Annals of the Association of American Geographers

Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/raag20

Spatial Optimization in Geography
Daoqin Tong a & Alan T. Murray b

a School of Geography and Development, University of Arizona
b School of Geographical Sciences and Urban Planning, Arizona State University


To cite this article: Daoqin Tong & Alan T. Murray (2012): Spatial Optimization in Geography, Annals of the Association of American Geographers, 102:6, 1290-1309

To link to this article: http://dx.doi.org/10.1080/00045608.2012.685044

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: http://www.tandfonline.com/page/terms-and-conditions

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae, and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand, or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.
Spatial Optimization in Geography

Daoqin Tong* and Alan T. Murray†

∗School of Geography and Development, University of Arizona
†School of Geographical Sciences and Urban Planning, Arizona State University

This article discusses spatial optimization in geography, focusing on contributions of geographers in explicit geographical contexts. An overview of spatial optimization is given, as well as illustrative examples. Many of the individuals contributing to this area of the discipline are identified, demonstrating the breadth of academic institutions spanning the globe where spatial optimization is represented in the research and curriculum of geographers. The article provides a characterization of what a spatial optimization problem is, but also properties, relationships, and challenges behind this. The ultimate purpose of this article is to highlight the spatial optimization subspecialty within geography and in doing so, highlight the need for continued spatial model development and application in the discipline. Further, there is also a need for research focused on techniques to solve spatial optimization problems, particularly in the context of geographic information systems. Key Words: GIS, spatial analysis, spatial optimization, spatial properties, spatial relationships, spatial representation.

Spatial optimization has long been an important subspecialty in the discipline of geography, contributing to the fields of transportation, location modeling, retail geography, medical geography, land use planning, political geography, GIScience, school districting, and others both within and outside the discipline. These fields rely on optimization techniques to structure and solve problems where spatial context is crucial. This has been to prescribe the best spatial arrangement or allocation of entities, resources, or goods for a defined planning period, but also to help understand the significance of a particular spatial arrangement or pattern. Such spatial arrangements might also be the subject of timing choices, phasing in or out, and staging, thus representing both static and dynamic domains.

Examples of where spatial optimization could, should, and have been applied are evident in our everyday lives. School districts across the globe regularly deal with declining or increasing enrollments, necessitating closures, additions, redistricting, and so on. One example is Columbus City Schools (2011), where they are simultaneously siting or renovating schools (five in 2010) and closing existing schools (nine closed in 2010). Issues associated with cost efficiencies, like student transportation, are important, but balancing school enrollments and managing class size, among other things, are central in siting new schools, closing existing

Initial submission, March 2011; revised submission, June 2011; final acceptance, July 2011
Published by Taylor & Francis, LLC.
Spatial Optimization in Geography

Spatial optimization is essential in public transit in urban areas. The City of Phoenix (Valley Metro 2011) has recently added rapid bus transit and light rail lines (with more to come), necessitating changes to support bus route services. Utilization of transit requires good access but also a well-performing system that moves people from where they are to where they need to go quickly and efficiently. In all of these cases, critical decisions about system configuration and operation can be cast as spatial optimization problems that seek to maximize system performance, minimize operational costs, and generally attain the greatest public value possible.

The significance of spatial optimization in geography no doubt stems from its obvious practical use and application. Not surprisingly, fundamental geographic theories, such as land rent relationships in the work of von Thunen, cost minimization in industrial location examined by Weber, and central place service hierarchies detailed in Christaller and Losch (see Kuby 1989; Church 2001; Church and Murray 2009), rely extensively on spatial optimization, either in explaining and understanding geographic patterns or prescribing best locations. To this end, spatial optimization can be considered inextricably grounded in geography, as well as a method relied on in various ways by geography’s quantitative revolution torchbearers—William Garrison, Peter Haggett, Richard Chorley, Alan Wilson, William Bunge, Brian Berry, Peter Gould, Waldo Tobler, Richard Morrill, Duane Marble, Reginald Golledge, and Gerard Rushton, among others. Spatial optimization was also a primary method utilized in the early careers of prominent geographers like David Harvey, Allen Scott, Eric Sheppard, and Kevin Cox, to name but a few.

Although Church (2001) noted that spatial optimization has been a concern of mankind from the beginning, academic reference to spatial optimization within geography can be found in Haggett (1975). The spatial optimization subspecialty continues to thrive today, building on rich disciplinary traditions and practical significance but also hand in hand with development and proliferation of geographic information systems (GIS) and GIScience. In general terms, optimization is a central component of GIS functionality, spatial statistical methods, cartography, clustering, and so on. Discussion here, however, is limited to explicit geographic contexts due to space constraints. Given its disciplinary significance, there are many individuals in geography specializing in spatial optimization. Table 1 lists many of the active faculty and researchers with spatial optimization as a primary or secondary focus, spanning the globe and representing a diversity of institutions. As noted in Table 1, the criterion for inclusion is a PhD from a geography program or an appointment in a geography program. In total, there are 135 people identified in Table 1, but that number does not include individuals at major research laboratories. There is a subset of institutions with a significant presence in spatial optimization, including Arizona State University, the University of California at Santa Barbara, and University College London, many with multiple faculty working in this area. Another notable feature in this list of people is the institution from which their PhD was obtained. A substantial number of these individuals (twenty-two) came out of Ohio State University. Other institutions of note include Johns Hopkins University (nine), the University of Iowa (nine), the University of California at Santa Barbara (seven), SUNY Buffalo (six), McMaster University (six), and the University of Washington (four). There are also a number with three each (Université Catholique de Louvain, University of Hong Kong, and University College London). In total, roughly half of the identified spatial optimization faculty were trained in geography programs at ten different institutions. A substantial number (almost half) came from a variety of universities outside of the ten most prominent producers of spatial optimization faculty in the field of geography.

The purpose of this article is to review and describe spatial optimization as a specialty area of geography, distinguishing it from a broader characterization as a “quantitative method” or “spatial analytical method,” both of which are often used as a contemporary catch-all in the discipline. The next section details a spatial optimization problem. This is followed by a discussion of solution methods for spatial optimization problems. Building on this, a discussion of the challenges facing those contributing to spatial optimization is undertaken, highlighting spatial representation issues, multiple objectives, and trade-off solutions and how closer coupling with GIS will evolve how problems are conceived as well as solved. Finally, conclusions are offered.
Table 1. Spatial optimization faculty

<table>
<thead>
<tr>
<th>University</th>
<th>Faculty</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arizona State University</td>
<td>Alan Murray (UCSB), Luc Anselin (Cornell), Michael Kuby (Boston U), Shubro Guhathakurta (UC Berkeley), Sergio Rey (UCSB), Paul Torrens (UCL)</td>
</tr>
<tr>
<td>California State University Monterey Bay</td>
<td>Yong Lao (Ohio State)</td>
</tr>
<tr>
<td>Cardiff University</td>
<td>Fulong Wu (U of HK), Naru Shiode (UCL)</td>
</tr>
<tr>
<td>Central Michigan University</td>
<td>Xiaolan Wu (Ohio State)</td>
</tr>
<tr>
<td>Chinese Academy of Sciences (CAS-IGSNRR)</td>
<td>Jingfeng Wang (CAS), Jin Fengjun (Renmin U)</td>
</tr>
<tr>
<td>Chinese Academy of Sciences (CAS-IPM)</td>
<td>Zheng Wang* (East Normal U)</td>
</tr>
<tr>
<td>Chinese University of Hong Kong</td>
<td>Bo Huang (CAS), Leung Yee (Colorado)</td>
</tr>
<tr>
<td>Clark University</td>
<td>Samuel Ratick (Hopkins)</td>
</tr>
<tr>
<td>Drexel University</td>
<td>Tony Grubesic (Ohio State)</td>
</tr>
<tr>
<td>Eastern Michigan University</td>
<td>Yichun Xie (SUNY Buffalo)</td>
</tr>
<tr>
<td>Edinboro University (PA)</td>
<td>Wook Lee (Ohio State)</td>
</tr>
<tr>
<td>Florida State University</td>
<td>Mark Horner (Ohio State)</td>
</tr>
<tr>
<td>Free University</td>
<td>Peter Nijkamp* (Erasmus)</td>
</tr>
<tr>
<td>George Mason University</td>
<td>Kevin Curtin (UCSB), Nigel Waters (Western Ontario), Kingsley Haynes* (Hopkins), Roger Stough* (Hopkins)</td>
</tr>
<tr>
<td>Griffith University</td>
<td>Douglas Ward* (Queensland)</td>
</tr>
<tr>
<td>Hofstra University</td>
<td>Jean-Paul Rodrigue (Montreal)</td>
</tr>
<tr>
<td>Indiana University–Purdue University</td>
<td>Rudy Banerjee (Iowa)</td>
</tr>
<tr>
<td>Johns Hopkins University</td>
<td>Benjamin Hobbs (Cornell), Justin Williams (Hopkins)</td>
</tr>
<tr>
<td>KTH Royal Institute of Technology</td>
<td>Takeshi Shirabe (U Penn)</td>
</tr>
<tr>
<td>Kyungpook National University</td>
<td>Kamyong Kim (Ohio State)</td>
</tr>
<tr>
<td>Louisiana State University</td>
<td>Fabui Wang (Ohio State)</td>
</tr>
<tr>
<td>Louisiana Tech University</td>
<td>Irene Casas (Ohio State)</td>
</tr>
<tr>
<td>McGill University</td>
<td>Raja Sengupta (Southern Illinois U)</td>
</tr>
<tr>
<td>McMaster University</td>
<td>Pavlos Karanoglou (McMaster), Darren Scott (McMaster), Antonio Paez (Tohoku)</td>
</tr>
<tr>
<td>Michigan State University</td>
<td>Ligmann-Zielinska (SDSU)</td>
</tr>
<tr>
<td>Missouri State University</td>
<td>Jun Luo (Wisconsin–Milwaukee)</td>
</tr>
<tr>
<td>National University of Ireland</td>
<td>Stewart Fortheringham (McMaster)</td>
</tr>
<tr>
<td>Northern Illinois University</td>
<td>Xuwei Chen (Texas State)</td>
</tr>
<tr>
<td>Ohio State University</td>
<td>Morton O’Kelly (McMaster), Ningchuan Xiao (Iowa), John Current* (Hopkins), David Schilling* (Hopkins)</td>
</tr>
<tr>
<td>Peking University</td>
<td>Yu Liu (Peking)</td>
</tr>
<tr>
<td>Portland State University</td>
<td>Jiunn-Der Duh (Michigan)</td>
</tr>
<tr>
<td>San Diego State University</td>
<td>Fiotr Jankowski (Washington)</td>
</tr>
<tr>
<td>State University of New York at Buffalo</td>
<td>Peter Rogerson (SUNY Buffalo), Enki Yoo (UCSB)</td>
</tr>
<tr>
<td>Sun Yat-sen University</td>
<td>Xia Li (U of HK)</td>
</tr>
<tr>
<td>Sungshin Women’s University</td>
<td>Gunhak Lee (Ohio State)</td>
</tr>
<tr>
<td>Tel Aviv University</td>
<td>Irzhak Benenson (Academy of Sciences of the USSR)</td>
</tr>
<tr>
<td>Texas State University</td>
<td>Ben Zhan (SUNY Buffalo)</td>
</tr>
<tr>
<td>Tsinghua University</td>
<td>Gu Chaolin (Nanjing)</td>
</tr>
<tr>
<td>Universite Catholique de Louvain</td>
<td>Dominique Peeters (Louvain), Isabelle Thomas (Louvain)</td>
</tr>
<tr>
<td>University Amsterdam</td>
<td>Jeroen Aerts* (Amsterdam)</td>
</tr>
<tr>
<td>University College London</td>
<td>Alan Wilson, Paul Dennish (Iowa), Michael Batty* (Wales), Paul Longley (Bristol)</td>
</tr>
<tr>
<td>University of Alberta</td>
<td>John Hodgson (Toronto)</td>
</tr>
<tr>
<td>University of Arizona</td>
<td>Daoqin Tong (Ohio State), Gordon Mulligan (UBC), David Plane (U Penn)</td>
</tr>
<tr>
<td>University of Auckland</td>
<td>David O’Sullivan (UCL)</td>
</tr>
<tr>
<td>University of California at Berkeley</td>
<td>John Radke (UBC)</td>
</tr>
<tr>
<td>University of California at Davis</td>
<td>Patricia Mokhtarian* (Northwestern)</td>
</tr>
<tr>
<td>University of California at Merced</td>
<td>Jeff Wright* (Hopkins)</td>
</tr>
<tr>
<td>University of California at Santa Barbara</td>
<td>Richard Church (Hopkins), Michael Goodchild (McMaster), Keith Clarke (Michigan), Waldo Tobler (Washington)</td>
</tr>
</tbody>
</table>
University Faculty

University of Central Missouri Hosep Cha (Ohio State)
University of Cincinnati Changjoo Kim (Ohio State)
University of Colorado Boulder Seth Spielman (Buffalo)
University of Connecticut Jeff Olslee (SUNY Buffalo), Robert Cromley (Ohio State)
University of Florida Grant Thrall (Ohio State)
University of Georgia Lan Mu (UC Berkeley)
University of Hong Kong Becky Loo (U of Hong Kong), Anthony Yeh (Syracuse)
University of Illinois Wenwu Tang (Iowa), Sara McLafferty (Iowa), Avijit Ghosh* (Iowa)
University of Iowa Gerard Rushton (Iowa), Mark Armstrong (Illinois), David Bennett (Iowa)
University of Leeds Mark Birkin (Leeds), Graham Clarke (Leeds), Martin Clarke
University of London Joana Barros (UCL)
University of Louisville Wei Song (Ohio State)
University of Maryland In-Young Yeo (Ohio State)
University of Michigan Dan Brown (UNC CH)
University of Missouri Tim Matisziw (Ohio State)
University of North Carolina–Charlotte Jean-Claude Thill (Louvain), Eric Delmelle (SUNY Buffalo), Wenwu Tang (Iowa)
University of North Texas Joseph Oppong (Alberta), Paul Hudak (UCSB)
University of Northern Iowa Tim Straus (Washington)
University of Oxford Bill Macmillan (Bristol)
University of Queensland David Pullar (Maine), Jonathan Corcoran (Glamorgan), Robert Stimson (Flinders)
University of South Carolina Chris Upchurch (Utah), Diansheng Guo (Penn State)
University of South Florida Joni Downs (Florida State), Hyun Kim (Ohio State), Michael Niedzielski (Ohio State)
University of Southern California Andrew Curtis (SUNY Buffalo)
University of Tennessee Bruce Ralston (Northwestern), Liem Tran (Hawaii)
University of Texas at Austin David Eaton (Hopkins)
University of Texas at Dallas Denis Dean (Virginia Tech)
University of Toledo Peter Linquist (Wisconsin–Milwaukee)
University of Toronto Ron Buliung (McMaster), John Miron (Toronto)
University of Utah Harvey Miller (Ohio State), Thomas Cova (UCSB)
University of Victoria Peter Keller (Western Ontario)
University of Washington Richard Morrill (Washington), Craig ZumBrunnen (UC Berkeley), Timothy Nyerges

Table 1. Spatial optimization faculty (Continued)

University Faculty

University of Western Ontario Jacek Malczewski
University of Wisconsin–Milwaukee Changshan Wu (Ohio State)
Vienna University Manfred Fischer (Friedrich-Alexander U)
Wageningen University Gerard Heuvelink (Utrecht)
Western Michigan University David Lemberg (UCSB)
Western Washington University Patrick Buckley (Boston U)
Wilfrid Laurier University Steven Roberts (Waterloo)
Wuhan University Zhixiang Fang (Wuhan)

Total 135

Note: Faculty were included based on either having a PhD awarded by a geography program or having a faculty appointment in a geography program. Institution where degree was obtained is indicated in parentheses.
*Indicates primary appointment in a nongeography department.

Characterizing a Spatial Optimization Problem

The use of optimization to study geographic phenomena in a formal manner began to take off in geography with advances in linear programming. Garrison (1959) published one of the first discussions in geography on structuring an explicit optimization problem that represented spatial questions or issues. Other formal texts and papers followed, most notably Scott (1971), Beaumont (1982), and, more recently, Church (2001), focusing on the use of optimization for spatial problems in geography. Similar to mathematics and engineering, spatial optimization problems in general consist of three components: an objective, decisions to be made, and constraining conditions. The objective relates to the
purpose of the problem context, often reflecting goals to be achieved, such as minimizing costs or maximizing achieved benefits. The problem objective is often structured using one or multiple objective functions. Decision variables correspond to the decision(s) to be made. Examples include where a power transmission corridor should be placed, how many sensors are needed to monitor an area, and when new supply distribution centers should be opened. Constraints establish conditions necessary to be satisfied associated with the problem under study. Such constraints might relate to budgetary limitations, production capacities, environmental impacts, and so on. Therefore, objectives, decision variables, and constraints are combined in some way to reflect the geographic problems of interest in either implicit or explicit terms, solved by exact or approximate techniques.

The objective(s) and constraints are generally specified using unambiguous mathematical notation. A generic model formulation is as follows:

Maximize \( g(x) \)  \hspace{1cm} (1)

subject to

\[ f_i(x) \leq b_i \quad \forall i \]  \hspace{1cm} (2)

\[ x \text{ conditions} \]  \hspace{1cm} (3)

where \( x \) is a vector of unknown decision variables, \( x = [x_1, x_2, x_3, \ldots, x_n] \), \( g() \) and \( f_i() \) are functions of \( x \), and \( b_i \) is a coefficient that limits the value of each associated function \( f_i() \). The conditions on \( x \) (Equation 3) typically consist of some combination of real, integer, or binary requirements; nonnegativity stipulations; or both.

What makes an optimization problem (Equations 1–3) spatial is the specification(s) of decision variables, coefficients, functions, and constraining conditions. Because geography and space become part of the model structure by design, this makes the resulting optimization problem unique. It is special in the sense that variables, coefficients, functions, and constraining conditions that are geographically based have interdependent, particularly spatially interdependent, relationships and properties that are often challenging to abstract and model, but they also tend to be inherently difficult to solve as a result. Thus, spatial optimization problems are typically of practical significance, either helping to describe or aiding in planning decision making, as well as requiring substantial technical skill and expertise to appropriately structure and solve for a given substantive context.

The most obvious spatial component of an optimization model is likely geographic decision variables. Problem context often necessitates that some decision be made regarding where something is placed or located. The decision might involve questions of how much should be placed somewhere, what path should be taken, or even what the pattern of a shipment or activity should be. Figure 1 provides a depiction of spatial decision variables along these lines. The decision variables in Figure 1A represent the best location to site a facility. This is precisely what is to be decided in the Weber problem, as an example. The variables therefore correspond to the coordinate pair \((x_1, x_2)\), defining the facility (factory) location, with \( x_1 \) referring to the \( x \)-coordinate and \( x_2 \) referring to the \( y \)-coordinate. Solution of the stipulated problem (beyond the decision variables depicted) would yield the best, or optimal, location as a coordinate pair. Another example is shown in Figure 1B, where there are twelve possible locations to site one or more facilities, so each indicated variable is used to reflect the decision of whether to select the location or not. That is, \( x_i = 1 \) could mean that site \( i \) is to be selected, whereas \( x_i = 0 \) means it is not to be selected. Thus, the seven variables correspond to decisions regarding which locations to select or not select, and the model solution would represent a siting plan of action. A third example is the variables shown in Figure 1C corresponding to decisions about which network arcs to traverse. Assuming a goal to get from \( A \) to \( B \) using arcs in the network, one might well be interested in the shortest path. Given this, the decision variables would indicate the route to be taken. For example, \( x_7 = x_{13} = x_{14} = x_{15} = 1 \) and all other \( x_i = 0 \) would represent one path and therefore a potential solution in this case. Again, all three examples relate decision variables to a geographic context, thereby making them spatial optimization problems. Such geographic variables can be continuous, involving an infinite number of potential locations (Figure 1A), or discrete, associated with a finite number of possibilities (Figures 1B and 1C).

Of course, spatial features of an optimization model extend beyond geographic variables, as coefficients, functions, and constraints could also be spatial in nature. In the remainder of this section, we offer discussion and examples of coefficients, functions, and constraining conditions in spatial optimization focusing on distance, adjacency, connectivity (contiguity),
containment, intersection, shape, districts, and pattern. Some of these terms are considered topological relationships or properties (see Worboys and Duckham 2004; Longley et al. 2011), but here they are characteristics of fundamental geographical interactions in spatial optimization that are relied on in isolation or in combination, likely having significance depending on scale.

Distance

Proximity between places, either in terms of physical distance or travel time, is a very typical concern in geographic problems. Further, Nystuen (1968) identified distance as a fundamental spatial concept. Depending on the context, one might be interested in average distances traveled by people in a region, the maximum distance any person must travel to obtain a service, the shortest path from one location to another, the fastest route between two or more places, and no doubt many others. Distance then is the displacement from a geographic location to another location and can be measured in some way, depending on the intended study purpose. There are, in fact, many ways to measure distance. In two dimensions, Euclidean, rectilinear, $l_p$, or network distances are all possible and have been used in spatial optimization application (see Church and Murray 2009). An example of some of the distance measures is illustrated in Figure 2. Of course, the unprojected world is three-dimensional, so the curvature of the earth must be taken into account. A generic metric like Euclidean distance has a 3-D equivalent—geodesic or great circle, depending on the assumed shape of the Earth. Technical reviews of distance measures can be found in Church and Murray (2009), Miller and Wentz (2003), and de Smith, Longley, and Goodchild (2011). Operationally, it is possible to derive or assess these or related distance measures using commercial GIS software, modules, or libraries.

In spatial optimization it is possible for distance to be incorporated as a function or constraint when facility locations are not known in advance, as is the case for the Weber problem using the Euclidean metric. It
is also possible to utilize distance as model coefficients, however, where they are derived a priori. Irrespective of whether distance is a coefficient, function, or constraint, it is invariably an important facet of model structure and design.

Representative examples in the geographic literature using distance as proximity measures in a spatial optimization context include locating service facilities or centers (Goodchild and Massam 1969; ReVelle and Swain 1970; Hillsman 1984; Densham and Rushton 1988; ReVelle and Elzinga 1989; Church 1990; Oppong and Hodgson 1994; Yeh and Chow 1996; Horner and Downs 2010) and sensitivity studies of problem solutions using various distance metrics (Peeters and Thomas 2000), among others.

Adjacency

The relationship that two areas are next to each other, or neighbors, is known as adjacency. Formally, when two polygons share a common edge or boundary, they are said to be adjacent (Church and Murray 2009). As illustrated in Figure 3, unit 1 is adjacent to units 2, 4, and 5. Adjacency can also be defined in terms of distance or travel time between two locations. Assessment of adjacency typically relies on GIS-based functionality to evaluate geometric intersection of polygon boundaries or, in some cases, distance or travel time.

In spatial optimization, adjacency typically reflects a limiting condition where it is necessary or desirable to avoid saturation of an area. Thus, given possible sites that would conflict if simultaneously chosen, as an example, constraints based on adjacency are imposed to ensure this does not happen. If one activity is sited, no other activities can be sited in adjacent areas that otherwise would constitute a conflict or oversaturation. Alternatively, one might be interested in minimizing/maximizing a spatial property that is defined by adjacency relationships.

Representative examples by geographers in the literature using adjacency as part of model structure in spatial optimization include the work of franchise systems evaluation (Zeller, Achabal, and Brown 1980), urban growth modeling (Clarke, Hoppen, and Gaydos 1997; Ward, Murray, and Phinn 2000), forest management (Snyder and ReVelle 1996; Murray 1999), and, more recently, nature reserve design (Fischer and Church 2003; Matisziw and Murray 2006), land use planning (Yeo, Guldmann, and Gordon 2007), urban form simulation (Xie, Batty, and Zhao 2007), sex offender residency modeling (Grubesic and Murray 2008), and wildlife management (Downs, Gates, and Murray 2008), among others.

Connectivity

The ability to get from one location to another or moving unimpeded from one location to another is known as connectivity, and it is also a fundamental spatial concept discussed in Nystuen (1968). Connectivity typically arises in a network, made up of nodes and incident arcs, where one is interested in whether a path exists along arcs in the network between a given set of nodes (or locations; see Haggett 1965). A related concept is contiguity, an early discussion of which can be found in Dacey (1965). A set of spatial objects is contiguous if it is possible to travel between all objects without leaving the object set (Church and Murray 2009). Connectivity and contiguity are related because the assessment of contiguity often involves the transformation of space to a corresponding network, where nodes represent the spatial objects and arcs between nodes are defined if the objects are adjacent. Figure 4 illustrates such a connected network where travel from any node is possible along links in the network. If link 1 is disabled or otherwise removed, nodes in region A will be disconnected from those in region B. Strictly interpreted, connectivity or contiguity is an all-or-nothing spatial property. A set of objects (or nodes) is either connected (contiguous) or not. Recent work by Wu and Murray (2008) has focused on relaxing the Boolean interpretation, introducing relative contiguity (relative
connectivity). Assessment or use of connectivity and contiguity measures can be carried out using GIS functions, generally requiring customized programming of some sort.

In spatial optimization a requirement might be to maintain connectivity. Thus, connectivity would be structured using constraints. Examples include conservation of flow constraints in path-based models used to identify a right-of-way corridor, but similarly in land acquisition too to create a reserve habitat for threatened or endangered species, a park, a community development, a landfill, a factory, and so on. In terms of relative contiguity, this could be approached as a function in the objective with a goal to minimize or maximize resulting connectivity.

Representative examples using connectivity or contiguity for spatial optimization by geographers include transportation network planning (MacKinnon and Hodgson 1970; O’Kelly 1987; H. Kim and O’Kelly 2009), land use planning and nature reserve design (Brookes 1997; Cova and Church 2000a; Williams 2002; Shirabe 2005b; Wu and Murray 2007; Ligmann-Zielinska, Church, and Jankowski 2008), and, more recently, security camera monitoring (Murray et al. 2007) and evacuation risk (Church and Cova 2000), among others.

**Containment**

Given two geographic objects, the condition where one object is completely within the other object is the containment relationship (see Longley et al. 2011). If one object is the trade area of a retail outlet, it is conceivable that a query of interest would be to find all households that are within the trade area. Assume that each household is represented as a point. Identifying those households within the trade area of the outlet involves an examination of containment: Is the point contained within the outlet’s trade area (polygon)? Figure 5 also illustrates the spatial containment of a warning siren where the evaluation involves the identification of the associated service area. Evaluation of containment therefore requires assessment of geometric relationships, a task easily accomplished using commercial GIS functionality.

In spatial optimization there is a need to account for various types of containment, thereby becoming part of the inherent process and model structure. Containment might focus on what areas are provided service coverage by current or potential service facilities, as an example. Thus, the containment relationship enables one to know who could be or would be served. This might then be a function that is optimized—locate facilities so that most of the population is contained within intended service areas. Alternatively, containment might be utilized to identify conflicts in a region. For example, if three waste disposal sites are located, containment relationships might be structured to ensure that no neighborhood is unduly impacted by more than one site. In this case, containment represents a constraint that limits spatial impacts.

Representative examples by geographers include emergency facilities siting (Toregas et al. 1971; Church and ReVelle 1974), health service provision (Osleeb and McCafferty 1992; Oppong and Hodgson 1994), ground water monitoring network design (Hudak, Loaiciga, and Schoolmaster 1993), reserve selection (Church, Stoms, and Davis 1996), and, more recently, warning siren siting (Murray 2005; Tong and Murray 2007).
Intersection

The property of intersection refers to a spatial location(s) where two objects simultaneously exist. For example, as Figure 6 shows, the intersection of two areas is that portion common to both areas (see Worboys and Duckham 2004; Longley et al. 2011). This term comes from set theory, along with union, difference, complement, and so on, and represents a formalized approach for relating spatial objects. Evaluation of intersection between spatial objects is a standard feature of most commercial GIS packages.

In spatial optimization, intersection has been used to account for service overlap and redundancy. If two or more sited facilities serve the same neighborhood, as an example, this might be viewed in a positive way, depending on the context. If the facilities are fire stations, then multiple fire stations capable of responding in a timely manner to a neighborhood means that the probability of delayed response or unavailability for that area in the time of an emergency decreases. To this end, it might be beneficial to optimize intersection, or overlap, associated with service areas when locating facilities. Thus, an objective in this case would involve the functional relationship of intersecting service areas. It is conceivable as well that one could constrain the level of intersection of service areas, too. This might be to ensure no more than a stipulated level of intersection but also could prohibit intersection altogether.

Representative examples in the geographic literature include the work of backup service provision (Hogan and ReVelle 1986) and, more recently, refueling facility siting (Kuby and Lim 2005), nature reserve design (Malcolm and ReVelle 2005), security monitoring (Murray et al. 2007; K. Kim, Murray, and Xiao 2008), traffic flow assessment (Zeng, Castillo, and Hodgson 2008), and emergency medical service provision (Erdemira et al. 2010), among others.

Shape

There has long been interest in shape in geography, particularly as it relates to a spatial feature or object (or collection thereof). An early detailed discussion of geographic shape can be found in Bunge (1962), Boyce and Clark (1964), and, more recently, in Wentz (2000). Although viewed as a nebulous term or property, it is reasonable to define shape as a characterization of a spatial object. Often, such a characterization has been to relate properties of an object’s observed shape to those of known shapes, like a circle, square, or rectangle.

In spatial optimization, shape has been a property considered by many applications. Political geography most certainly has focused on shape, but it is also a factor in nature reserve design and land acquisition applications. The reason is that shape has significant implications, both in terms of the meaning of an area created as well as its inherent characteristics and value. In some contexts an elongated area might be viewed in positive terms but as a negative in others. As an example, gerrymandered legislative districts characterized by their elongated shape have been (and are) considered problematic, often created to achieve a particular political gain through a strategic representation of individuals. In contrast, some species are known to prefer “corridor” spaces that are elongated in shape, such as a corridor (Figure 7). Spatial optimization concerned with shape has often focused on compactness, or a circle-like property of shape. Shape has been conceived of as a function to be optimized using the perimeter of the resulting shape. Enclosing the same area, a shape with the least perimeter is a circle and obviously the most compact shape possible. It is also possible to utilize one or more derived measures of shape as constraining conditions, where they are limited to certain possible values.

Representative examples in the geographic literature using shape in spatial optimization include land use...
planning (Wright, ReVelle, and Cohon 1983; Aerts et al. 2003; Shirabe 2005a), nature reserve design (Williams, ReVelle, and Levin 2004), and evaluation of service areas (Massam and Goodchild 1971), among others.

Districts

There has long been interest in districts (or partitions or regions) in geography. Discussion of geographic districts can be found in Garrison (1959), Berry (1961), Yeates (1963), and Morrill (1973). Although varied in terms of concern, the importance of districts has generally been from the perspective of regionalization of space, administrative organization, or both. Examples include post office service areas, school districts, and political jurisdictions. As a spatial property or relationship, the districts of an area can be thought of as a partitioning of space into a finite number of subregions. Figure 8 shows a case where the region is partitioned into three subregions. These subregions might be required to be contiguous or they might not. Further, there might be desired attributes for the districts, such as equitable, balanced, compact, and so on (see Lemberg 2004).

In spatial optimization, districts are significant in many application settings. Often, districts are couched as a spatial outcome in the sense that one is looking for a partitioning of geographic space into a finite number of districts. To this end, districts might be sought that have other desirable properties, with respect to either attributes or spatial features. For example, uniform districts with an equivalent population in each could ensure balanced or equitable service. Alternatively, districts with particular spatial properties might be preferred, such as those that are contiguous or most compact.

Representative examples in the geographic literature focused on districting in spatial optimization include school districting (Schoepe and Church 1991; Church and Murray 1993; Lemberg and Church 2000), political redistricting (Morrill 1976; Macmillan and Pierce 1994; Openshaw and Rao 1995; Williams 1995; Barkan, Densham, and Rushton 2006), and, more recently, unit grouping (Duque, Church, and Middleton 2011), among others.

Pattern

A description of a distribution of spatial objects is often referred to as a pattern. Thus, pattern reflects a property of how objects are organized across space and in many cases is conceived of in terms of being random, clustered, or dispersed. An early discussion of pattern can be found in King (1962), focusing on human settlement, with direct connections to location theory and central place hierarchies. Of course, a major concern in early location theory was settlement patterns and the socioeconomic processes behind them. The quantification of pattern in some circles has become fairly well accepted, typically relying on measures of point pattern (quadrat, nearest neighbor, K functions, etc.), spatial autocorrelation (join counts, Moran’s I, Getis-Ord G, etc.), fragmentation, and geostatistical models (see O’Sullivan and Unwin [2003] for a discussion of selected measures). Figure 9 plots examples of clustered,
In spatial optimization, pattern is particularly important in a number of contexts. As an example, when identifying land for nature reserve designation to protect threatened flora and fauna, a common concern or goal is to ensure that the land parcels identified are clustered. Alternatively, if disaster relief facilities were sited in a city, it would no doubt be desirable to disperse the facilities to increase the likelihood that most would be operational after a natural disaster or terrorist attack. In either case, or others, measurement along the preceding lines is possible, but the ideals of clustered and dispersed, and anywhere in between, can be operationalized as well.

Representative examples in the geographic literature where pattern is addressed using spatial optimization include facility dispersion (Kuby 1987; Curtin and Church 2006), settlement patterns (Church and Bell 1988), obnoxious facility location (Ratick and White 1988), and, more recently, nature reserve design (Matisziw and Murray 2006) and backup facility siting (Ratick, Meacham, and Aoyama 2008).

**Solving Spatial Optimization Problems**

The previous section detailed the features that contribute to making a spatial optimization problem unique. The specification and formulation of an optimization problem to appropriately abstract reality and represent a problem or issue of interest is typically not a trivial task, especially in a spatial context. Some even consider structuring a model to be an art, requiring skill and creativity (Church and Murray 2009). The decisions to be made and complicated relationships and properties of a spatial problem need to be constructed using mathematics or mathematical principles in a logical manner. How to abstract reality is neither easy nor straightforward in most cases. Sufficient detail is needed for a realistic reflection of the problem of interest, and this must be brought together in a way that makes sense.

Armstrong (2000) highlighted that many geographic models are “intrinsically computationally intensive.” This is particularly true with spatial optimization. Comparatively, spatial optimization problems are challenging and difficult to solve for a number of reasons. The structure of a spatial optimization problem often involves dealing with geographic features and properties, as discussed in the previous section, including the specification of connectivity, evaluation of containment and intersection, and derivation of shape or pattern. Incorporating these spatial elements often involves
more complex formulations, and additional constraints and variables, which inevitably requires more effort in solving the corresponding problem. In addition, with advances in GIS and other technologies (remote sensing, Global Positioning System, etc.), more spatial data are available with higher resolution, necessitating even more detail and resulting in even larger problem sizes.

Solving a spatial optimization problem is therefore complicated by a number of issues. First, problem articulation and specification might be difficult. Second, assuming that the problem can be communicated, the inherent spatial structure might make the spatial optimization problem difficult to solve. Third, again assuming that the problem can be formalized, the size of the associated spatial optimization problem (number of decisions and number of constraints) could make it difficult to solve. Finally, the decision-making context might contribute to making it difficult to solve. Some situations might require decisions to be arrived at in real time, on the spot, so an approach would be required to identify decisions in seconds or minutes. In other situations, more time might be available for problem solution, perhaps hours, days, weeks, or months.

A solution to an optimization problem ultimately indicates the values identified for the associated decision variables. The constraints define solution feasibility. That is, a solution satisfying all problem constraints would be considered feasible. All or some of the feasible solutions might be of interest, depending on the problem. Whether it is a good or bad solution, however, depends on the measured quality of the solution, typically reflected in the objective function. Given that the objective function provides the measure of solution quality, it is a matter of somehow scrutinizing this function to infer decision variable values that are of most interest. Most people have likely encountered some optimization problem solution principles in a formal manner through basic calculus when determining the derivative of a function. The derivative set equal to zero theoretically gives the so-called critical points of a function. Typically of interest in this set of critical points are the relative extrema (minima and maxima). In a strict sense, however, absolute minima and maxima are the sought-after solutions, depending on whether one seeks to minimize or maximize the objective. Unfortunately, it is not always possible (or easy) to find the derivative, not to mention maintaining constraint feasibility. As a result, other approaches are necessary.

There are two basic strategies for solving spatial optimization problems, exact and heuristic methods (Scott 1971; Miller and Shaw 2001; Church and Murray 2009). Exact methods are those that exhaust all possibilities or exploit problem properties, ensuring that the optimal solution is found. Alternatively, heuristics are problem-specific ad hoc strategies, finding characteristically good solutions but generally lacking any capacity to verify or validate solution quality.

**Exact Methods**

A solution approach to an optimization problem is considered an exact method if the best possible, or optimal, solution is guaranteed to be identified using the approach. That is, it can be proven that the solution found by an exact method produces the best value, so no other decision variable values would result in a better objective while maintaining problem feasibility. There are, in fact, many possible exact approaches, including derivative-based techniques, enumeration, linear programming, integer programming with branch and bound, and so on, as well as more problem-specific methods like the Hungarian algorithm, transportation simplex, out-of-kilter algorithm, Dijkstra’s algorithm, and so on. A few of these approaches are now discussed in more detail.

Enumeration is somewhat straightforward, depending on the problem, as it identifies all feasible solutions, evaluates them, and then allows concluding which is the best. This makes a lot of sense, if solutions can be readily identified and the number of different solution combinations is not computationally prohibitive. Further, one can guarantee optimality or exactness. Unfortunately, many problems cannot be approached using enumeration for two reasons. First, some problems have an infinite number of feasible solutions, making evaluation of each impossible. Second, the number of feasible solutions, although finite in number, might simply be too plentiful to evaluate in a reasonable amount of time.

Another popular exact method is linear programming, an approach that is based on linear algebra and solving a system of linear equations. It assumes that the objective and all constraints are linear functions and that decision variables take on continuous values. There are many commercial software packages capable of solving linear programming problems with millions of decision variables and millions of constraints (e.g., Gurobi, Cplex, LINDO, etc.). The only real limitation is that some problems do not maintain the assumption of linearity, either in functions or decision variable values.

To address one aspect of limiting capabilities in linear programming, integer programming with branch and
bound is an exact method for solving problems where decision variables are restricted to integer values. Although also commercially available, the problem sizes that can be addressed are substantially more limited for integer programs.

What makes spatial optimization problems difficult (or special) to solve by exact methods is the mathematical structure attributable to geographic relationships being modeled. All of the topological relationships noted previously are examples, including distance, adjacency, connectivity, containment, intersection, shape, districts, and pattern. Rosing, ReVelle, and Rosing-Vogelaar (1979) is an early example recognizing that different mathematical structures are possible to reflect user assignment in the $p$-median problem (constraints in the model), with some approaches being virtually impossible to solve, whereas others are extremely easy using linear programming (see Church 2003, 2008). For a dispersion problem, Murray and Church (1997) were able to derive constraints that exploited spatial relationships, resulting in problem structures (constraints) that enabled applications to be solved that otherwise would not have been solvable. Another example is work by Williams (2002) and Shirabe (2005b) who devised exact approaches for mathematically structuring contiguity conditions (constraints) in land use planning models.

Heuristics
There are many reasons why exact methods for solving an optimization problem might not be feasible or appropriate. The most obvious is that no exact method can solve the problem at hand. Of course, there are other reasons as well, such as computational effort exceeding allotted time, limitations with exact approaches in finding alternative or near-optimal solutions, expense (both computing time and solver software), and others.

As noted earlier, a heuristic is an approach based on rules of thumb, strategies, or ad hoc procedures to solve an optimization problem. The idea is to explore the solution space in some way, identifying feasible solutions from which a best solution can be found. Some heuristics start with one or multiple feasible solutions. According to the specific search rule of the heuristic, a new solution(s) is identified and the current solution(s) is updated. The search ends when the termination condition is satisfied. Through the search over many iterations, a high-quality solution(s), possibly an optimal one, is sought. Many heuristics have been used for solving spatial optimization problems, including interchange, greedy, simulated annealing, tabu search, and population-based heuristics, such as genetic algorithms and ant colony heuristics.

What makes spatial optimization problems difficult (or special) to solve by heuristics are similar to exact methods—problem structure associated with topological relationships. When heuristics are relied on, however, spatial features can also be utilized for solving the problems more efficiently and effectively. In zoning design, Openshaw and Rao (1995) implemented simulated annealing and tabu search heuristics where the modification of a zone only occurs on the boundary units to ensure contiguity, thereby using geographic knowledge to enhance the techniques for solving the problem. Densham and Rushton (1992) showed that computational efficiencies could be achieved in heuristic performance if spatial structure was exploited in solving the $p$ median problem (see related work by Sorensen and Church 1996). A continuous space coverage problem was addressed in Murray et al. (2008) that relied on spatial functions and properties and Voronoi polygons combined with derived medial axes in a developed heuristic. Spatial structure of the maximal covering location problem was exploited in Tong, Murray, and Xiao (2009) in a genetic algorithm heuristic, using a crossover operator based on spatial proximity of candidate facility sites. In another genetic algorithm application, Fraley, Jankowska, and Jankowski (2010) were able to spatially diversify solutions through promotion, resulting in improvement of the overall heuristic performance in delineating neighborhood boundaries.

Future Spatial Optimization Challenges
The inherent complexity of geographical problems poses great challenges to solving spatial optimization problems. Research developing exact and heuristic solution approaches for spatial optimization problems will likely remain an important and active area. There is an increasing exigency in geography to develop techniques for solving spatial optimization problems efficiently and effectively. This is driven by growing needs to better reflect real-world problems but also the recognition that even small economic efficiencies can result in millions of dollars in savings for businesses, municipalities, and governments. Thus, problems are inherently more difficult to solve because of larger planning applications and greater spatial detail. Heuristics are appealing in this regard, but identifying the most feasible solution will continue to warrant serious consideration in times
when better efficiency can translate to significant monetary savings.

It is clear, then, that problem solution will continue to challenge spatial optimization application, as new models are defined, existing models are enhanced, and more challenging problem instances are encountered. This goes beyond the issue of whether to use an exact or heuristic approach, suggesting that solution development and the continued search for better methods will always be important. There are other challenges facing spatial optimization researchers though, such as abstraction and representation issues, recognition of multiple problem objectives and the need for considering alternative optima, and further exploitation of topological relationships accessed through GIS. These issues are now discussed in more detail.

Abstraction and Representation Issues

Although existing spatial optimization models are typically capable of reflecting important aspects of geographic relationships, uncertainties and errors in the abstraction process remain. Such uncertainties and errors are often associated with the way we abstract geographical space. With advances in GIS, the representation of geographic space is increasingly seen as an important issue in spatial analysis (Miller 1996; Church 1999; Goodchild and Haining 2004; Church and Murray 2009). It is well recognized that findings can be highly dependent on how space is abstracted and represented. This can be due to the way we partition or conceptualize space.

Similar to other methods in geography, spatial optimization problems often are subject to the modifiable area unit problem (MAUP) regarding scale of analysis and unit definition (see Openshaw and Taylor 1981). For example, in an application of the p median problem, Fotheringham, Densham, and Curtis (1995) highlighted that solutions were highly dependent on the way geographical units are defined, suggesting caution when using aggregated spatial units. Although Murray and Gottsegen (1997) showed this to be somewhat exaggerated, the issue nevertheless remains. Horner and Murray (2002) illustrated scale and unit definition complications in the assessment of urban excess commuting using the transportation problem. Murray and O’Kelly (2002) provided examples of abstraction and MAUP issues in coverage optimization, and Cromley and Mrozsinski (2002) provided supporting evidence of this as well.

In spatial optimization, geographical units are often abstracted as points, such as centroids for census tracts, blocks, cities, traffic analysis zones, and so on. Compared to complex spatial objects, a point representation makes data handling and problem solution easier given that evaluation of spatial relationships based on points is more straightforward (Miller 1996; Church 1999; Murray 2003), but loss of geographical detail introduces uncertainties and errors in point simplification. For example, Hillsman and Rhoda (1978) detailed sources of error associated with proximity when areal units are represented as points in location models (see also Goodchild 1979; Rodriguez-Bachiller 1983; Rushton 1989; Miller 1996; Church 1999).

Recognizing that errors are inherent in representation strategies used to abstract geographical space, alternative representation schemes with more spatial detail are essential. The use of the most disaggregated data available has been suggested in spatial optimization modeling (Rushton 1989; Horner and Murray 2002). Recent advances in computational capabilities have made representation and evaluation of various spatial object-based relationships feasible (Miller and Wentz 2003). Instead of aggregating two-dimensional geographical areal units into points, studies have also examined the use of objects of various geometric shapes, including lines and polygons, to represent geographical entities in spatial optimization (Murray and Tong 2007). Some advances in this area are even focusing on new model structure to capture important spatial object features and relationships, such as the work of Murray (2005), K. Kim and Murray (2008), and Tong and Murray (2009). Using objects of various shapes to represent space allows for more geographic detail to be captured in spatial optimization models. The complex features and relationships of spatial objects in models, however, make solving them more challenging. For example, K. Kim et al. (2008) and Tong, Murray, and Xiao (2009) reported difficulty in solving modest problem instances by exact methods, necessitating the development and use of heuristics.

The difficulty in modeling geographical units and solving the associated problem instances presents challenges to spatial optimization. Much research is needed to reduce or alleviate errors and uncertainties in abstracting geographical space. One potential research direction for addressing this issue is to develop new abstraction schemes that represent geographical units more realistically. Recent studies by Murray (2005) and Tong and Murray (2009) show promise along these lines. New spatial abstraction strategies, however, will
no doubt bring about a need for the development of new models and more efficient approaches to solve them.

**Multiple Objectives and Trade-Offs**

Spatial problems are complex, as they reflect real-world issues to be addressed. Concerns from different perspectives could arise for most problems of interest, including social, environmental, economical, political, legal, and other types. Given this, there is increasing recognition of multiple objectives that should be addressed. For example, in land use planning, criteria for land acquisition can include proximity to transportation facilities, costs, and environmental impact, among others. As a result, problem solutions often reflect the trade-offs between different objectives. Schilling et al. (1980) recognized the importance of alternative objectives in spatial optimization, with Rushton (1989) also raising the issue of considering alternative or near optima.

Different objectives often conflict with each other and there rarely exists a solution that optimizes all the objectives simultaneously. With one solution, one objective might be optimized but usually the others are not. Hence, a multiobjective optimization problem often has a set of optimal solutions where no single solution is superior to the other, with each representing some trade-off or compromise between objectives (Co-hon 1978). The collection of all of the optimal solutions to a multiobjective problem forms the Pareto front. Any solution that is not on the Pareto front is considered an inferior solution, dominated by at least one optimal solution on the Pareto front. Solving a multiobjective optimization problem brings about the need to identify the Pareto front. The weighting method is one technique that has been commonly used for solving a multiobjective optimization problem. The method assigns weights (preferences) to objectives. The problem then can be conceived of as a single objective optimization problem. By varying the set of weights assigned to objectives, solutions can often be obtained to form or approximate the Pareto front. This method has been used in a variety of studies, including security monitoring (Murray et al. 2007), design of a relief distribution network (Horner and Downs 2007), and urban land use allocation and development (Ligmann-Zielinska, Church, and Jankowski 2008), among others. Although straightforward and widely applied, the weighting method is not without limitations (see Xiao 2008; K. Kim et al. 2008).

Heuristics are another commonly used strategy for solving a multiobjective optimization problem. Among other heuristics, evolutionary and genetic algorithms are widely used for solving multiobjective optimization problems. In contrast to the weighting method with a single combined objective incorporating different weight assignments solved multiple times, evolutionary algorithms are able to provide multiple solutions in one application of the heuristic. Evolutionary algorithms for solving multiobjective spatial optimization problems are detailed in Bennett, Wade, and Armstrong (1999); Xiao, Bennett, and Armstrong (2002); Li and Yeh (2005); K. Kim et al. (2008); and Wu, Murray, and Xiao (2011), among others.

Many challenges exist in multiobjective spatial optimization. Similar to single-objective spatial optimization problems, problem formulation and solution are not easy tasks. For example, research on how to simultaneously model aspects of compactness, contiguity, and shape in land acquisition remains challenging. Solving multiobjective spatial optimization problems is significantly more difficult compared to its single-objective counterpart. For a multiobjective problem, the final selection of an optimal solution for implementation can also be challenging. Solutions on the Pareto front reflect different trade-offs of the objectives, and identifying the one that reconciles all parties’ preferences is not easily accomplished.

**Integration with GIS**

As spatial optimization explicitly addresses problems with an inherent spatial context, the importance of GIS has been increasingly recognized. This includes management of spatial data, evaluation of spatial properties and relationships, assistance in problem solution, and solution display and assessment (Longley and Batty 1996; Church 1999, 2002; Malczewski 1999; Church and Murray 2009; Murray 2010). A significant amount of spatial data are available in GIS and readily serve as input into a spatial optimization model (Church 2002; Murray 2010). Such data include demographic information, land uses, elevation, administrative boundaries, transportation networks, flows, and so on. In addition, GIS provides functionalities for analyzing spatial features and evaluating various spatial relations that are critical in spatial optimization, including distance calculation, adjacency and contiguity assessment, evaluation of a shape and pattern, and so on.

In addition to the aforementioned support that GIS offers to any spatial analysis, advances in GIS can change how we look at and conceive of problems. For
example, GIS data models often dictate how space is abstracted and represented but also present possibilities for new conceptualizations. Murray (2005) proposed an object-based spatial representation in coverage location modeling and demonstrated that the new representation scheme was more reliable and less sensitive to MAUP compared with commonly used abstraction strategies based on points. Similar advances are found in K. Kim and Murray (2008) and Tong and Murray (2009). Perhaps equally important is that GIS facilitates representation and solution of some problems that for various reasons have characteristics that cannot be well articulated by mathematical formulas. Examples include the work of Brookes (1997) but also growth and development problems addressed by Clarke, Hoppen, and Gaydos (1997), and Benenson and Torrens (2004), among others.

The GIS environment can also enhance our capability to solve various spatial optimization problems, either exactly or heuristically. On the exact method side, work by Murray and Tong (2007) exploits spatial knowledge and relationships to simplify the problem and ultimately solve it. Murray et al. (2008) and Matisziw and Murray (2009) illustrated the potential to extract features of space that enable a more focused and mathematically structured search for optimal solutions. All of this is only possible through use and integration with GIS.

In addition, multiobjective spatial optimization problems provide a basis for greater integration and understanding through GIS. Recently, there has been continued interest in integrating GIS and multiobjective spatial optimization as spatial decision support systems (Malczewski 2006). On one hand, GIS provides tools and functionalities for spatial data management and for decision makers to map solutions and interactively modify preferences and trade-offs between various objectives (Armstrong et al. 1992; Jankowski, Andrienko, and Andrienko 2001). Furthermore, GIS can also be used to facilitate constructing and solving a multiobjective spatial optimization problem. Representative examples are Eastman et al. (1995), Cromley and Hanink (1999), Cova and Church (2000b), and Chakhar and Mousseau (2007). On the other hand, advances in multiobjective spatial optimization help to enhance GIS capabilities. Representative work includes Church et al. (2003) and Huang et al. (2008).

Conclusions

Spatial optimization is an important subspecialty in the discipline of geography, with a vast array of fields relying on optimization techniques to structure and solve problems where spatial context is crucial. Geography faculty working in the area of spatial optimization truly span the globe, providing education to the next generation of researchers, planners, and decision makers, but also applying spatial optimization to important problems of our day. It is also worth noting that other disciplines have contributed to the growth of spatial optimization, including regional science, industrial engineering, operations research, ecology, and so on. As the discipline of geography continues to be more interdisciplinary, it is true of spatial optimization as well.

This article provides an overview of spatial optimization as a subspecialty in the discipline of geography. We highlight the characteristics of spatial optimization that distinguish it from spatial analysis or spatial analytical methods, as well as optimization. Further, much emphasis was on defining what spatial optimization is and what makes encountered problems spatial. Methods of problem solution were reviewed, including exact and heuristic techniques. This was followed by future challenges ranging from representation and abstraction issues to multiple objectives to GIS integration. All are further complicated by a future where location-based services and the so-called data avalanche are a reality. Of course, there is an increasing demand for providing real-time solutions where problems are constructed and solved on the fly. In other contexts, temporal components need to be incorporated explicitly (Miller and Bridwell 2009; Osleeb and Ratick 2010; Tong, Ren, and Mack 2012), often resulting in a tremendous increase in problem complexity. All of this calls for the development of more efficient solution algorithms that are able to solve large-sized problems quickly.

Notes

1. Because of the limited scope, no discussion of optimization used in facets of GIScience, statistics, spatial statistics, cartography, geostatistics, and so on is included. See Murray (2007) for a more general discussion that includes these areas.
2. This might not always be the case, as it is possible to structure a heuristic solution approach without technically formulating the actual problem of interest.
3. The \( p \) median problem aims to find \( p \) facility locations that minimize overall demand weighted travel distance.
4. The maximal covering location problem seeks to identify the spatial configuration of a fixed number of service facilities so that the overall coverage provided is maximized.
References


Correspondence: School of Geography and Development, University of Arizona, Tucson, AZ 85721, e-mail: daoqin@email.arizona.edu (Tong); School of Geographical Sciences, Arizona State University, Tempe, AZ 85287, e-mail: atmurray@asu.edu (Murray).