$= \frac{\lambda_1\lambda_2\sqrt{1+\rho}}{\lambda_2\cos\frac{\theta}{2} - \lambda_1\sin\frac{\theta}{2}}$$

$$\left\{ \frac{\cos\frac{\theta}{2}}{\lambda_1\sqrt{1+\rho}} \left[1 - e^{-\frac{t\tan\frac{\theta}{2}\lambda_1\sqrt{1+\rho}}{\cos\frac{\theta}{2}}} \right] - \frac{\sin\frac{\theta}{2}}{\lambda_2\sqrt{1+\rho}} \left[1 - e^{-\frac{t\lambda_2\sqrt{1+\rho}}{\cos\frac{\theta}{2}}} \right] \right\}$$

The second integral is

$$I_2(t) = \lambda_1\lambda_2\sqrt{\frac{1+\rho}{1-\rho}} \int\limits_{t\tan\frac{\theta}{2}}^{t} e^{-\frac{X\left\{\lambda_1\cos\frac{\theta}{2} - \lambda_2\sin\frac{\theta}{2}\right\}}{\sqrt{1-\rho}}} \int\limits_{x\tan\frac{\theta}{2}}^{t} e^{-\frac{Y\left\{\lambda_2\cos\frac{\theta}{2} - \lambda_1\sin\frac{\theta}{2}\right\}}{\sqrt{1-\rho}}} \, dy\, dx$$

$$= \frac{\lambda_1\lambda_2\sqrt{1+\rho}}{\lambda_2\cos\frac{\theta}{2} - \lambda_1\sin\frac{\theta}{2}} \int\limits_{t\tan\frac{\theta}{2}}^{t} e^{-\frac{X\left\{\lambda_1\cos\frac{\theta}{2} - \lambda_2\sin\frac{\theta}{2}\right\}}{\sqrt{1-\rho}}}$$

$$\left\{ e^{-\frac{x\tan\frac{\theta}{2}\left\{\lambda_2\cos\frac{\theta}{2} - \lambda_1\sin\frac{\theta}{2}\right\}}{\sqrt{1-\rho}}} - e^{-\frac{t\left\{\lambda_2\cos\frac{\theta}{2} - \lambda_1\sin\frac{\theta}{2}\right\}}{\sqrt{1-\rho}}} \right\} dx$$

$$= \frac{\lambda_1\lambda_2\sqrt{1+\rho}}{\lambda_2\cos\frac{\theta}{2} - \lambda_1\sin\frac{\theta}{2}}$$

$$\left\{ \int\limits_{t\tan\frac{\theta}{2}}^{t} e^{-\frac{x\lambda_1\sqrt{1+\rho}}{\cos\frac{\theta}{2}}} dx - e^{-\frac{t\left\{\lambda_2\cos\frac{\theta}{2} - \lambda_1\sin\frac{\theta}{2}\right\}}{\sqrt{1-\rho}}} \int\limits_{t\tan\frac{\theta}{2}}^{t} e^{-\frac{X\left\{\lambda_1\cos\frac{\theta}{2} - \lambda_2\sin\frac{\theta}{2}\right\}}{\sqrt{1-\rho}}} dx \right\}$$

$$= \frac{\lambda_1\lambda_2\sqrt{1+\rho}}{\lambda_2\cos\frac{\theta}{2} - \lambda_1\sin\frac{\theta}{2}} \left\{ \frac{\cos\frac{\theta}{2}}{\lambda_1\sqrt{1+\rho}} \left[e^{-\frac{t\tan\frac{\theta}{2}\lambda_1\sqrt{1+\rho}}{\cos\frac{\theta}{2}}} - e^{-\frac{t\lambda_1\sqrt{1+\rho}}{\cos\frac{\theta}{2}}} \right] \right.$$

$$- \frac{\sqrt{1-\rho}}{\lambda_1\cos\frac{\theta}{2} - \lambda_2\sin\frac{\theta}{2}} e^{-\frac{t\left\{\lambda_2\cos\frac{\theta}{2} - \lambda_1\sin\frac{\theta}{2}\right\}}{\sqrt{1-\rho}}}$$

$$\left. \left[e^{-\frac{t\tan\frac{\theta}{2}\left\{\lambda_1\cos\frac{\theta}{2} - \lambda_2\sin\frac{\theta}{2}\right\}}{\sqrt{1-\rho}}} - e^{-\frac{t\left\{\lambda_1\cos\frac{\theta}{2} - \lambda_2\sin\frac{\theta}{2}\right\}}{\sqrt{1-\rho}}} \right] \right\}$$

$$= \frac{\lambda_1\lambda_2\sqrt{1+\rho}}{\lambda_2\cos\frac{\theta}{2}-\lambda_1\sin\frac{\theta}{2}}\left\{\frac{\cos\frac{\theta}{2}}{\lambda_1\sqrt{1+\rho}}\left[e^{-\frac{t\tan\frac{\theta}{2}\lambda_1\sqrt{1+\rho}}{\cos\frac{\theta}{2}}}-e^{-\frac{t\lambda_1\sqrt{1+\rho}}{\cos\frac{\theta}{2}}}\right]\right.$$
$$\left.-\frac{\sqrt{1-\rho}}{\lambda_1\cos\frac{\theta}{2}-\lambda_2\sin\frac{\theta}{2}}\left[e^{-\frac{t\lambda_2\sqrt{1+\rho}}{\cos\frac{\theta}{2}}}-e^{-t(\lambda_1+\lambda_2)}\right]\right\}$$

Therefore,

$$F(t)=I_1(t)+I_2(t)=\frac{\lambda_1\lambda_2\sqrt{1+\rho}}{\lambda_2\cos\frac{\theta}{2}-\lambda_1\sin\frac{\theta}{2}}\left\{\frac{\cos\frac{\theta}{2}}{\lambda_1\sqrt{1+\rho}}\left[1-e^{-\frac{t\tan\frac{\theta}{2}\lambda_1\sqrt{1+\rho}}{\cos\frac{\theta}{2}}}\right]\right.$$
$$-\frac{\sin\frac{\theta}{2}}{\lambda_2\sqrt{1+\rho}}\left[1-e^{-\frac{t\lambda_2\sqrt{1+\rho}}{\cos\frac{\theta}{2}}}\right]$$
$$+\frac{\cos\frac{\theta}{2}}{\lambda_1\sqrt{1+\rho}}\left[e^{-\frac{t\tan\frac{\theta}{2}\lambda_1\sqrt{1+\rho}}{\cos\frac{\theta}{2}}}-e^{-\frac{t\lambda_1\sqrt{1+\rho}}{\cos\frac{\theta}{2}}}\right]$$
$$\left.-\frac{\sqrt{1-\rho}}{\lambda_1\cos\frac{\theta}{2}-\lambda_2\sin\frac{\theta}{2}}\left[e^{-\frac{t\lambda_2\sqrt{1+\rho}}{\cos\frac{\theta}{2}}}-e^{-t(\lambda_1+\lambda_2)}\right]\right\}$$
$$=\frac{\lambda_1\lambda_2\sqrt{1+\rho}}{\lambda_2\cos\frac{\theta}{2}-\lambda_1\sin\frac{\theta}{2}}\left\{\frac{\cos\frac{\theta}{2}}{\lambda_1\sqrt{1+\rho}}\left[1-e^{-\frac{t\lambda_1\sqrt{1+\rho}}{\cos\frac{\theta}{2}}}\right]\right.$$
$$-\frac{\sin\frac{\theta}{2}}{\lambda_2\sqrt{1+\rho}}\left[1-e^{-\frac{t\lambda_2\sqrt{1+\rho}}{\cos\frac{\theta}{2}}}\right]$$
$$\left.-\frac{\sqrt{1-\rho}}{\lambda_1\cos\frac{\theta}{2}-\lambda_2\sin\frac{\theta}{2}}\left[e^{-\frac{t\lambda_2\sqrt{1+\rho}}{\cos\frac{\theta}{2}}}-e^{-t(\lambda_1+\lambda_2)}\right]\right\}$$

We get

$$\frac{dF(t)}{dt}=\frac{\lambda_1\lambda_2\sqrt{1+\rho}}{\lambda_2\cos\frac{\theta}{2}-\lambda_1\sin\frac{\theta}{2}}\left\{e^{-\frac{t\lambda_1\sqrt{1+\rho}}{\cos\frac{\theta}{2}}}-\tan\frac{\theta}{2}e^{-\frac{t\lambda_2\sqrt{1+\rho}}{\cos\frac{\theta}{2}}}\right.$$
$$\left.+\frac{\lambda_2\sqrt{1-\rho^2}}{(\lambda_1\cos\frac{\theta}{2}-\lambda_2\sin\frac{\theta}{2})\cos\frac{\theta}{2}}e^{-\frac{t\lambda_2\sqrt{1+\rho}}{\cos\frac{\theta}{2}}}-\frac{\sqrt{1-\rho}(\lambda_1+\lambda_2)}{\lambda_1\cos\frac{\theta}{2}-\lambda_2\sin\frac{\theta}{2}}e^{-t(\lambda_1+\lambda_2)}\right\}$$
$$\int_0^\infty\frac{dF(t)}{dt}t\,dt=\frac{\lambda_1\lambda_2\sqrt{1+\rho}}{\lambda_2\cos\frac{\theta}{2}-\lambda_1\sin\frac{\theta}{2}}\left\{\frac{\cos^2\frac{\theta}{2}}{(1+\rho)\lambda_1^2}-\frac{\tan\frac{\theta}{2}\cos^2\frac{\theta}{2}}{(1+\rho)\lambda_2^2}\right.$$

$$+\frac{\sqrt{1-\rho^2}\cos\frac{\theta}{2}}{(\lambda_1\cos\frac{\theta}{2}-\lambda_2\sin\frac{\theta}{2})\lambda_2(1+\rho)}-\frac{\sqrt{1-\rho}}{(\lambda_1\cos\frac{\theta}{2}-\lambda_2\sin\frac{\theta}{2})(\lambda_1+\lambda_2)}\Bigg\}$$

$$\begin{aligned}
E[\max(X,Y)] &= \frac{\lambda_1\lambda_2}{(\lambda_2\cos\frac{\theta}{2}-\lambda_1\sin\frac{\theta}{2})\sqrt{1+\rho}}\left\{\frac{\cos^2\frac{\theta}{2}}{\lambda_1^2}-\frac{\rho}{2\lambda_2^2}\right\}\\
&\quad+\frac{\lambda_1\sqrt{1-\rho}\cos\frac{\theta}{2}}{(\lambda_2\cos\frac{\theta}{2}-\lambda_1\sin\frac{\theta}{2})(\lambda_1\cos\frac{\theta}{2}-\lambda_2\sin\frac{\theta}{2})}\\
&\quad-\frac{\lambda_1\lambda_2\sqrt{1-\rho^2}}{(\lambda_2\cos\frac{\theta}{2}-\lambda_1\sin\frac{\theta}{2})(\lambda_1\cos\frac{\theta}{2}-\lambda_2\sin\frac{\theta}{2})(\lambda_1+\lambda_2)}\\
&= \frac{\lambda_1\lambda_2}{(\lambda_2\cos\frac{\theta}{2}-\lambda_1\sin\frac{\theta}{2})\sqrt{1+\rho}}\left\{\frac{\cos^2\frac{\theta}{2}}{\lambda_1^2}-\frac{\rho}{2\lambda_2^2}\right\}\\
&\quad+\frac{\lambda_1\sqrt{1-\rho}\cos\frac{\theta}{2}}{\lambda_1\lambda_2-\frac{\rho}{2}(\lambda_1^2+\lambda_2^2)}-\frac{\lambda_1\lambda_2\sqrt{1-\rho^2}}{(\lambda_1\lambda_2-\frac{\rho}{2}(\lambda_1^2+\lambda_2^2))(\lambda_1+\lambda_2)}\\
&= \frac{1}{(\lambda_1\lambda_2-\frac{\rho}{2}(\lambda_1^2+\lambda_2^2))\sqrt{1+\rho}}\\
&\quad\left\{\frac{\lambda_1\cos\frac{\theta}{2}-\lambda_2\sin\frac{\theta}{2}}{\lambda_1\lambda_2}\left\{\lambda_2^2\cos^2\frac{\theta}{2}-\frac{\rho}{2}\lambda_1^2\right\}\right.\\
&\quad\left.+\lambda_1\sqrt{1-\rho^2}\cos\frac{\theta}{2}-\frac{\lambda_1\lambda_2\sqrt{(1-\rho^2)(1+\rho)}}{(\lambda_1+\lambda_2)}\right\}\\
&= \frac{1}{(\lambda_1\lambda_2-\frac{\rho}{2}(\lambda_1^2+\lambda_2^2))\sqrt{1+\rho}}\\
&\quad\left\{\lambda_2\cos^3\frac{\theta}{2}-\frac{\lambda_2^2}{\lambda_1}\cos^2\frac{\theta}{2}\sin\frac{\theta}{2}-\frac{\rho}{2}\frac{\lambda_1^2}{\lambda_2}\cos\frac{\theta}{2}+\frac{\rho}{2}\lambda_1\sin\frac{\theta}{2}\right.\\
&\quad\left.+\lambda_1\sqrt{1-\rho^2}\cos\frac{\theta}{2}-\frac{\lambda_1\lambda_2\sqrt{(1-\rho^2)(1+\rho)}}{(\lambda_1+\lambda_2)}\right\}\\
&= \frac{1}{(\lambda_1\lambda_2-\frac{\rho}{2}(\lambda_1^2+\lambda_2^2))\sqrt{1+\rho}}\\
&\quad\left\{(\lambda_1+\lambda_2)\cos^3\frac{\theta}{2}-\frac{\lambda_1^3+\lambda_2^3}{\lambda_1\lambda_2}\cos^2\frac{\theta}{2}\sin\frac{\theta}{2}\right.\\
&\quad\left.-\frac{\lambda_1\lambda_2\sqrt{(1-\rho^2)(1+\rho)}}{(\lambda_1+\lambda_2)}\right\}
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{(\lambda_1\lambda_2 - \frac{\rho}{2}(\lambda_1^2+\lambda_2^2))} \\
&\quad \Bigg\{(\lambda_1+\lambda_2)\cos^2\frac{\theta}{2} - \frac{(\lambda_1^2+\lambda_2^2)(\lambda_1+\lambda_2)}{\lambda_1\lambda_2\sqrt{1+\rho}}\cos^2\frac{\theta}{2}\sin\frac{\theta}{2} \\
&\quad -\frac{\lambda_1\lambda_2\sqrt{(1-\rho^2)}}{(\lambda_1+\lambda_2)}\Bigg\} \\
&= \frac{1}{(\lambda_1\lambda_2 - \frac{\rho}{2}(\lambda_1^2+\lambda_2^2))}\left\{\frac{(\lambda_1+\lambda_2)\cos^2\frac{\theta}{2}}{\lambda_1\lambda_2}\left[\lambda_1\lambda_2 - \frac{(\lambda_1^2+\lambda_2^2)}{\sqrt{1+\rho}}\sin\frac{\theta}{2}\right]\right. \\
&\quad \left.-\frac{\lambda_1\lambda_2\sqrt{(1-\rho^2)}}{(\lambda_1+\lambda_2)}\right\} \\
&= \frac{1}{(\lambda_1\lambda_2 - \frac{\rho}{2}(\lambda_1^2+\lambda_2^2))}\left\{\frac{(\lambda_1+\lambda_2)\cos^2\frac{\theta}{2}}{\lambda_1\lambda_2}\left[\lambda_1\lambda_2 - \frac{\rho}{2}(\lambda_1^2+\lambda_2^2)\right.\right. \\
&\quad \left.\left.+(\lambda_1^2+\lambda_2^2)\frac{\cos\frac{\theta}{2}\sin\frac{\theta}{2}\sqrt{1+\rho}-\sin\frac{\theta}{2}}{\sqrt{1+\rho}}\right] - \frac{\lambda_1\lambda_2\sqrt{(1-\rho^2)}}{(\lambda_1+\lambda_2)}\right\} \\
&= (\frac{1}{\lambda_1}+\frac{1}{\lambda_2})\frac{1+\sqrt{1-\rho^2}}{2} \\
&\quad +\frac{1}{(\lambda_1\lambda_2 - \frac{\rho}{2}(\lambda_1^2+\lambda_2^2))}\left\{\frac{(\lambda_1+\lambda_2)(\lambda_1^2+\lambda_2^2)\rho^2\sqrt{1-\rho}}{4\lambda_1\lambda_2\sqrt{1+\rho}}\right. \\
&\quad \left.-\frac{\lambda_1\lambda_2\sqrt{1-\rho^2}}{(\lambda_1+\lambda_2)}\right\}
\end{aligned}
$$

$$
\begin{aligned}
E[\max(X,Y)] &= (\frac{1}{\lambda_1}+\frac{1}{\lambda_2})\frac{1+\sqrt{1-\rho^2}}{2} - \frac{1}{\lambda_1+\lambda_2}\left\{\frac{\sqrt{1-\rho^2}}{1-\rho\frac{\lambda_1^2+\lambda_2^2}{2\lambda_1\lambda_2}}\right\} \\
&\quad + (\frac{1}{\lambda_1}+\frac{1}{\lambda_2})\rho^2\frac{(\lambda_1^2+\lambda_2^2)\sqrt{1-\rho}}{4\sqrt{1+\rho}(\lambda_1\lambda_2-\frac{\rho}{2}(\lambda_1^2+\lambda_2^2))}
\end{aligned}
$$

The formulas are valid for $\lambda_1\lambda_2 - \frac{\rho}{2}(\lambda_1^2+\lambda_2^2) > 0$ or $\rho < \frac{2\lambda_1\lambda_2}{\lambda_1^2+\lambda_2^2} = 1 - \frac{(\lambda_1-\lambda_2)^2}{\lambda_1^2+\lambda_2^2}$.

By Theorem 14.3 $E[\max(X, Y)] + E[\min(X, Y)] = \frac{1}{\lambda_1} + \frac{1}{\lambda_2}$ and therefore:

$$E[\min(X, Y)] = (\frac{1}{\lambda_1} + \frac{1}{\lambda_2})\frac{1 - \sqrt{1-\rho^2}}{2} + \frac{1}{\lambda_1 + \lambda_2}\left\{\frac{\sqrt{1-\rho^2}}{1 - \rho\frac{\lambda_1^2+\lambda_2^2}{2\lambda_1\lambda_2}}\right\}$$
$$- (\frac{1}{\lambda_1} + \frac{1}{\lambda_2})\rho^2\frac{(\lambda_1^2 + \lambda_2^2)\sqrt{1-\rho}}{4\sqrt{1+\rho}(\lambda_1\lambda_2 - \frac{\rho}{2}(\lambda_1^2 + \lambda_2^2))}$$

References

Arnold, B. C., & Strauss, D. (1988). Bivariate distributions with exponential conditionals. *Journal of the American Statistical Association, 83*, 522–527.

Cetin, M., & List, G. (2004). Investigating effects of correlated service times on system performance: Implications in border crossing operations. *Transportation Research Record: Journal of the Transportation Research Board, 1870*, 70–76.

Downton, F. (1970). Bivariate exponential distributions in reliability theory. *Journal of the Royal Statistical Society. Series B (Methodological), 32*, 408–417.

Drezner, T., Drezner, Z., & Zerom, D. (2018). Competitive facility location with random attractiveness. *Operations Research Letters, 46*, 312–317.

Drezner, Z., Turel, O., & Zerom, D. (2010). A modified Kolmogorov-Smirnov test for normality. *Communications in Statistics - Simulation and Computation, 39*, 693–704.

Drezner, Z., & Zerom, D. (2016), A simple and effective discretization of a continuous random variable. *Communications in Statistics - Simulation and Computation, 45*, 3798–3810.

Gelenbe, E., Pujolle, G., & Nelson, J. (1998). *Introduction to queueing networks*. Chichester: Wiley.

Gumbel, E. J. (1960). Bivariate exponential distributions. *Journal of the American Statistical Association, 55*, 698–707.

Kim, B., & Kim, J. (2018). The waiting time distribution for a correlated queue with exponential interarrival and service times. *Operations Research Letters, 46*, 268–271.

Massey, F. J. Jr. (1951). The Kolmogorov-Smirnov test for goodness of fit. *Journal of the American Statistical Association, 46*, 68–78.

Mitchell, C., Paulson, A., & Beswick, C. (1977). The effect of correlated exponential service times on single server tandem queues. *Naval Research Logistics Quarterly, 24*, 95–112.

Nadarajah, S., & Kotz, S. (2008). Exact distribution of the max/min of two Gaussian random variables. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems, 16*, 210–212.

Index

H. A. Eiselt, V. Marianov (eds.), *Contributions to Location Analysis*, International Series in Operations Research & Management Science 281,
https://doi.org/10.1007/978-3-030-19111-5

W