Transportation Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Continuous Facility Location with Backbone Network Costs

John Gunnar Carlsson, Fan Jia

TRANSPORTATION SCIENCE
Articles in Advance, pp. 1–19
ISSN 0041-1655 (print) | ISSN 1526-5447 (online)

http://dx.doi.org/10.1287/trsc.2013.0511

Continuous Facility Location with Backbone Network Costs

John Gunnar Carlsson, Fan Jia
Department of Industrial and Systems Engineering, University of Minnesota, Minneapolis, Minnesota 55455
{jgc@umn.edu, jiaxx049@ie.umn.edu}

We consider a continuous facility location problem in which our objective is to minimize the weighted sum of three costs: (1) fixed costs from installing the facilities, (2) backbone network costs incurred from connecting the facilities to each other, and (3) transportation costs incurred from providing services from the facilities to the service region. We first analyze the limiting behavior of this model and derive the two asymptotically optimal configurations of facilities: one of these configurations is the well studied *honeycomb heuristic*, and the other is an Archimedean spiral. We then give a fast constant-factor approximation algorithm for finding the placement of a set of facilities in any convex polygon that minimizes the sum of the three aforementioned costs.

Keywords: continuous facility location; backbone network design
History: Received: February 2012; revisions received: August 2012, February 2013; accepted: June 2013. Published online in *Articles in Advance*.

1. Introduction

Broadly speaking, a set of facilities providing service to a geographic region often incurs costs from three major sources:

1. *fixed costs* from installing the facilities,
2. *backbone network costs* from connecting the facilities to each other, and
3. *transportation costs* from providing services from the facilities to the region.

Letting $X = \{x_1, \ldots, x_k\}$ denote the set of facilities and R the service region, the optimal location problem is therefore given by

$$\underset{X}{\text{minimize}} \; \{\text{Fixed}(X,R) + \text{Backbone}(X,R) + \text{Transportation}(X,R)\}. \qquad \text{(FBT)}$$

The novelty of problem (FBT) comes from the property that the fixed and backbone network costs will often increase as more facilities X are added (since increasing the number of facilities in a region usually increases the fixed costs in the region and makes the backbone network longer, although exceptions to this principle certainly exist), but the transportation costs should decrease because there are more facilities located in the region to provide service. In this paper, we consider the case where the service region is a convex polygon C and the facilities X represent facilities whose customers or clients are uniformly distributed in the region. An application of (FBT) was introduced in Cachon (2014), where the goal is to minimize the total amount of carbon emissions that are produced by a supply chain network of retail stores and the customers they serve. In this paper, we approximate the three quantities above by making the following assumptions:

1. The costs due to *facilities* (the "fixed costs") take the form $\gamma_k \cdot k$, where $k = |X|$ and $\{\gamma_k\}$ is a sequence whose kth entry γ_k represents the fixed cost per facility when we build a total of k facilities. It is natural to assume that γ_k is decreasing to reflect the intuitive notion that a collection of small facilities is cheaper per facility (or, as might be inferred from Cachon 2014, produces fewer emissions per facility) in the language of Cachon (2014) than a single large, central facility (e.g., a single facility with a capacity to serve 1,000 customers produces more emissions than a facility with a capacity to serve only 500 customers). It is also natural to suppose that $\gamma_k \cdot k$ should *increase* as k becomes larger as a result of the usual economies of scale (e.g., a single facility with a capacity to serve 1,000 customers produces fewer emissions than *two* facilities each having a capacity to serve 500 customers).
2. The costs due to the *backbone network* are proportional to TSP(X), a traveling salesman tour of X. This models the case where a single warehouse (coincident with one of the facilities) supplies goods to the facilities X using a single truck to transport goods from the warehouse to the facilities. This transportation is facilitated via line haul transportation on a so-called *peddling route* (Langevin, Mbaraga, and Campbell 1996)—that is, a route consisting of multiple stops.

3. The costs due to *transportation* from the facilities to the service region are proportional to $\mathrm{Dir}(X, C) := \iint_C \min_i \|x - x_i\| \, dA$, that is, the average length of a direct trip between a point x sampled uniformly in C and its nearest facility x_i (scaled proportionally to the area of C). This models the case where we have a continuum of customers distributed uniformly in C and each one uses their nearest facility, making a single direct round trip to and from the facility. To emphasize that the trip consists of both an outbound and an inbound leg, it is perhaps more appropriate to model the local transportation costs as $2 \cdot \mathrm{Dir}(X, C)$; to keep our notation compact, we will suppress the coefficient "2" and assume that it is included in the relevant input parameters, which are introduced in the next paragraph.

It follows that we can model the total costs due to the facilities X providing service to customers in C as

$$F(X) = \gamma_{|X|} \cdot |X| + \phi \,\mathrm{TSP}(X) + \psi \,\mathrm{Dir}(X, C), \qquad (1)$$

where $\{\gamma_k\}$ is a given sequence and ϕ and ψ are given constants; see Figure 1. The salient property of this model is that line haul transportation via the backbone network benefits from an *economy of scale* because a TSP tour of the points X has decreasing marginal costs as $|X|$ becomes larger (see Beardwood, Halton, and Hammersley 1959, for example, which explains that, for uniformly distributed point sets X, the length of $\mathrm{TSP}(X)$ increases proportionally to $\sqrt{|X|}$ as $|X| \to \infty$). On the other hand, the local transportation costs are modeled via direct trips, and thus the workload at each facility does not benefit from such an economy of scale. This model might be contrasted with Daganzo and Newell (1986), which uses peddling routes (i.e., diminishing marginal returns) in both the backbone network and the local transportation, or Campbell (1990), which uses direct routes in the backbone network but peddling routes in the local transportation (the exact opposite of our setting). One natural instance of our problem arises when one considers the problem faced by stores selling items that require delivery or installation, such as appliances, electronics, food, or furniture: large freight trucks distribute products to showrooms, which are then put on display to customers. These customers then schedule a delivery and installation of their desired product. Because of the time sensitivity of such requests and other complications, economies of scale are much harder to leverage at the local store-to-customer level than at the transshipment level (Campbell 1990, for example, quantifies this when one of the "complications" considered is capacities on the second-stage vehicles). In our work, we make this distinction clear by modeling the facility-to-customer transportation costs using direct trips from facilities to customers but modeling the backbone network line haul between facilities using peddling routes. In §2.1, we show how to generalize this model to address the case where facility-to-customer transportation costs are incurred using multistop trips instead of direct trips under certain assumptions on the lengths of these multistop trips.

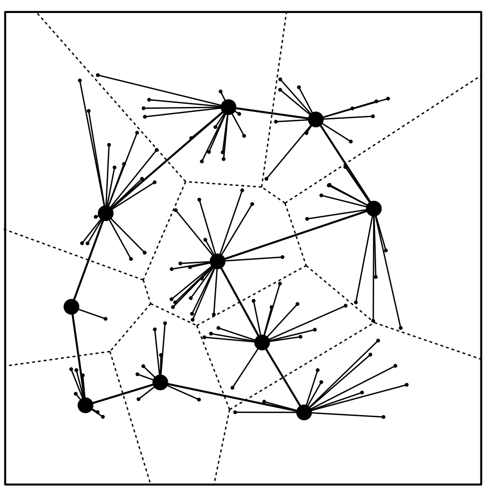

Figure 1 **The Relevant Quantities in Objective Function (1) Where C Is a Square**

Note. The thick black lines indicate the backbone network (a TSP tour), the large points indicate facilities X, the small points indicate various customer locations (which we do not deal with explicitly in our formulation, as they are assumed to be continuously and uniformly distributed in the region), and the dashed lines indicate the service subregions (i.e., the Voronoi cells) of the facilities.

The units of the problem input parameters are given in Table 1, which takes into account the fact that the frequency of trips between the facilities and customers may differ from trips along the backbone network. We reiterate that we allow $k = |X|$ to vary also, i.e., to choose how many facilities to build. This objective function was first introduced in Cachon (2014), in which the author uses several regular polygonal tilings of the plane to estimate the "carbon penalty" of facility configurations—that is, the difference between the carbon cost to a firm and the true externality cost of the total emissions generated. The author shows that, using reasonable estimates of the input parameters, the realized penalty is negligible. In this paper, we make two contributions: In §2, we analyze the objective function (1), and we show that an asymptotically optimal configuration is to distribute the facilities either in a hexagonal tiling or equidistantly along an Archimedean spiral, depending on the nature of the input parameters. Next, we derive two lower bounds for the function (1) for a given convex region C, which we use in a constant-factor approximation algorithm for placing the points X in C so as to minimize the total cost of the facility configuration.

Table 1 Units of the Parameters γ_k, ϕ, and ψ in the Formulation (1)

Parameter	Units	Comments
γ_k	$\frac{\text{cost/facility}}{\text{week}}$	When k facilities are providing service
ϕ	$\frac{\text{cost}}{\text{mile}} \cdot \frac{\text{\# trips required}}{\text{week}}$	For trucks on the backbone network
ψ	$2 \cdot \frac{\text{cost}}{\text{mile}} \cdot \frac{\text{(\# trips required)/customer}}{\text{week}} \cdot \text{(\# customers)}$	For customers in C

Notes. Note that the term "# trips required" may differ from ϕ to ψ, which models the case where the frequency of trips between the facilities and the customers is different from the frequency of trips along the backbone network. Note also the coefficient "2" of ψ, which reflects the assumption that transportation between facilities and customers occurs via direct trips to and from the facility. The model introduced in Cachon (2014) measures "cost" in terms of pounds of CO_2.

1.1. Other Applications

It is worth mentioning that the objective function (1) is a sensible model for carbon emissions as in Cachon (2014), but it also works more generally as a model for spatial one-to-many distribution models with transshipment; elements of such models have previously been examined in Campbell (1990, 1993) and Geoffrion (1979) for example. Rather than minimizing emissions, one might attempt to minimize the actual financial costs incurred by the company in placing its retail stores, warehouses, or other such facilities. In this case, the fixed and backbone network costs are easily understood, though the transportation costs in (1) do not have an obvious counterpart because customers generally bear the cost of travel to retail stores rather than vice versa (with exceptions being businesses that primarily deliver goods to their customers such as large appliances, electronics, food, or furniture), so the company does not directly incur the cost $\psi\,\mathrm{Dir}(X, C)$. In such a case, we should use an alternative model for the transportation costs that possesses some kind of spatial demand component (i.e., that customers near a facility are more likely to use it than those farther away); we will discuss one such model in §2.1.

One might also consider applying this model to the design of an urban transportation network, such as a high-speed rail line. The major difference in this case is that high-speed rail is, for the most part, a "many-to-many" phenomenon, which is distinct from our problem. Our model would perhaps be best suited for the case where a large collection of passengers emanates from a single source, such as a central business district. In this case, rather than using a traveling salesman tour as a backbone network, we might use a minimum spanning tree or a Steiner tree. Our analysis here actually carries over to these alternative backbone networks as well, and we discuss them where appropriate.

1.2. Related Work

The canonical location problem that is most closely related to (FBT) is clearly the *uncapacitated facility location problem* (Li 2011), although the two differ substantially by the inclusion of backbone network costs. A more directly related model to (FBT) is in the seminal paper by Magnanti and Wong (1984), which describes several discrete and continuous models and algorithms for simultaneous facility location and routing. The first explicit formulation of (FBT) to our knowledge is found in Melkote and Daskin (2001), which demonstrates how to solve a hybrid location or routing location problem on a graph as a mixed-integer program; further developments on network formulations of problems of type (FBT) have since emerged (Arnold, Peeters, and Thomas 2004; Miranda and Garrido 2004).

The formulation (1) has previously been discussed (but not solved) in Langevin, Mbaraga, and Campbell (1996), which gives a taxonomy of six classes that differentiate the various continuous approximation models developed for freight distribution problems. The problem of minimizing objective function (1) belongs to class IV, "one-to-many distribution with transshipments," which we can readily observe in Figure 2 of that paper. One important distinction between the models of Langevin, Mbaraga, and Campbell (1996) and our own is that we use the expression $\mathrm{Dir}(X, C) = \iint_C \min_i \|x - x_i\|\, dA$ to model the transportation costs, whereas the corresponding models in Langevin, Mbaraga, and Campbell (1996) use traveling salesman tours originating at the facilities. In §2.1, we will show that the conclusions we derive here are more or less applicable to the approach used therein. Along the same lines, §§5 and 6 of Newell (1973) provide an elegant theoretical justification, using continuum mechanics, for the continuous approximation that this paper employs to describe approximate global optima to the objective functions used herein.

A relatively new branch in location theory deals with *location routing* problems (LRPs) that pay special attention to vehicle routing issues in facility placement (Nagy and Salhi 2007). Such problems are substantially more difficult than the canonical location models because, as Berman, Jaillet, and Simchi-Levi (1995, p. 431) observe,

> [T]he facility ... must be "central" relative to the ensemble of the demand points, as ordered by the (yet unknown) tour through all of them. By contrast, in the classical problems the facility ... must be located by considering distances to individual demand points, thus making the problem more tractable.

One important distinction between the LRP and our problem (FBT) is that we think of the backbone network as connecting the *facilities* together, whereas the LRP considers networks that connect the *customers* together (i.e., vehicle tours that provide service to the customers). In our formulation (1), a parallel argument to the quotation above would be as follows: minimizing the transportation costs, $\mathrm{Dir}(X, C)$, would dictate that we should spread the facilities X as uniformly as possible throughout C and thus be "central" with respect to the customers. However, by pursuing such a strategy too aggressively, we incur large fixed and backbone network costs $\gamma_{|X|} \cdot |X| + \phi\,\mathrm{TSP}(X)$.

Section 2 studies the limiting behavior of the optimal solution to our problem (1) as the transportation coefficient ψ becomes large. As such, our analysis closely resembles other research on the asymptotic behavior of Euclidean optimization problems, such as the traveling salesman problem (Beardwood, Halton, and Hammersley 1959; Bertsimas and Simchi-Levi 1996; Haimovich and Rinnooy Kan 1985) and general *subadditive Euclidean functionals* (Redmond and Yukich 1994; Steele 1981) as well as the *k-center* and *k-medians* problems (Hochbaum 1984; Zemel 1984). Although our analysis is deterministic (as opposed to the cited works that are probabilistic), the spirit of our contribution is most closely related to the aforementioned results.

1.3. Notational Conventions

A quantity that we will use frequently in this paper is the *Fermat–Weber value* of a shape S, $\mathrm{Dir}(S)$, which refers to the quantity

$$\mathrm{Dir}(S) = \inf_{x_0 \in S} \iint_S \|x - x_0\|\, dA,$$

so that x_0 is the point that minimizes the average direct-trip distance between a uniformly selected point in S; i.e., x_0 is the *geometric median* of S. For any shape S, we define $\mathrm{diam}(S)$ to be the *diameter* of S, i.e., $\sup_{x,\, y \in S} \|x - y\|$. For any shape S and any point x_0, we define the distance function

$$D(x_0, S) = \inf_{x \in S} \|x - x_0\|,$$

and for any set $S \subset \mathbb{R}^2$, let $N_\epsilon(S)$ denote the set of all points x within ϵ of S; i.e., $N_\epsilon(S) = \{x\colon D(x, S) \le \epsilon\}$.

For any (possibly infinite) set of points X in a convex region C, we define

$$\mathrm{Dir}(X, C) = \iint_C \inf_{x' \in X} \|x - x'\|\, dA.$$

For any scalar x, we let $\lfloor x \rfloor$ and $\lceil x \rceil$ denote the floor and ceiling functions of x, we let $\lfloor x \rceil$ denote the rounding function of x, and we let $\log(x)$ denote the natural log of x. Finally, we make use of the following four common conventions in asymptotic analysis:

1. We say that $f(x) \in \mathcal{O}(g(x))$ if there exists a constant c and a value x_0 such that $f(x) \le c \cdot g(x)$ for all $x \ge x_0$.
2. We say that $f(x) \in o(g(x))$ if $\lim_{x\to\infty} f(x)/g(x) = 0$.
3. We say that $f(x) \in \omega(g(x))$ if $\lim_{x\to\infty} f(x)/g(x) = \infty$.
4. We say that $f(x) \sim g(x)$ if $\lim_{x\to\infty} f(x)/g(x) = 1$.

2. Asymptotic Analysis

We begin by considering the optimal configurations for minimizing (1) as the various parameters γ_k, ϕ, and ψ change; without loss of generality, we assume that $\mathrm{Area}(C) = 1$. As $\psi \to 0$, representing a very sparsely populated region in which the transportation cost to consumers becomes negligible, the optimal configuration is clearly to place a single facility ($k = 1$) in the region C. We shall devote most of this section to the impact of increasing population density, i.e., $\psi \to \infty$, although before doing so, we will consider two special cases of (1).

Case 1 ($\phi = 0$). If $\phi = 0$, then we do not incur any penalty for the backbone network $\mathrm{TSP}(X)$ of X. The optimal number of points k to place will depend on the behavior of γ_k, but clearly once we have selected k, our objective is merely to distribute those k points as uniformly as possible in C (minimizing $\mathrm{Dir}(X, C)$), without regard to their backbone network. We thus have a kind of "soft constrained" instance of the well-studied *k-medians* problem, and as the optimal k becomes large (equivalently, as $\psi \to \infty$), the optimal solution is known to be the *honeycomb heuristic* (Hochbaum 1984; Newell 1973), which is shown in Figure 2, panel (a).

Case 2 ($\gamma_k = 0$). If $\gamma_k = 0$ for all k, then we do not incur any penalty for placing facilities in the region if they do not lengthen the backbone network. Thus, our optimal configuration will be to place infinitely many facilities along the backbone network (see Figure 3), although we have not yet discussed what the shape of the backbone network should be. One possibility is to use an *Archimedean spiral*, shown in Figure 2, panel (b), with the equation given in polar coordinates as $r = a\theta$. In §A of the online supplement (available as supplemental material at http://dx.doi.org/10.1287/trsc.2013.0511), we show

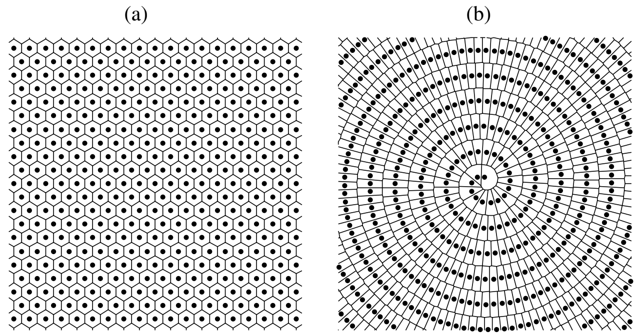

Figure 2 Facility Placement Using the Two Asymptotically Optimal Configurations, the Honeycomb and Archimedes Heuristics

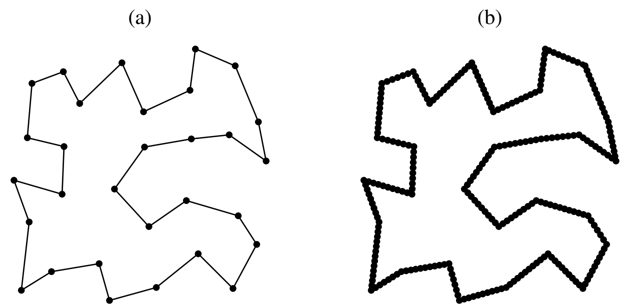

Figure 3 Facility Placement When $\gamma_k = 0$ for All k

Note. If $\gamma_k = 0$ for all k, then the absence of fixed costs for facilities implies that configuration (a) is strictly worse than (b); that is, we should place infinitely many facilities along the backbone network.

that by setting $a = \sqrt{\phi/\psi/\pi}$, the length of the backbone network is $\mathrm{TSP}(X) \sim \sqrt{\psi/\phi}/2$ and therefore that the overall cost, $\phi\,\mathrm{TSP}(X) + \psi\,\mathrm{Dir}(X, C)$, approaches $\sqrt{\phi\psi}$. Using a lower bound, we show in §3 that this configuration, which we call the *Archimedes heuristic*, is actually optimal as $\psi \to \infty$. The Archimedean spiral was previously identified as being an optimal configuration for a related sensor location problem described in Campbell and Nickerson (2011); in that paper, rather than explicitly incurring backbone network costs $\phi\,\mathrm{TSP}(X)$, the objective is to determine a policy for a collection of mobile robots with limited communication radii to convene into a configuration such that all robots can communicate with one another. It turns out that the spiral parameter a varies on the communication radii of the robots.

For notational simplicity, we assume for the remainder of this section that $\phi = 1$ (this is done without loss of generality, since we are examining the limiting behavior as $\psi \to \infty$). Having discussed the two preceding cases, we now sketch a proof of the following claims, which relate γ_k and ψ.

Claim 1. *Suppose that $\gamma_k \in \omega(k^{-1/2})$. As $\psi \to \infty$, the honeycomb heuristic (with appropriately chosen values of $k = |X|$) is an asymptotically optimal configuration for minimizing* (1).

Claim 2. *Conversely to Claim* 1, *if $\gamma_k \in o(k^{-1/2})$, then as $\psi \to \infty$, the Archimedes heuristic (with appropriately chosen values of $k = |X|$) is an asymptotically optimal configuration for minimizing* (1).

We first prove Claim 1 by showing that, as $\psi \to \infty$, the backbone network cost $\mathrm{TSP}(X)$ is dwarfed by the fixed costs $\gamma_{|X|} \cdot |X|$ and the facility-to-customer transportation cost $\psi\,\mathrm{Dir}(X, C)$. Let $k(\psi)$ denote the optimal number of facilities that we should place if our objective is merely to minimize $\gamma_{|X|} \cdot |X| + \psi\,\mathrm{Dir}(X, C)$, i.e., assuming $\phi = 0$ (these must be in a honeycomb configuration, by our earlier argument). Consider the objective cost of a honeycomb configuration of $k(\psi)$ facilities but with $\phi = 1$ as in our original assumption:

$$\gamma_{k(\psi)} \cdot k(\psi) + \mathrm{TSP}(X) + \psi\,\mathrm{Dir}(X, C)$$

(for clarification, we reiterate that, in the above expression, $k(\psi)$ is always defined by assuming $\phi = 0$). Since the Fermat–Weber value of a hexagon H with unit area is given by

$$\alpha_1 := \mathrm{Dir}(H) = \frac{3^{3/4}(4 + 3\log 3)\sqrt{6}}{108} \approx 0.37721,$$

it is straightforward to see that the Fermat–Weber value of a regular hexagon with area A is $\alpha_1 A^{3/2}$. In our case, we have $k(\psi)$ hexagons with area $1/k(\psi)$, and therefore the total Fermat–Weber value of our $k(\psi)$ facilities is simply $k(\psi) \cdot \alpha_1(1/k(\psi))^{3/2} = \alpha_1/\sqrt{k(\psi)}$. Each of these hexagons has sides of length $(\sqrt{2}/3^{3/4}) \cdot (1/\sqrt{k(\psi)})$, and therefore each point has an associated TSP tour segment of length $\beta_1/\sqrt{k(\psi)}$, with $\beta_1 := 3^{-1/4}\sqrt{2} \approx 1.0746$. We therefore find that as $\psi \to \infty$, $k(\psi)$ becomes large, and therefore the objective function value $F(X)$ approaches

$$F(X) \sim \gamma_{k(\psi)} \cdot k(\psi) + \beta_1\sqrt{k(\psi)} + \psi\alpha_1/\sqrt{k(\psi)}.$$

Since $\gamma_k \in \omega(k^{-1/2})$, or equivalently, $\gamma_k \cdot k \in \omega(\sqrt{k})$, we can select a value of ψ large enough so that $k(\psi)$ becomes sufficiently large as to force the quantity

$$\frac{\beta_1\sqrt{k(\psi)}}{\gamma_{k(\psi)} \cdot k(\psi)}$$

to be arbitrarily small. The ratio

$$\frac{\gamma_{k(\psi)} \cdot k(\psi) + \beta_1\sqrt{k(\psi)} + \psi\alpha_1/\sqrt{k(\psi)}}{\gamma_{k(\psi)} \cdot k(\psi) + \psi\alpha_1/\sqrt{k(\psi)}}$$

therefore converges to 1, which proves the asymptotic optimality of the honeycomb heuristic since the denominator of the above expression is clearly a lower bound for our original problem.

To prove Claim 2, we consider a set X of k points distributed equidistantly in C along an Archimedean spiral having length l.

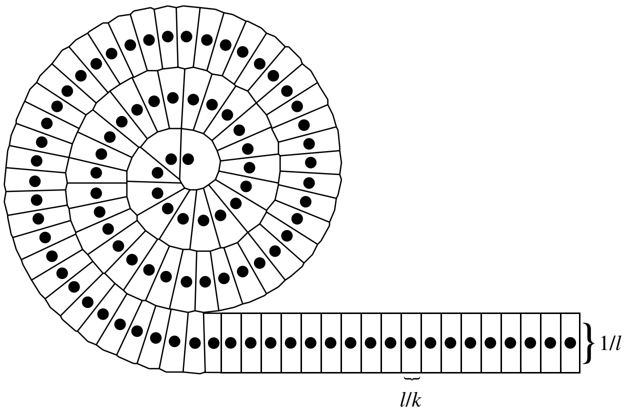

Figure 4 **A Sufficiently Long Archimedian Spiral of Length l, Where a Set X of k Points Is Distributed Equidistantly in C**

Note. For this spiral, it is easy to see that the Voronoi cells of each point are approximately rectangular, with dimensions $l/k \times 1/l$ (this is because there are k such cells, and we assume that the area of C is 1).

As k becomes large, the Voronoi cells of these points can be approximated by rectangles R_i having dimensions $l/k \times 1/l$, as shown in Figure 4. We can approximate the Fermat–Weber value of such a rectangle as follows:

$$\begin{aligned}\mathrm{Dir}(R_i) &= \int_{-1/(2l)}^{1/(2l)} \int_{-l/(2k)}^{l/(2k)} \sqrt{x^2+y^2}\,dx\,dy \\ &\approx \int_{-1/(2l)}^{1/(2l)} \int_{-l/(2k)}^{l/(2k)} |x|+|y|\,dx\,dy \\ &= \frac{1}{4kl} + \frac{l}{4k^2}\end{aligned}$$

so that $\mathrm{Dir}(X, C) \approx k \cdot \mathrm{Dir}(R_i) \approx 1/(4l) + l/(4k)$ (our assumption that k is large means that the R_i's will be skinny, which justifies our approximation of the integrand $\sqrt{x^2+y^2} \approx |x|+|y|$ above). When we set $l = \sqrt{\psi}/2$ (which is equivalent to using $a = \sqrt{1/\psi}/\pi$ in the polar equation $r = a\theta$), we find that $\mathrm{Dir}(X, C) \sim 1/2\sqrt{\psi} + \sqrt{\psi}/(8k)$, so that the total objective cost is

$$\begin{aligned}F(X) &\approx \gamma_k \cdot k + \underbrace{\sqrt{\psi}/2}_{\mathrm{TSP}(X)} + \psi \underbrace{\left(\frac{1}{2\sqrt{\psi}} + \frac{\sqrt{\psi}}{8k}\right)}_{\mathrm{Dir}(X,C)} \\ &= \gamma_k \cdot k + \sqrt{\psi} + \frac{\psi^{3/2}}{8k}.\end{aligned}$$

Note that the above expression depends only on $\{\gamma_k\}$ and ψ, which are given, and k, which we are free to choose. Using the fact that $\gamma_k \in o(k^{-1/2})$, or equivalently, $\gamma_k \cdot k \in o(\sqrt{k})$, we show that

$$\lim_{\psi\to\infty} \min_k \left\{\gamma_k \cdot k + \sqrt{\psi} + \frac{\psi^{3/2}}{8k}\right\} \sim \sqrt{\psi},$$

or equivalently, that

$$\lim_{\psi\to\infty} \min_k \left\{\gamma_k \cdot k + \frac{\psi^{3/2}}{8k}\right\} \in o(\sqrt{\psi}).$$

Suppose for the sake of contradiction that the above limit does not hold. If this is the case, then there exists a constant $c > 0$ and an increasing sequence $\{\psi_i\} \to \infty$ such that

$$\gamma_k \cdot k + \frac{\psi_i^{3/2}}{8k} \geq c\sqrt{\psi_i} \tag{2}$$

for all i and k. As $\gamma_k \cdot k \in o(\sqrt{k})$, there exists a threshold $\bar{k}$ such that $\gamma_k \cdot k < (c^{3/2}/8)\sqrt{k}$ for all $k \geq \bar{k}$. Now set $k^* = 4\psi_i/c$ and assume without loss of generality that $k^* \geq \bar{k}$ (otherwise, we simply remove the first few elements of the sequence $\{\psi_i\}$). We find that

$$\gamma_{k^*} \cdot k^* + \frac{\psi^{3/2}}{8k^*} < \frac{9c}{32}\sqrt{\psi_i}$$

for all i, which contradicts (2) because $9c/32 < c$. We therefore find that $F(X) \sim \sqrt{\psi}$ as $\psi \to \infty$ when k is chosen appropriately under the Archimedes heuristic with $a = \sqrt{1/\psi}/\pi$. In §3, we use a lower bounding argument to verify that this configuration is, in fact, optimal, which completes the proof of Claim 2.

Remark 1. Using the lower bound in §3.2, we show that Claims 1 and 2 also hold when we use a minimum spanning tree or Steiner tree as our backbone network instead of a TSP tour. For visual purposes, the Archimedes heuristic may appear somewhat unsatisfying because of the terminal end point in the center of the region. To work around this, an alternative configuration is the "double spiral" shown in Figure 5, panel (a), which inherits the same objec-

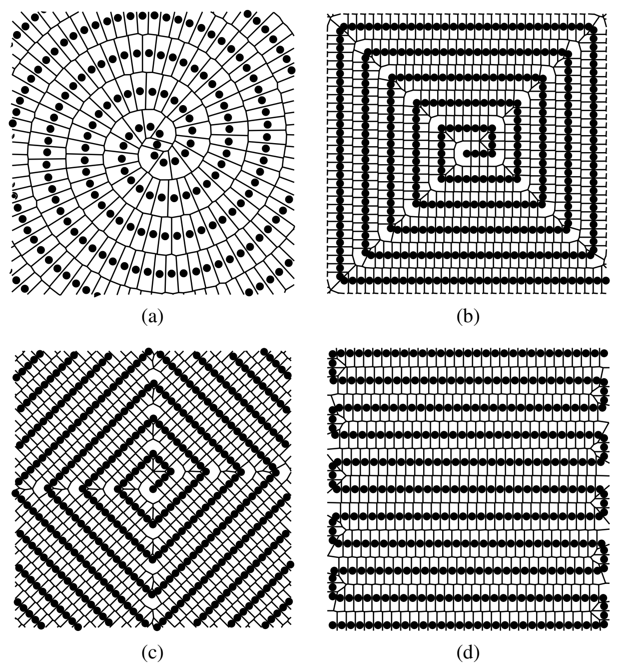

Figure 5 **The "Double Spiral" (a), Facility Placement Under the Square Spirals (b) and Diamond Spirals (c), and "Zig-Zag" Configuration (d)**

Note. The key property that all these share is that, like the Archimedes heuristic, we can "unravel" the service region into a long skinny region with dimensions that are actually optimal for the input parameters, as shown in Figure 4.

tive function properties as the single spiral, without this central singularity.

Remark 2. If we assume that the direct transportation cost $\mathrm{Dir}(X, C)$ is taken under the l_1 or l_∞ norm instead of the Euclidean norm, two optimal structures are the square and diamond spirals shown in Figure 5, panels (b) and (c). Another possibility would be the "zig-zag" configuration shown in Figure 5, panel (d). The key property that all of these configurations share is that we can "unravel" them into a long skinny region while essentially retaining the same objective value, as shown in Figure 4. Since the l_2 norm is rotationally invariant, these configurations are also optimal for our original case where $\mathrm{Dir}(X, C)$ is Euclidean.

Remark 3. Cachon (2014) considers the problem of minimizing (1) for the special case where $\gamma_k = 0$ for all k, and we constrain X to follow the honeycomb heuristic, a square grid, or an equilateral triangular tiling. The author's analysis is summarized as follows: suppose that k facilities are distributed according to the honeycomb heuristic on a region of area 1. Since the Fermat–Weber value of a hexagon H with unit area is given by

$$\alpha_1 := \mathrm{Dir}(H) = \frac{3^{3/4}(4+3\log 3)\sqrt{6}}{108} \approx 0.37721,$$

it is straightforward to see that the Fermat–Weber value of a regular hexagon with area A is $\alpha_1 A^{3/2}$. In our case, we have k hexagons with area $1/k$, and therefore the total Fermat–Weber value of our k facilities is simply $k \cdot \alpha_1 (1/k)^{3/2} = \alpha_1/\sqrt{k}$. Each of these hexagons has sides of length $(\sqrt{2}/3^{3/4}) \cdot (1/\sqrt{k})$, and therefore each point has an associated TSP segment of length $\beta_1/\sqrt{k}$, with $\beta_1 := 3^{-1/4}\sqrt{2} \approx 1.0746$. We therefore find that, as k becomes large, the objective function value $F(X)$ approaches

$$F(X) \approx \underbrace{\phi}_{=1} \beta_1 \sqrt{k} + \psi \frac{\alpha_1}{\sqrt{k}},$$

which is minimized at $k = (\psi\alpha_1)/\beta_1$, at which point the objective function value is

$$2\sqrt{\alpha_1 \beta_1 \psi} \approx 1.2733\sqrt{\psi}. \tag{3}$$

For the square grid, the relevant coefficients (analogous to α_1 and β_1) turn out to be $\alpha_2 \approx 0.38260$ and $\beta_2 = 1$; for the triangular tiling, they are $\alpha_3 \approx 0.40365$ and $\beta_3 \approx 0.87738$. The optimal objective function values for these configurations are $1.2371\sqrt{\psi}$ for the square grid and $1.1902\sqrt{\psi}$ for the triangular layout. Thus, we see that when $\gamma_k \in o(k^{-1/2})$, as $\psi \to \infty$, the Archimedes heuristic outperforms the regular tilings by more than 19%. As an intellectual exercise, we may also consider an irregular configuration such as the

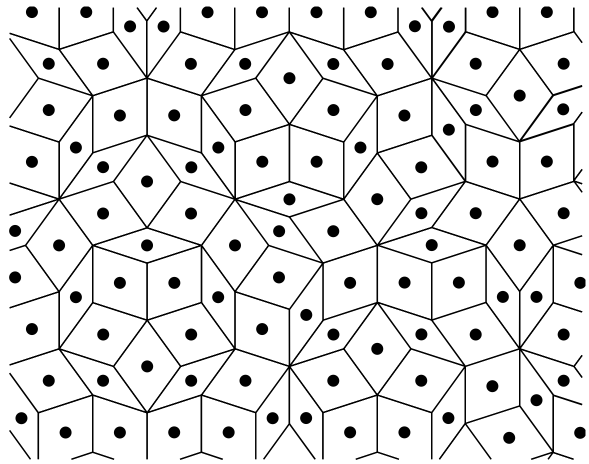

Figure 6 **A Penrose Tiling of Rhombi**

Penrose tiling (Grünbaum and Shephard 2012) shown in Figure 6. We find that the relevant coefficients turn out to be $\alpha_4 \approx 0.4560$ and $\beta_4 \approx 0.9578$, giving an optimal objective function value of $1.3217\sqrt{\psi}$.

Remark 4. It is natural to consider the case where $\gamma_k = ck^{-1/2}$ for some constant c, so that neither Claim 1 nor Claim 2 holds. The optimal configuration in such a case ought to depend on the value of c; for very small values of c, a "spiral-like" configuration ought to be near optimal, and for larger values of c, we expect honeycomb-type configurations to be preferable, as shown in Figure 7.

2.1. Alternative Cost Models

2.1.1. A Gravity Model of Demand. The *gravity hypothesis* (Sheppard 1978) is a well-known geographic theory that states that the "interaction" between two points x and y decays at a rate proportional to the inverse square of the distance between them, i.e., $1/\|x-y\|^2$. Here, "interaction" might be measured by economic activity (Bergeijk and Brakman 2010) or transport (Rodrigue, Comtois, and Slack 2009), for example. We can design a spatial utility model based around the gravity hypothesis by postulating that, if a customer at point x is within the service region of point x_i (i.e., nearer to x_i than any other facility point x_j), then

$$\Pr(\text{customer at } x \text{ uses } x_i) = \frac{1}{(1+\alpha\|x-x_i\|)^2},$$

where α is a decay parameter (the "1+" term in the denominator ensures that we have quadratic decay but that the customer uses the facility with probability 1 if $x = x_i$). The total amount of demand served by the facilities X is then proportional to $\tilde{D}(X, C)$, defined as

$$\tilde{D}(X, C) := \iint_C \frac{dA}{(1+\alpha \min_i\{\|x-x_i\|\})^2}$$

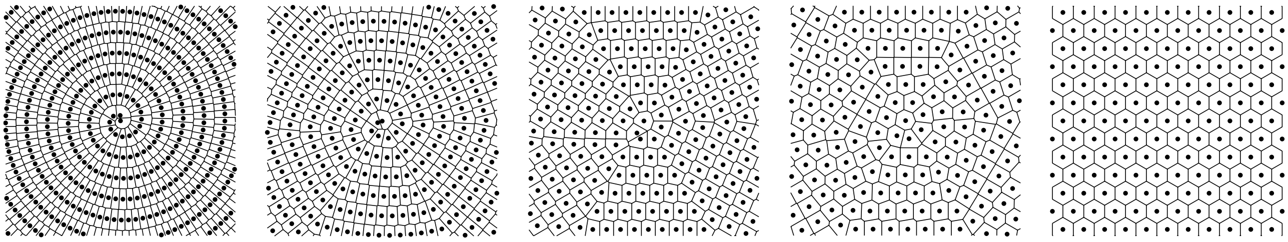

Figure 7 Near-Optimal Configurations for (1) When $\gamma_k = ck^{-1/2}$ for Increasing Values of c

$$= \sum_{i=1}^{k} \iint_{V_i} \frac{dA}{(1+\alpha\|x-x_i\|)^2},$$

where $\mathcal{V} = \{V_1, \ldots, V_k\}$ denotes a Voronoi partition of C with respect to the points X. Since a firm wants $\tilde{D}(X, C)$ to be as large as possible while keeping fixed costs and backbone network costs small, we thus consider the alternative model of (1) given by minimizing

$$\bar{F}(X) = \gamma_{|X|} \cdot |X| + \phi\,\mathrm{TSP}(X) - \psi\tilde{D}(X, C).$$

As in the preceding section, we can analyze the asymptotic behavior of this model when $\psi \to \infty$ by considering the optimal facility placement under the special cases where $\phi = 0$ and $\gamma_k = 0$. Applying a monotonicity argument to that of Hochbaum (1984), it is intuitive that when $\phi = 0$, the optimal solution is again the honeycomb heuristic. When $\gamma_k = 0$, the optimal solution is an Archimedean spiral with length $\sqrt{\alpha\psi/2} - \alpha/2 \sim \sqrt{\alpha\psi/2}$. We can easily verify the counterparts to Claims 1 and 2 accordingly.

2.1.2. Multitrip Costs. As described in §1, the facility-to-customer transportation costs $\psi\,\mathrm{Dir}(X, C)$ model the case where we have a continuum of customers distributed uniformly in C, and the cost due to each customer is proportional to the distance between that customer and its nearest facility x_i (a single direct round trip between the facility and the customer). We can extend this model to consider the case where a vehicle makes multiple trips to customers, starting and ending at the facility, if we adopt the same assumptions as in Newell and Daganzo (1986), which are explained below.

More specifically, we suppose that a total of N customers are distributed uniformly in C, and let $\psi' = \psi/N$ so that the transportation costs (when direct trips are used) can equivalently be written as $\psi' N \cdot \mathrm{Dir}(X, C)$; this allows us to describe the transportation costs in terms of the number of customers. In our alternative model, we assume that a service vehicle can visit m customers before a return trip to the facility is required. We let V_i denote the service subregion (the Voronoi cell) associated with x_i, and suppose that a vehicle based at x_i visits the customers in V_i, of which there are N_i in total. The main assumption of Newell and Daganzo (1986) (as stated in the opening paragraph thereof) that we will adopt is that $m \ll N_i$; this simply models the case where there are many required vehicle tours in each subregion V_i, which might be imposed by lengthy service times at customer locations or limited vehicle capacities (see Campbell 1990 for a thorough discussion and Bertsimas and Simchi-Levi 1996; Haimovich and Rinnooy Kan 1985 for probabilistic and worst-case analyses under the same assumptions).

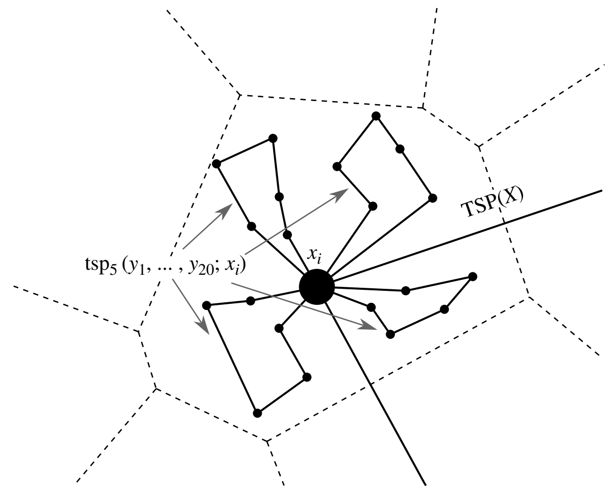

Figure 8 The Service Region V_i Associated with Facility x_i and the Four Tours of the "Service Vehicle" to Visit the $N_i = 20$ Customers y_i in the Region ($m = 5$ in This Example)

Let $Y = \{y_1, \ldots, y_{N_i}\}$ denote the set of customers distributed in V_i; we write the cost of servicing the set of customers in V_i as $\mathrm{tsp}_m(Y; x_i)$, where we use the lowercase notation "tsp" to reflect the notion that this travel is happening locally *within* V_i, as opposed to the backbone network costs $\mathrm{TSP}(X)$, which occur between facilities; see Figure 8. Each vehicle's route will consist of at most $m+1$ stops, namely, a set of m customers plus the starting point x_i; the original case where transportation costs are $\psi\,\mathrm{Dir}(X, C)$ is simply the special case of this new formulation in which $m = 1$.

Since we assume that the customers are distributed uniformly at random in V_i, we define

$$\mathrm{Etsp}_m(V_i, x_i) = \mathbf{E}[\mathrm{tsp}_m(Y; x_i)]$$

as the expected distance that the vehicle serving region V_i will traverse. The total expected transportation cost to customers is then given by $(\psi'/2)\,\mathrm{Etsp}_m(V_i, x_i)$, where we have used a coefficient $\psi'/2$ rather than ψ' because of the implicit multiplier "2" that we introduced when we defined ψ in Table 1. The purpose of this multiplier was to account for the inbound and outbound components of travel (which we no longer have to consider in such a fashion in the multistop model). Theorem 4 of Haimovich and Rinnooy Kan (1985) states that, provided m is fixed, we have

$$\mathrm{Etsp}_m(V_i, x_i) \sim \frac{2N_i}{m} \cdot \frac{\iint_{V_i} \|x - x_i\|\, dA}{\mathrm{Area}(V_i)}$$

as $N_i \to \infty$. The survey by Bertsimas and Simchi-Levi (1996) provides an intuitive justification for this relationship in the following passage, where we have replaced some of the original notations with our own for the sake of consistency:

> Any solution for the capacitated VRP [vehicle routing problem] has two cost components; the first component is proportional to the total "radial" cost between the depot and the customers. The second component is proportional to the "circular" cost; the cost of traveling between customers. This cost is related to the cost of the optimal traveling salesman tour. It is well known (Beardwood, Halton, and Hammersley 1959) that, for large N_i the cost of the optimal traveling salesman tour grows like $\sqrt{N_i}$, while the total radial cost between the depot and the customers grows like N_i because the number of vehicles used in any solution is at least $\lceil N_i/m \rceil$. Therefore, it is intuitive that when the number of customers is large enough the first cost component will dominate the optimal solution value. (p. 288)

Returning to our problem, we see that since $N_i \sim N \cdot \mathrm{Area}(V_i)$ as $N \to \infty$ (with probability 1), it must follow that the total transportation cost in the region is then

$$\begin{aligned}&\sum_{i=1}^{k} (\psi'/2)\,\mathrm{Etsp}_m(V_i, x_i)\\ &\quad\sim \sum_{i=1}^{k} (\psi'/2)\left(\frac{2N_i}{m} \cdot \frac{\iint_{V_i} \|x - x_i\|\, dA}{\mathrm{Area}(V_i)}\right)\\ &\quad\sim \sum_{i=1}^{k} \frac{\psi' \cdot N}{m} \iint_{V_i} \|x - x_i\|\, dA = \frac{\psi}{m}\,\mathrm{Dir}(X, C),\end{aligned}$$

which differs from the transportation cost in our initial formulation merely by a factor of $1/m$. Thus, we find that the introduction of a multistop model for transportation cost within subregions does not alter our model in a fundamental way, provided that the number of stops allowed on a vehicle tour, m, is small relative to the total number of clients in each subregion, N_i.

2.1.3. Competition with Backbone Network Costs. One might also view the preceding result within the context of competitive location problems, in which the objective is to find the best location for a new facility to attract the most buying power away from existing facilities, and conversely, to determine the optimal location for the "defending" facilities to minimize the attractive power of the new facility. When the number of facilities is fixed and $\phi = 0$, the honeycomb heuristic is currently the best-known "defensive" configuration for facilities in the plane, and in Drezner and Zemel (1993), it is shown that it is within 2.5% of a lower bound for the optimal such solution.

Another variation would be the problem of constructing a "defensive" configuration of facilities that also takes into account the cost of the backbone network. We can write this problem as

$$\underset{X}{\text{minimize}} \left\{ \gamma_{|X|} \cdot |X| + (\phi = 1)\,\mathrm{TSP}(X) + \psi \max_{p \in C} S(p \mid X) \right\}, \tag{4}$$

where $S(p \mid X)$ denotes the amount of area that "attacking" facility p "steals" from the facilities X (see Figure 9). From Drezner and Zemel (1993), we have bounds for $S(p \mid X)$ for the square grid, triangular tiling, and honeycomb heuristic, which are reproduced in Table 2.[1] Using the values ζ_i as in the table, and reintroducing the values β_i from earlier (the "TSP coefficients"), we find that our objective can be written as

$$\underset{k}{\text{minimize}} \{ \gamma_k \cdot k + (\phi = 1)\beta_i \sqrt{k} + \psi \zeta_i / k \},$$

whose asymptotic behavior we will again analyze in terms of $\{\gamma_k\}$ as $\psi \to \infty$. As before, it is obvious that if $\gamma_k \in \omega(k^{-1/2})$, then the fixed costs dwarf the backbone network costs as $\psi \to \infty$ (since $k \to \infty$ also), and the honeycomb heuristic is therefore asymptotically optimal (assuming that the honeycomb heuristic is indeed the optimal competitive configuration when there is no backbone network cost, as conjectured in Drezner and Zemel 1993).

The case where $\gamma_k \in o(k^{-1/2})$ is slightly more involved. If we consider the problem where fixed costs are omitted, i.e.,

$$\underset{X}{\text{minimize}} \left\{ (\phi = 1)\,\mathrm{TSP}(X) + \psi \max_{p \in C} S(p \mid X) \right\},$$

then we find that optimal objective value is $(3 \cdot 2^{1/3}/2)\beta_i^{2/3}\psi^{1/3}\zeta_i^{1/3}$ for each of the regular tilings, namely, $1.587\psi^{1/3}$ for the honeycomb placement, $1.560\psi^{1/3}$ for the square grid, and $1.513\psi^{1/3}$ for the

[1] The paper by Drezner and Zemel (1993) actually proves that $\zeta_3 \geq 2/3$, although numerical simulations strongly suggest that equality holds.

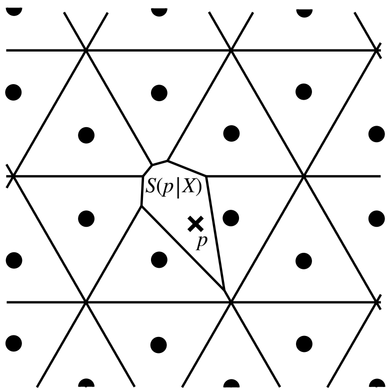

Figure 9 **Competitive Placement of a Facility p, Where "Defending" Facilities Are Distributed in a Triangular Tiling (the Above Placement of p Is Suboptimal)**

triangular layout. This is because we can write the problem as

$$\underset{X}{\text{minimize}}\{(\phi=1)\beta_i\sqrt{k}+\psi\zeta_i/k\}$$

and subsequently solve for the optimal k.

However, if we place an infinite number of facilities together along an Archimedean spiral of length l, then as Figure 10 shows, the maximum area that the attacker p can "steal" from the defending facilities X, $S(p \mid X)$, is approximately $1/(3l^2)$. Thus we consider the problem

$$\underset{l}{\text{minimize}}\left\{(\phi=1)l+\frac{\psi}{3l^2}\right\},$$

which has an optimal solution $l^* = 18^{1/3}\psi^{1/3}/3$, at which point the objective value is $c\psi^{1/3}$, where $c = 18^{1/3}/2 \approx 1.310$. We conjecture that the Archimedean spiral is an optimally competitive configuration for this problem, although a rigorous proof appears difficult. In §B of the online supplement, we

Table 2 **Upper Bounds for $S(p \mid X)$, the Amount of Area That an "Attacking" Facility p Can "Steal" from the "Defending" Facilities X**

Configuration	Constant	Upper bound
Honeycomb	ζ_1	0.5127
Square	ζ_2	0.5625
Triangular	ζ_3	2/3

Notes. The values ζ_i are defined as follows: if $|X| = k$ and Area$(C) = 1$, then each facility's service region (Voronoi cell) will have area $1/k$. An attacking facility can "steal" at most ζ_i/k from X; in other words, ζ_i represents the amount of area that an attacking facility can steal, measured as a fraction of the area of any defending facility's service region.

show that if $\gamma_k \in o(k^{-1/2})$, then the optimal objective function cost is $c\psi^{1/3} + o(\psi^{1/3})$.

3. Lower Bounds in a Convex Region

In this section, we introduce some lower bounds for the objective function $F(X)$ defined on a given convex region C. We begin with a collection of bounds relating $\text{Dir}(X, C)$, $\text{TSP}(X)$, and $|X| = k$.

3.1. Bounds for $\text{Dir}(X, C)$ and $\text{TSP}(X)$

Lemma 1. *For any region C with area A, we have*

$$\text{Dir}(C) \geq \frac{2A^{3/2}}{3\sqrt{\pi}}.$$

Proof. It is well known that, for a fixed area A, a disk with radius $\sqrt{A/\pi}$ is the region with minimal Fermat–Weber value. The above expression is merely the Fermat–Weber value of such a disk. □

Corollary 1. *For any region C with area A containing a set of points $X = \{x_1, \ldots, x_k\}$, we have* $\text{Dir}(X, C) \geq 2A^{3/2}/(3\sqrt{\pi k})$.

Theorem 1. *Suppose that $X = \{x_1, \ldots, x_k\}$ is a set of points in a convex polygon C such that* $\text{TSP}(X) = l$

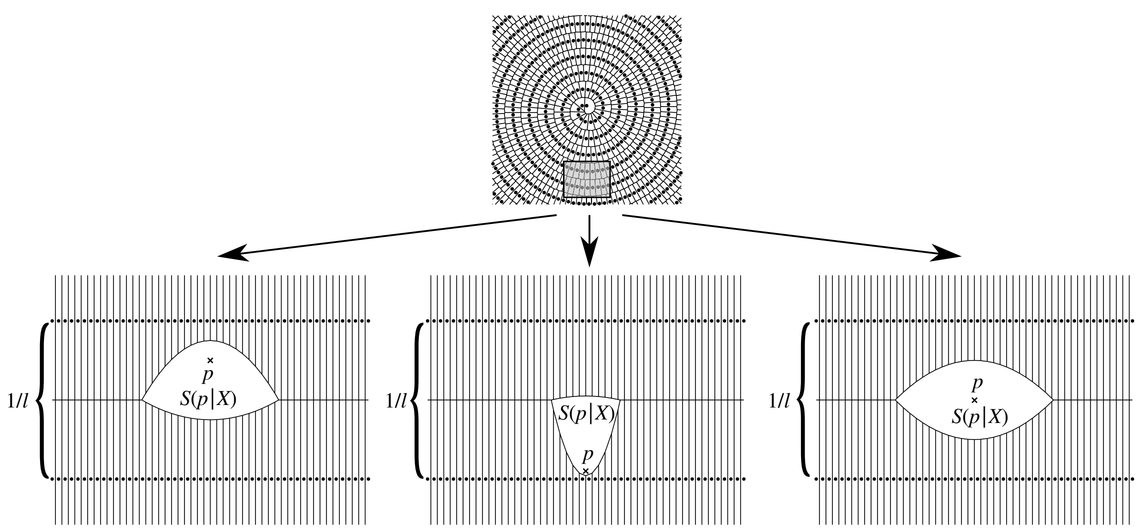

Figure 10 **Three Possible Locations for the Attacking Facility Against the Archimedes Configuration**

Note. Not surprisingly, the third location (exactly between the two arcs of the spiral) maximizes the amount of area that the attacking facility can steal.

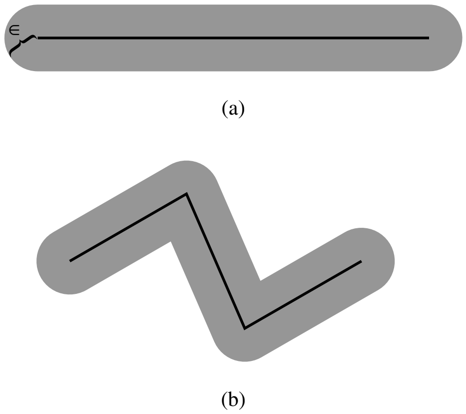

Figure 11 **The Neighborhoods $N_\epsilon(P)$ for Two Paths of the Same Length, a Line Segment (a) and a Collection of Segments (b)**

and $\mathrm{Area}(C) = A$*. Then* $\mathrm{Dir}(X, C) \geq 2A^{3/2}/(3\sqrt{\pi k})$ *and* $\mathrm{Dir}(X, C) \geq 2A^2/(8l + 3\sqrt{\pi A})$.

Proof. The first inequality follows from Lemma 1. The second follows via two simple lemmas, which we will now prove. We first make the observation that if P is a TSP tour of the points X, then obviously $\mathrm{Dir}(X, C) \geq \mathrm{Dir}(P, C)$ (since $X \subset P$), and therefore we can consider bounding the quantity $\mathrm{Dir}(P, C)$ over all paths with a given length l.

Lemma 2. *For any path P of length l and any ϵ, we have* $\mathrm{Area}(N_\epsilon(P)) \leq \pi\epsilon^2 + 2\epsilon l$*, which is tight when P is a line segment.*

Proof. Note this lemma applies to *all* paths, and not merely those that are closed (that would be the most appropriate setting for TSP tours, although the analysis thereof appears difficult). We prove this by induction on the number of line segments n that comprise P. The base case $n = 1$ is simply a line segment for which $N_\epsilon(P)$ is shown in Figure 11, panel (a). To complete the induction, consider a path consisting of n line segments, which we can think of as the union of a path P' with length l' with $n-1$ line segments and a line segment s of length l'' such that $l = l' + l''$. Let $P' \cup s = P$ denote their union. Since P' and s are joined at a point, the neighborhoods $N_\epsilon(P')$ and $N_\epsilon(s)$ must both contain a ball of radius ϵ about their point of intersection; in other words, we have $\mathrm{Area}(N_\epsilon(P') \cap N_\epsilon(s)) \geq \pi\epsilon^2$, and therefore we find that

$$\begin{aligned}
&\mathrm{Area}(N_\epsilon(P))\\
&\quad = \mathrm{Area}(N_\epsilon(P' \cup s)) = \mathrm{Area}(N_\epsilon(P') \cup N_\epsilon(s))\\
&\quad = \underbrace{\mathrm{Area}(N_\epsilon(P'))}_{\leq \pi\epsilon^2 + 2\epsilon l'} + \underbrace{\mathrm{Area}(N_\epsilon(s))}_{\leq \pi\epsilon^2 + 2\epsilon l''} - \underbrace{\mathrm{Area}(N_\epsilon(P') \cap N_\epsilon(s))}_{\geq \pi\epsilon^2},
\end{aligned}$$

and the desired result follows. □

Lemma 3. *Let P denote a path with length l and let C denote a planar region with area A containing P. Further, let L denote a line segment with length l and let $\epsilon_L^{\max}$ be chosen so that* $\mathrm{Area}(N_{\epsilon_L^{\max}}(L)) = A$*. Then* $\mathrm{Dir}(L, N_{\epsilon_L^{\max}}(L)) \leq \mathrm{Dir}(P, C)$*; in other words, for a given area A, among all paths with fixed length l, a line segment and its appropriately chosen neighborhood have the minimal Fermat–Weber value.*

Proof. Assume without loss of generality that $A = 1$, and let $\epsilon_P^{\max}$ be chosen so that $\mathrm{Area}(N_{\epsilon_P^{\max}}(P)) = A = 1$. It is obvious that $\mathrm{Dir}(P, N_{\epsilon_P^{\max}}(P)) \leq \mathrm{Dir}(P, C)$ for all regions C with area 1. Thus it will suffice to show that $\mathrm{Dir}(P, N_{\epsilon_P^{\max}}(P)) \geq \mathrm{Dir}(L, N_{\epsilon_L^{\max}}(L))$. Consider a random variable ϵ_P defined by setting $\epsilon_P := D(z, P)$, where z is a random variable sampled uniformly from $N_{\epsilon_P^{\max}}(P)$, and define ϵ_L similarly. Note that the cumulative distribution functions for ϵ_P and ϵ_L are given by

$$F_P(\epsilon_P) = \min\{1, \mathrm{Area}(N_{\epsilon_P}(P))\},$$
$$F_L(\epsilon_L) = \min\{1, \mathrm{Area}(N_{\epsilon_L}(L))\}.$$

By Lemma 2, for any $\epsilon > 0$, we have $F_L(\epsilon) \geq F_P(\epsilon)$. Next note that

$$\mathbf{E}(\epsilon_L) = \int_0^\infty \big(1 - F_L(\epsilon)\big)\, d\epsilon \leq \int_0^\infty \big(1 - F_P(\epsilon)\big)\, d\epsilon = \mathbf{E}(\epsilon_P),$$

a well-known result of first-order stochastic dominance (see p. 249 of Pinedo 2008, for instance). The proof is complete by observing that, by definition, $\mathbf{E}(\epsilon_L) = \mathrm{Dir}(L, N_{\epsilon_L}(L))$ and $\mathbf{E}(\epsilon_P) = \mathrm{Dir}(P, N_{\epsilon_P}(P))$. □

Having established the two preceding lemmas, we next note that, for any line segment L with length l and any ϵ, we can compute

$$\begin{aligned}
\mathrm{Area}(N_\epsilon(L)) &= \pi\epsilon^2 + 2\epsilon l,\\
\mathrm{Dir}(L, N_\epsilon(L)) &= \frac{2\pi\epsilon^3}{3} + \epsilon^2 l.
\end{aligned} \tag{5}$$

Solving Equation (5) in terms of $\epsilon > 0$ and substituting, we find that

$$\begin{aligned}
&\mathrm{Dir}(L, N_\epsilon(L))\\
&\quad = \big(-2l^3 - 3\pi l \mathrm{Area}(N_\epsilon(L)) + (2l^2 + 2\pi \mathrm{Area}(N_\epsilon(L)))\\
&\qquad \cdot \sqrt{l^2 + \pi \mathrm{Area}(N_\epsilon(L))}\big)\big/(3\pi^2)\\
&\quad \geq \frac{2A^2}{8l + 3\sqrt{\pi A}},
\end{aligned}$$

where $A = \mathrm{Area}(N_\epsilon(L))$. (We have performed some routine calculations in the inequality above, which we omit for brevity.)

Our proof of Theorem 1 is complete; if we let P be a TSP tour of any point set X, contained in a region C with area A, then

$$\begin{aligned}
\mathrm{Dir}(X, C) &\geq \mathrm{Dir}(P, C) \geq \mathrm{Dir}(P, N_{\epsilon_P^{\max}}(P))\\
&\geq \mathrm{Dir}(L, N_{\epsilon_L^{\max}}(L)) \geq \frac{2A^2}{8l + 3\sqrt{\pi A}},
\end{aligned}$$

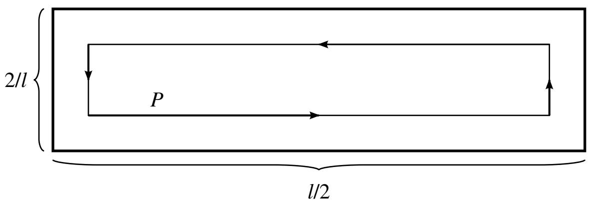

Figure 12 **Rectangle of Dimensions $(l/2)\times(2/l)$ Containing a Path P of Length l**

Note. When C is the rectangle shown above and P is the path indicated, we find that as $l\to\infty$, we have $\mathrm{Dir}(P, C)\sim 1/(4l)$, so that the second lower bound of Theorem 1 becomes tight.

where length(P) = length(L) and $\epsilon_P^{\max}$ and $\epsilon_L^{\max}$ are chosen so as to induce the appropriate areas. □

The second bound of Theorem 1 is useful because it establishes the inverse proportionality between the backbone line haul network length $l = \mathrm{TSP}(X)$ and the transportation cost to customers, $\mathrm{Dir}(X, C)$. More broadly, one could use this result to understand the best-possible marginal improvement to local transportation $\mathrm{Dir}(X, C)$ one could obtain by lengthening the backbone network.

Remark 5. In addition to giving us a lower bound, Theorem 1 also suggests what kind of *region* ought to be efficient for our original problem: consider a rectangle of dimensions $(l/2)\times(2/l)$ containing a path P of length l, as shown in Figure 12. It is not hard to verify that $\mathrm{Dir}(P, C)\sim 1/(4l)$ as $l\to\infty$, so that the second bound of Theorem 1 becomes tight. This equivalently suggests that such regions ought to be optimal for instances of our original problem (1), much in the same spirit as the design of "zones" in Figure 2 of Newell and Daganzo (1986).

Before describing our approximation algorithm, we must introduce an additional lower bound, which applies when C is a particularly long skinny region. To quantify this, we orient C so that $\mathrm{diam}(C)$ is aligned with the coordinate x axis, and we assume without loss of generality that C is contained in a box of dimensions $(\mathrm{diam}(C) = w)\times h$, where $w\ge h$. By convexity of C, it immediately follows that

$$wh/2 \le \mathrm{Area}(C) \le wh.$$

Lemma 4. *For any region C with area A contained between two lines a distance h apart, we have* $\mathrm{Dir}(C)\ge A^2/(4h)$.

Proof. Assume without loss of generality that the two lines in question are horizontal and consider any point $x_0 = (x_1^0, x_2^0)$ between them. Clearly, for any point $x = (x_1, x_2)$, we have $\|x_0 - x\| \ge |x_1^0 - x_1|$. It is easy to see that the region that minimizes $\iint_C |x_1^0 - x_1|\,dA$ (subject to the constraint that C be contained between the two lines) is simply a rectangle with height h and width A/h. The value $\iint_C |x_1^0 - x_1|\,dA$, which is a lower bound on the Fermat–Weber value of such a rectangle, is precisely $\int_0^h\int_{-A/(2h)}^{A/(2h)} |x|\,dx\,dy = A^2/(4h)$ as desired. □

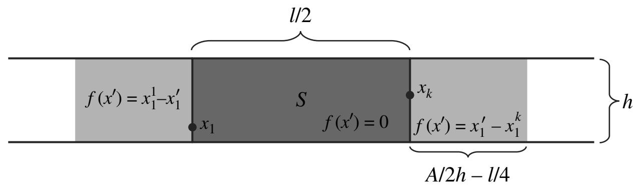

Figure 13 **The Distribution of Area That Minimizes $f(x')$**

Theorem 2. *Suppose that $X = \{x_1, \ldots, x_k\}$ is a set of points in a convex polygon C such that* $\mathrm{TSP}(X) = l < A/h$. *Then* $\mathrm{Dir}(X, C)\ge A^2/(4hk)$ *and* $\mathrm{Dir}(X, C)\ge (A - hl)^2/(4h)$.

Proof. The first inequality follows immediately from Lemma 4. Assume without loss of generality that $x_1 = (x_1^1, x_2^1)$ is the leftmost point in X and $x_k = (x_1^k, x_2^k)$ is the rightmost point in X. Clearly, $x_1^k - x_1^1 \le l/2$ since a TSP tour must return to its starting point. The maximum amount of area of C that can be contained in the slab S between the lines $l_1 = \{(x_1, x_2)\colon x_1 = x_1^1\}$ and $l_2 = \{(x_1, x_2)\colon x_1 = x_1^k\}$ is $hl/2$. Thus, we have at least $A - hl/2$ units of area of C remaining to distribute outside S. Let $x' = (x_1', x_2')$ be a dummy variable and consider the function $f(x')$ defined by

$$f(x') = \begin{cases} x_1^1 - x_1' & \text{if } x_1' \le x_1^1,\\ 0 & \text{if } x_1^1 < x_1' < x_1^k,\\ x_1' - x_1^k & \text{otherwise,}\end{cases}$$

and note that clearly $f(x')\le \min_{x_i\in X}\|x' - x_i\|$. We now consider the problem of distributing the remaining $A - hl/2$ units of area in the rectangle so as to minimize the integral of $f(x')$. The obvious solution is to distribute $A/2 - hl/4$ units of area to the right and left of S in rectangles of dimensions $A/(2h) - l/(4\times h)$, as shown in Figure 13. The integral of $f(x')$ over this shape is precisely $(A - hl/2)^2/(4h)$ as desired. □

3.2. Minimizing $F(X)$

Having proven Theorems 1 and 2, we can derive lower bounds on the objective function $F(X)$ of (1) by solving the following optimization problems, where we let l and z denote variables corresponding to

TSP(X) and Dir(X, C), and we assume (for now) that $k = |X|$ is fixed:

$$\begin{aligned}
&\underset{l,z}{\text{minimize}} \quad \{\gamma_k \cdot k + \phi l + \psi z\}\\
&\text{s.t.} \quad z \geq \frac{2A^{3/2}}{3\sqrt{\pi k}},\\
&\qquad z \geq \frac{2A^2}{8l + 3\sqrt{\pi A}},\\
&\qquad z, l \geq 0;
\end{aligned} \tag{6}$$

$$\begin{aligned}
&\underset{l,z}{\text{minimize}} \quad \{\gamma_k \cdot k + \phi l + \psi z\}\\
&\text{s.t.} \quad z \geq \frac{A^2}{4hk},\\
&\qquad z \geq \frac{(A - hl/2)^2}{4h} \mathbb{1}(l/2 < A/h),\\
&\qquad z, l \geq 0.
\end{aligned} \tag{7}$$

Here, $\mathbb{1}(\cdot)$ denotes the indicator function. It is easy to show that the optimal objective function value to (6) is

$$\Phi^1_{\mathrm{LB}}(A, \phi, \psi, k) = \gamma_k \cdot k + \begin{cases} \dfrac{3\sqrt{A\pi}}{8}(\sqrt{k/\psi} - 1)\phi + \dfrac{2A^{3/2}\psi^{3/2}}{3\sqrt{\pi k}} & \text{if } \phi \leq \dfrac{16A\psi^2}{9\pi k},\\[2ex] A\sqrt{\phi\psi} - \dfrac{3\phi\sqrt{\pi A}}{8} & \text{if } \dfrac{16A\psi^2}{9\pi k} < \phi \leq \dfrac{16A\psi}{9\pi},\\[2ex] \dfrac{2\psi A^{3/2}}{3\sqrt{\pi}} & \text{otherwise,} \end{cases} \tag{8}$$

and the optimal objective function value to (7) is

$$\Phi^2_{\mathrm{LB}}(A, \phi, \psi, h, k) = \gamma_k \cdot k + \begin{cases} \dfrac{2(1 + 1/\sqrt{k})A}{h}\phi + \dfrac{A^2}{4hk}\psi & \text{if } \phi \leq \dfrac{A\psi}{4\sqrt{k}},\\[2ex] \dfrac{2\phi A}{h} - \dfrac{4\phi^2}{\psi h} & \text{if } \dfrac{A\psi}{4\sqrt{k}} < \phi \leq \dfrac{A\psi}{4},\\[2ex] \dfrac{A^2}{4h}\psi & \text{otherwise;} \end{cases} \tag{9}$$

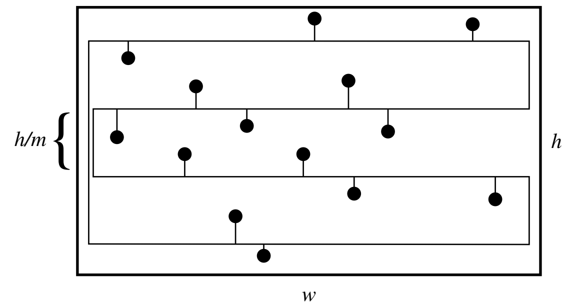

Figure 14 **The TSP Path Construction That Satisfies (11) ($m = 4$ in This Example)**

therefore, a lower bound for our objective function (1) is as follows:

$$\Phi_{\mathrm{LB}}(A, \phi, \psi, h, k) = \max\{\Phi^1_{\mathrm{LB}}(A, \phi, \psi, k), \Phi^2_{\mathrm{LB}}(A, \phi, \psi, h, k)\}. \tag{10}$$

Note that the second case of Φ^1_{LB} can be used to prove that the Archimedes heuristic is asymptotically optimal as $\psi \to \infty$ (for the case where $\gamma_k \in o(k^{-1/2})$), since we have $\Phi^1_{\mathrm{LB}} \sim \sqrt{\psi}$ when $A = 1$, $\phi = 1$, and ψ and k become large as in §2.

4. Upper Bounds in a Convex Region

In this section, we give some upper bounds for TSP(X) and Dir(X, C), which we will use in proving that our algorithm produces a solution within a constant factor of optimality.

Theorem 3. *Suppose that $X = \{x_1, \ldots, x_k\}$ is a set of points distributed in a rectangle R having dimensions $w \times h$. For any positive even integer m, we have*

$$\mathrm{TSP}(X) \leq G(w, h, k) := hk/m + mw + 2\left(\frac{m-1}{m}\right)h. \tag{11}$$

Proof. This is due to Few (1955). We will show how to construct a path through the points X whose length is bounded by the above quantity. One way to form such a path is to move along R horizontally m times, making vertical diversions to touch each of the points, as shown in Figure 14. It is easy to see that, if we ignore the additional work needed to visit the points X, the length of the path is $mw + 2((m-1)/m)h$. The distance from each point x_i to the main path is at most $h/(2m)$, and therefore each such point introduces no more than h/m additional work to the path, which gives the desired result. □

Remark 6. Note that as $hk/w \to \infty$, by setting $m = 2 \cdot \lfloor\sqrt{hk/w}/2\rceil$, we find that $G(w, h, k) \sim 2\sqrt{whk}$. This proportionality to $\sqrt{k}$ is consistent with well-known results in geometric probability such as in Redmond and Yukich (1994) and Steele (1981). A more precise discussion of path constructions of these types can be found in Daganzo (1984).

We can now define a function $H(A,w,h)$, which is an upper bound on the Fermat–Weber value of a convex region with area A contained in a box of dimensions $w\times h$, assuming that $w\geq h$.

Theorem 4. *Suppose that C is a convex region with area A in a rectangle R having dimensions $w\times h$, with $w\geq h$. We have*

$$\begin{aligned}
\mathrm{Dir}(C) &\leq H(A,w,h)\\
&:=\left[\log\left(\frac{h+\sqrt{w^2+h^2}}{wa+w\sqrt{1+a^2}}\right)-a\sqrt{1+a^2}\right]\frac{w^3}{12}\\
&\quad+\left[\log\left(\frac{bw+b\sqrt{w^2+h^2}}{h+h\sqrt{b^2+1}}\right)-\frac{\sqrt{b^2+1}}{b^2}\right]\frac{h^3}{12}\\
&\quad+\frac{wh\sqrt{w^2+h^2}}{6} \qquad \text{if } A<wh-\frac{h}{2}\sqrt{w^2-h^2}\\
&:=\log\left(\frac{h+\sqrt{w^2+h^2}}{w}\right)\cdot\frac{w^3}{12}+\left[\log\left(\frac{cw+c\sqrt{w^2+h^2}}{h+h\sqrt{c^2+1}}\right)\right.\\
&\qquad\left.-\frac{\sqrt{1+c^2}}{c^2}\right]\frac{h^3}{12}+\frac{1}{6}wh\sqrt{w^2+h^2} \qquad \text{otherwise,}
\end{aligned}$$

where

$$a=\frac{w^3h-wh^3-2(wh-A)\sqrt{(w^2+h^2)^2-8hwA+4A^2}}{2Awh-2w^2h^2-w^2\sqrt{(w^2+h^2)^2-8hwA+4A^2}}$$

$$b=\frac{2(wh^3-Ah^2)+wh\sqrt{h^4+2w^2h^2+w^4-8Awh+4A^2}}{w^4+3w^2h^2-8Awh+4A^2}$$

$$c=\frac{h^2}{2(wh-A)}.$$

Proof. See §C of the online supplement. □

Remark 7. It is not hard to show that, if we fix the product wh, then as $h/w\to 0$, we have

$$H(A,w,h)\sim\frac{2}{3}Aw-\frac{1}{12}w^2h-\frac{1}{3}\cdot\frac{A^2}{h} \tag{12}$$

for all A (see §C of the online supplement). It can also be shown that, for fixed w and h, the function $H(A,w,h)$ is concave in A.

5. An Approximation Algorithm

In this section, we describe our simple algorithm for placing facilities in C so as to (approximately) minimize the objective function of (1). In the asymptotic analysis, we assumed (without loss of generality) that $\phi=1$ and that $A=1$; in this section, we shall assume that $\psi=1$ so that ϕ is allowed to vary instead (this simplifies some of the notation and limits the use of fractions). Thus, the input to our algorithm is a convex polygon C, a scalar ϕ, and a sequence $\{\gamma_k\}$ such that $\gamma_k\cdot k$ is increasing. We first describe a subroutine that constructs, for any given integer k, an approximately optimal placement of k facilities; we then apply this subroutine with strategic values of k so as to produce an overall approximation guarantee.

Assume that C is aligned so that its diameter coincides with the coordinate x axis. We then enclose C in an axis-aligned box $\square C$ of dimensions $w\times h$, where we assume without loss of generality that $w=1/h=\mathrm{diam}(C)$; the convexity of C immediately implies that $A=\mathrm{Area}(C)\in[1/2,1]$. In our algorithm, we then let $k_1=\lfloor k/2\rfloor$ and $k_2=\lceil k/2\rceil$ and divide $\square C$ into two pieces of areas $(k_1/k)\cdot\mathrm{Area}(\square C)=k_1/k$ and $(k_2/k)\cdot\mathrm{Area}(\square C)=k_2/k$, respectively, using a vertical line. This is performed recursively (with the option to split using a horizontal line, if the height of an intermediate subregion exceeds its width) until all regions have area $1/k$.

Algorithm 1 takes as input an axis-aligned rectangle R and a positive integer k. (This is used as a subroutine in Algorithm 2.) Algorithm 2 takes as input a convex polygon C and an integer k.

Algorithm 1 (Algorithm RectanglePartition(R,k))
Input: An Axis-aligned retangle R and an integer k.
Output: A partition of R into k rectangles, each of whose area is Area $(R)/k$.

```
if k = 1 then
  return R;
else
  Set k_1 = ⌊k/2⌋ and k_2 = ⌈k/2⌉;
  Let w denote the width of R and h the height;
  if w > h then
    With a vertical line, divided R into two
    pieces R_1 and R_2 with area k_1/k · Area (R)
    on the right and k_2/k · Area (R) on the left;
  else
    With a horizontal line, divided R into two
    pieces R_1 and R_2 with area k_1/k · Area (R)
    on the top and k_2/k · Area R on the bottom;
  end
  return RectanglePartition(R_1, k_1) ∪
    RectanglePartition(R_2, k_2);
end
```

This is described in Algorithms 1 and 2 and Figures 15 and 16. As an aside, it turns out that Algorithm 2 is a constant-factor approximation algorithm for the continuous k-medians problem in a convex polygon (that is, minimizing $\mathrm{Dir}(X,C)$ in a given convex polygon C with a constraint that $|X|=k$), with approximation constant 2.74 (Carlsson, Jia, and Li 2014).

Algorithm 2 (Algorithm ApproxFW(C,k))
Input: A convex polygon C and an integer k.
Output: The locations of k points p_i in C that approximately minimize $\mathrm{Dir}(C,k)$ within a factor of 2.74.

```
Align a diameter of C with the coordinate x-axis;
Let □C denote an axis-aligned box of dimensions
  w × h, where w = diam(C);
```

Let $R_1, \ldots, R_k = \text{RectanglePartition}\ (\square C, k)$;
for $i \in \{1, \ldots, k\}$ **do**
  Let c_i denote the center of R_i;
  if $c_i \in C$ **then**
    Set $p_i = c_i$;
  else
    if $R_i \cap C$ *is nonempty* **then**
      Let R_i' be the minimum axis-aligned bounding box of $R_i \cap C$ and let c_i' denote its center;
      Set $p_i = c_i'$;
    else
      Place p_i anywhere in C;
    end
  end
end
return $p_1, \ldots, p_k$.

Definition 1. The *aspect ratio* of a rectangle R, written $\mathrm{AR}(R)$, is the ratio of the longer side of R to the shorter side.

Before defining the appropriate values of k that should be passed to Algorithm 2 to solve our problem, we state the following claim.

Algorithm 3 (Algorithm $\text{FacilityPlacement}(C, \phi, \{\gamma_k\})$)
Input: A convex polygon C with area $A \in [1/2, 1]$ contained in a minimum bounding box of dimensions $(\mathrm{diam}(C) = 1/h = w) \times h$, a positive scalar ϕ, and a sequence γ_k such that $\gamma_k \cdot k$ is increasing.
Output: The locations of a finite set of points X in C that approximately minimize $F(X)\gamma_{|X|} \cdot |X| + \phi\,\mathrm{TSP}(X) + \mathrm{Dir}(X, C)$ within a factor of 3.93.
Let $\alpha{:=}H(A, \sqrt{3}, 1/\sqrt{3})$ and $K{:=}\max[\lceil \alpha/(2\phi)\rceil, \lceil 1/h^2\rceil]$;
Set $v^*\infty$;
for $k \in \{1, \ldots, k\}$ **do**
  Let $X = \text{ApproxFW}(C, K)$;
  if $F(X) \le v^I$ **then**
    Set $v^* := F(X)$ and $X^* := X$;
  end
end
return X^I.

Claim 3. *Suppose that $\square C$ is a box of dimensions $w \times h$, where $w \ge h$, and that $R_1, \ldots, R_k$ is the output of Algorithm 1. If $k \ge w/(3h)$, then we have $\mathrm{AR}(R_i) \le 3$ for all rectangles R_i. If $k < w/(3h)$, then $\mathrm{AR}(R_i) = w/(hk)$.*

Proof. This is straightforward and explained in §D of the online supplement. □

Using the preceding results, we can now present our approximation algorithm for placing facilities so as to minimize (1), which is given in Algorithm 3. Algorithm 3 takes as input a convex polygon C, a sequence $\{\gamma_k\}$, and a positive scalar ϕ.

In the following section, we will prove that this is a constant-factor approximation algorithm.

6. Proof of Approximation Bounds

In this section, we show that Algorithm 3 produces a constant-factor approximation for the problem of minimizing the function $f(X)$ defined by

$$f(X) = \phi\,\mathrm{TSP}(X) + \mathrm{Dir}(X, C),$$

which differs from the original objective function $F(X)$ in (1) in that we have assumed that $\psi = 1$ (without loss of generality) and that $\gamma_k = 0$ for all k. The proof for general sequences $\{\gamma_k\}$ follows the same line of reasoning as this section and is given in §E of the online supplement. We reiterate that we also assume that the input region C is contained in a bounding box of dimensions $(\mathrm{diam}(C) = 1/h = w) \times h$, which implies that $\mathrm{Area}(C) = A \in [1/2, 1]$. Throughout this section, we will use the terms w and $1/h$ interchangeably to simplify exposition. To begin, we define the function $\alpha(A)$ as

$$\alpha(A) = H(A, \sqrt{3}, 1/\sqrt{3}) \in (0.2943, 0.4753) \quad \text{for } A \in [1/2, 1]$$

and the function $\beta(A)$ as

$$\beta(A) = H(A, 1, 1) \in (0.2092, 0.3826) \quad \text{for } A \in [1/2, 1],$$

which we will abbreviate from now on as α and β, respectively, suppressing the dependence on A in the interests of brevity. It is not hard to verify that α and β are concave in A.

Note that Algorithm 3 requires that we iterate through various values of k from 1 through $K = \max\{\lfloor 1/h^2\rceil, \lfloor \alpha/(2\phi)\rceil\}$. In fact, it turns out that we can verify that the desired approximation ratio holds by only checking *three* strategic values of k, namely, $k = 1$, $k = \lfloor 1/h^2\rceil$, and $k = \lfloor \alpha/(2\phi)\rceil$, which are shown in Figure 17. Since we are assuming that $\psi = 1$ and $\gamma_k = 0$ in this section, we can condense the lower bounding functions Φ^1_{LB} and Φ^2_{LB} to

$$\Phi^1_{\mathrm{LB}}(A, \phi) = \begin{cases} A\sqrt{\phi} - \dfrac{3\phi\sqrt{\pi A}}{8} & \text{if } \phi \le \dfrac{16A}{9\pi}, \\[2ex] \dfrac{2A^{3/2}}{3\sqrt{\pi}} & \text{otherwise,} \end{cases}$$

and

$$\Phi^2_{\mathrm{LB}}(A, \phi, h) = \begin{cases} \dfrac{2A\phi - 4\phi^2}{h} & \text{if } \phi \le \dfrac{A}{4}, \\[2ex] \dfrac{A^2}{4h} & \text{otherwise,} \end{cases}$$

where we have used the fact that the assumption that $\gamma_k = 0$ implies that k should be ∞ in our lower bounds (since they are both decreasing in k). We subsequently define $\Phi_{\mathrm{LB}}(A, \phi, h) = \max\{\Phi^1_{\mathrm{LB}}(A, \phi), \Phi^2_{\mathrm{LB}}(A, \phi, h)\}$.

We can define upper bounding functions Φ^1_{UB}, Φ^2_{UB}, and Φ^3_{UB} as follows:

Carlsson and Jia: *Continuous Facility Location with Backbone Network Costs*
Transportation Science, *Articles in Advance*, pp. 1–19, © 2014 INFORMS

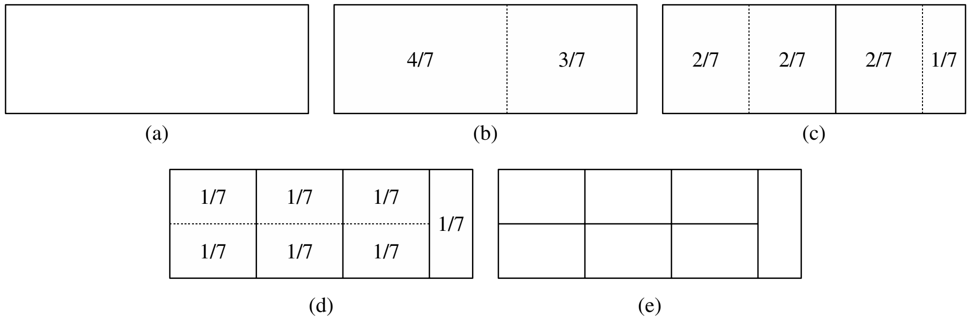

Figure 15 **The Input and Output of Algorithm 1**

Notes. We begin in (a) with a rectangle R and an integer $k = 7$; here we assume that $\text{Area}(R) = 1$. In panels (b)–(d), we subdivide R into smaller rectangles by a recursive subdivision; the areas of each subrectangle are shown. Panel (e) shows the output.

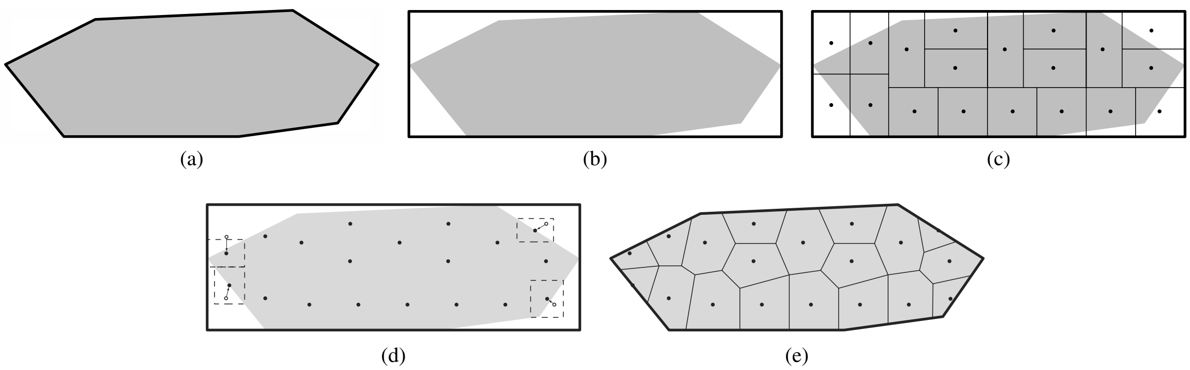

Figure 16 **The Input and Output of Algorithm 2**

Notes. We begin in (a) with a convex polygon C, whose axis-aligned bounding box $\square C$ is computed in (b). The bounding box is then partitioned into $k = 19$ equal area pieces in (c) using Algorithm 1. Some of the centers of these pieces are then relocated in (d), and (e) shows the output and Voronoi partition.

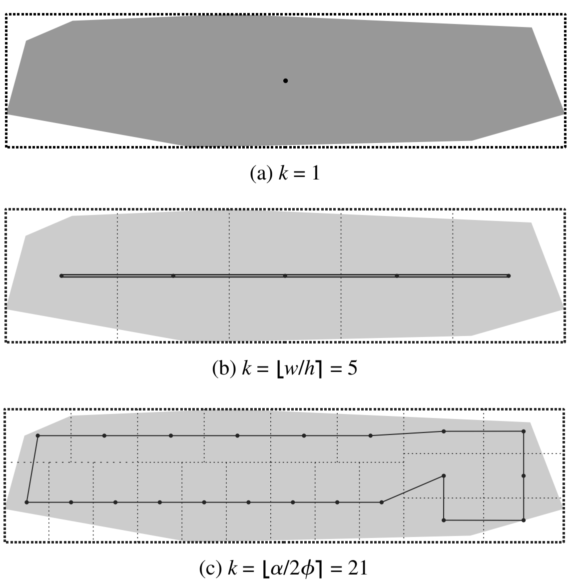

Figure 17 **The Three Candidate Outputs That Are Passed to Algorithm 2, with TSP Paths Shown**

Note. For visual clarity, the Voronoi diagrams of the point sets have been omitted, and instead we show the rectangular partition from Algorithm 1 that led to their placement.

- If we apply Algorithm 2 with only $k = 1$ point, then there is no backbone network, and therefore the objective function cost is at most $H(A, 1/h, h)$. We consequently define $\Phi^1_{\text{UB}}(A, \phi, h) = H(A, 1/h, h)$. As mentioned in Claim 7, we have

$$\Phi^1_{\text{UB}}(A, \phi, h) \sim \frac{2A}{3h} - \frac{1}{12h} - \frac{A^2}{3h}$$

as $h \to 0$.

- If we apply Algorithm 2 with $k = \lceil 1/h^2 \rceil = \lceil w/h \rceil$ points, then we have a backbone network with length $2 \cdot ((k-1)/(hk))$ and a collection of k equidistantly spaced points $X = \{x_1, \ldots, x_k\}$, each of which is contained in a rectangle R_i (produced by Algorithm 1) of dimensions $w/k \times h$. By convexity of $H(\cdot)$ in A, we know that the maximum value of $\text{Dir}(X, C)$ is therefore bounded above by $k \cdot H(A/k, w/k, h)$, so that we may define

$$\Phi^2_{\text{UB}}(A, \phi, h) = 2\left(\frac{k-1}{hk}\right)\phi + k \cdot H\left(\frac{A}{k}, \frac{1}{hk}, h\right)$$

with $k = \lceil 1/h^2 \rceil$. Note that as $h \to 0$, the aspect ratios of the rectangles R_i approach 1 (since $\text{AR}(R_i) = w/k/h = 1/h^2/\lceil 1/h^2 \rceil \to 1$), so that

$$\Phi^2_{\text{UB}}(A, \phi, h) \sim \frac{2\phi}{h} + \frac{\beta}{\sqrt{k}} \sim \frac{2\phi}{h} + \beta h$$

for small h by applying a simple scaling argument to the definition of β.

• If we apply Algorithm 2 with $k = \lfloor \alpha/(2\phi) \rceil$ and we also have $\lfloor \alpha/(2\phi) \rceil \geq 1/(3h^2)$, then by Claim 3, we know that all of the rectangles $R_1, \ldots, R_k$ produced by Algorithm 1 have an aspect ratio of at most 3. We therefore know that the maximum value of $\mathrm{Dir}(X, C)$ is bounded above by $k \cdot H(A/k, \sqrt{3/k}, 1/\sqrt{3k})$, which is equal to $\alpha/\sqrt{k}$ by applying a simple scaling argument to the definition of α. Using Theorem 3, we can therefore define

$$\Phi^3_{\mathrm{UB}}(A, \phi, h) = \begin{cases} \phi\left[\dfrac{hk}{m} + \dfrac{m}{h} + 2\,\dfrac{m-1}{m}\,h\right] + \alpha/\sqrt{k} \\ \qquad \text{if } \left\lfloor \dfrac{\alpha}{2\phi} \right\rceil \geq \dfrac{1}{3h^2}, \\ \infty \qquad \text{otherwise,} \end{cases}$$

with $k = \lfloor \alpha/(2\phi) \rceil$ and $m = 2 \cdot \max\{1, \lfloor \sqrt{\alpha} h/(2\sqrt{2\phi}) \rceil\}$. Note that as $\phi/h^2 \to 0$, we find that $\Phi^3_{\mathrm{UB}}(A, \phi, h) \sim 2\sqrt{2\alpha\phi}$. The choice of $k = \lfloor \alpha/(2\phi) \rceil$ rectangles to make our upper bound small follows the same spirit as that of the districting routing strategy of Daganzo and Newell (1986).

We subsequently define

$$\Phi_{\mathrm{UB}}(A, \phi, h) = \min\{\Phi^1_{\mathrm{UB}}(A, \phi, h), \Phi^2_{\mathrm{UB}}(A, \phi, h), \Phi^3_{\mathrm{UB}}(A, \phi, h)\}$$

and next show that $\Phi_{\mathrm{UB}}(A, \phi, h)/\Phi_{\mathrm{LB}}(A, \phi, h) \leq 3.93$ for all possible inputs.

6.1. Decomposing the Input Domain

It is clear that our proof will be complete if we can verify that $\Phi_{\mathrm{UB}}/\Phi_{\mathrm{LB}} \leq 3.93$ on the domain $(A, \phi, h) \in [1/2, 1] \times (0, \infty) \times (0, 1]$. To make this domain bounded, we first observe that if $\phi > A/4$ (which is obviously bounded below by 1/8), then $\Phi^1_{\mathrm{UB}}/\Phi^2_{\mathrm{LB}} \leq 3.4$. This is because neither bound depends on ϕ in this range, so we merely have to check the domain $(A, h) \in [1/2, 1] \times (0, 1]$. We can conclude that $\Phi^1_{\mathrm{UB}}/\Phi^2_{\mathrm{LB}} \leq 3.4$ on this domain by verifying the desired result computationally on the compact domain $(A, h) \in [1/2, 1] \times [\epsilon, 1]$ for small ϵ, then observing that as $h \to 0$, we have

$$\frac{\Phi^1_{\mathrm{UB}}}{\Phi^2_{\mathrm{LB}}} \sim \frac{2A/(3h) - 1/(12h) - A^2/(3h)}{A^2/4h} = \frac{8A - 4A^2 - 1}{3A^2} < 3 \quad \text{for } A \in [1/2, 1]$$

as desired. It will therefore suffice to verify that $\Phi_{\mathrm{UB}}/\Phi_{\mathrm{LB}} \leq 3.93$ on the *bounded* domain $(A, \phi, h) \in [1/2, 1] \times (0, A/4] \times (0, 1]$. We will prove this by decomposing the domain in question into five subdomains, as shown in Figure 18.

1. Subdomain I is compact and has strictly positive values of A, ϕ, and h. Thus, we can verify

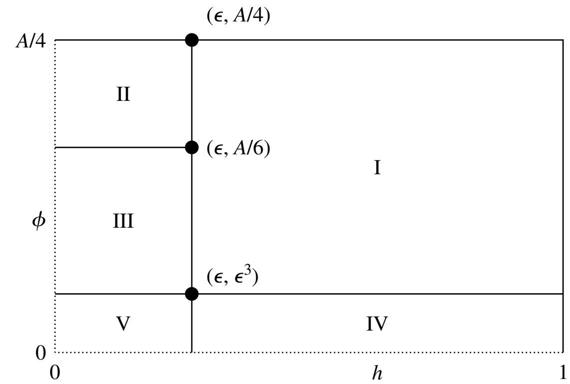

Figure 18 **The Bounded Domain $(\phi, h) \in (0, A/2] \times (0, 1]$, Where ϵ Is a Small Quantity (Say, 10^{-6}) (Not Drawn to Scale)**

Note. Here, we assume that $A \in [1/2, 1]$ is fixed, and we consider the approximation ratio as a function of ϕ and h; the dashed lines reflect the fact that the domain of interest is open on one end (i.e., we cannot have $h = 0$ or $\phi = 0$).

that $\Phi_{\mathrm{UB}}/\Phi_{\mathrm{LB}} \leq 3.93$ computationally using a branch-and-bound procedure; this is fairly straightforward because all upper and lower bounds are increasing in ϕ and A and decreasing in h. Figure 19 shows an actual surface plot of this ratio so that we can also visually confirm the bound.

2. On subdomain II, we have $\phi \in [A/6, A/4]$ and h is small, so that

$$\frac{\Phi^1_{\mathrm{UB}}}{\Phi^2_{\mathrm{LB}}} \sim \frac{2A/3 - 1/12 - A^2/3}{2A\phi - 4\phi^2} \leq \frac{2A/3 - 1/12 - A^2/3}{2A(A/6) - 4(A/6)^2} \leq 3 \quad \text{for } A \in [1/2, 1].$$

3. On subdomain III, we have ϕ bounded below by ϵ^3 and h is small, so that

$$\frac{\Phi^2_{\mathrm{UB}}}{\Phi^2_{\mathrm{LB}}} \sim \frac{2\phi/h + \beta h}{(2A\phi - 4\phi^2)/h} \sim \frac{2\phi/h}{(2A\phi - 4\phi^2)/h} = \frac{1}{A - 2\phi} \leq \frac{1}{A - 2(A/6)} \leq 3 \quad \text{for } A \in [1/2, 1].$$

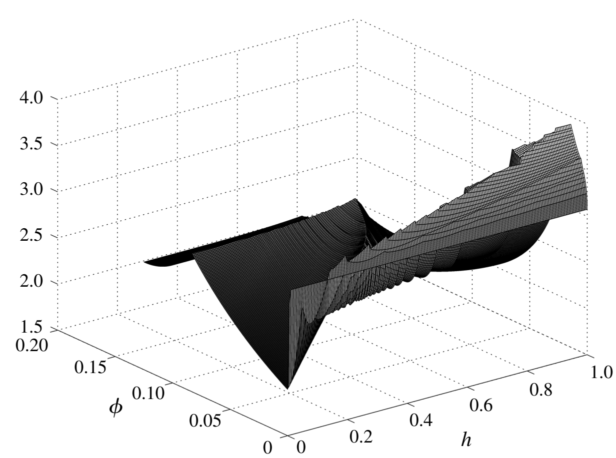

Figure 19 **The Ratio $\Phi_{\mathrm{UB}}/\Phi_{\mathrm{LB}}$ for $A = 1/2$ and $(\phi, h) \in (0, A/4] \times (0, 1]$**

4. On subdomain IV, we have $\phi \ll h^2$, so that (provided ϵ is small enough to ensure that $\alpha/(2\epsilon^3) > 1/3$, i.e., that Φ^3_{UB} is finite) the approximation ratio is

$$\frac{\Phi^3_{\mathrm{UB}}}{\Phi^1_{\mathrm{LB}}} \sim \frac{2\sqrt{2\alpha\phi}}{A\sqrt{\phi} - (3\phi\sqrt{\pi A})/8}$$
$$\sim \frac{2\sqrt{2\alpha}}{A} < 3.2 \quad \text{for } A \in [1/2, 1].$$

5. On subdomain V, the analysis is somewhat more involved because ϕ and h are both small but the ratio between them may be arbitrarily large. We consider the family of curves in subdomain V of the form $\{(\phi, h)\colon \phi = ch^2\}$ over varying c. If $c \geq 1/8$, then we find that

$$\frac{\Phi^2_{\mathrm{UB}}}{\Phi^2_{\mathrm{LB}}} \sim \frac{2\phi/h + \beta h}{(2A\phi - 4\phi^2)/h}$$
$$= \sim \frac{1}{A} + \frac{\beta}{2cA} \leq 3.7 \quad \text{for } A \in [1/2, 1].$$

If $c < 1/8$, then we find that

$$\frac{\Phi^3_{\mathrm{UB}}}{\Phi^1_{\mathrm{LB}}} = \frac{\phi[hk/m + m/h + 2((m-1)/m)h] + \alpha/\sqrt{k}}{A\sqrt{\phi} - (3\phi\sqrt{\pi A})/8}$$
$$\sim \frac{4(\alpha/m + 2cm + 4ch^2(1-1/m) + 2\sqrt{2\alpha c})}{(8A\sqrt{c} - 3ch\sqrt{A\pi})},$$

where $m = \lfloor\sqrt{\alpha/(2c)}\rceil$. As $c \to 0$ and thus $m \to \infty$, the above expression is approximately

$$\frac{4(\alpha/\sqrt{\alpha/(2c)} + 2c\sqrt{\alpha/(2c)} + 4ch^2[1 - 1/\sqrt{\alpha/(2c)}] + 2\sqrt{2\alpha c})}{(8A\sqrt{c} - 3ch\sqrt{A\pi})}$$
$$\sim \frac{2\sqrt{2\alpha}}{A} < 3.2 \quad \text{for } A \in [1/2, 1].$$

The nonlimiting case for c (e.g., $c \in [10^{-3}, 1/8]$) can be handled computationally.

This completes the proof that Algorithm 3 is a factor 3.93 approximation algorithm for minimizing the objective function of (1) for the special case where $\gamma_k = 0$ for all k. In Figure 19, we show a plot of $\Phi_{\mathrm{UB}}/\Phi_{\mathrm{LB}}$ for $A = 1/2$ and $(\phi, h) \in (0, 1/4] \times (0, 1]$.

Theorem 5. *Algorithm* 3 *is an approximation algorithm for minimizing the objective function of* (1), *with approximation constant* 3.93. *Its running time is* $\mathcal{O}(\omega n + \omega^2 \log n)$, *where* $\omega = \max\{1/h^2, 1/\phi\}$ *and* n *is the number of vertices of* C.

Proof. See §E of the online supplement for a generalization of the preceding proof for general sequences $\{\gamma_k\}$. □

7. Conclusions

We have considered the problem of designing an optimal facility location configuration in a convex planar region to minimize a weighted combination of the fixed costs, backbone network costs, and transportation costs in the region. We first showed that the two asymptotically optimal configurations that minimize these costs are the well-studied honeycomb heuristic and the Archimedes heuristic, which we introduced here. After analyzing several variations on our initial model, we then gave a fast constant-factor approximation algorithm for placing facilities in any convex polygonal region to minimize the costs described herein.

A natural direction for further research would be to investigate the optimal solutions to variations of our problem under different assumptions on the backbone network topology. The results in this paper describe the optimal solutions when the backbone network is a TSP tour or a minimum spanning or Steiner tree, although other possibilities remain, such as a *complete graph* or *star network* on the facilities. We expect the optimal solutions for these problems to have a distinctly different behavior, and we intend to study them in the near future.

Supplemental Material

Supplemental material to this paper is available at http://dx.doi.org/10.1287/trsc.2013.0511.

Acknowledgments

The authors gratefully acknowledge funding from the Defense Advanced Research Projects Agency [Young Faculty Award N66001-12-1-4218], National Science Foundation [Grant CMMI-1234585], and Office of Naval Research [Grant N000141210719]. The authors also thank Gérard Cachon for his initial formulation of this problem and Saif Benjaafar for introducing the topic, as well as Robert Eberhart, Z. Max Shen, Jiawei Zhang, the associate editor, and two anonymous referees for their helpful comments.

References

Arnold P, Peeters D, Thomas I (2004) Modelling a rail/road intermodal transportation system. *Transportation Res. Part E* 40(3): 255–270.

Beardwood J, Halton JH, Hammersley JM (1959) The shortest path through many points. *Math. Proc. Cambridge Philos. Soc.* 55(4): 299–327.

Bergeijk PAG, Brakman S (2010) *The Gravity Model in International Trade: Advances and Applications* (Cambridge University Press, Cambridge, UK).

Berman O, Jaillet P, Simchi-Levi D (1995) Location-routing problems with uncertainty. Drezner Z, ed. *Facility Location: A Survey of Applications and Methods* (Springer-Verlag, Berlin), 427–452.

Bertsimas DJ, Simchi-Levi D (1996) A new generation of vehicle routing research: Robust algorithms, addressing uncertainty. *Oper. Res.* 44(2):286–304.

Cachon GP (2014) Retail store density and the cost of greenhouse gas emissions. *Management Sci.* 60(8):1907–1925.

Campbell JF (1990) Designing logistics systems by analyzing transportation, inventory and terminal cost tradeoffs. *J. Bus. Logist.* 11(1):159–179.

Campbell JF (1993) One-to-many distribution with transshipments: An analytic model. *Transportation Sci.* 27(4):330–340.

Campbell JF, Nickerson JV (2011) Optimal arrangements for assembling a network in an emergency. *Hawaii Internat. Conf. System Sci.* (Institute of Electrical and Electronics Engineers, Washington, DC), 1–10.

Carlsson JG, Jia F, Li Y (2014) An approximation algorithm for the continuous k-medians problem in a convex polygon. *INFORMS J. Comput.* 26(2):280–289.

Daganzo CF (1984) The length of tours in zones of different shapes. *Transportation Res. Part B* 18(2):135–145.

Daganzo CF, Newell GF (1986) Configuration of physical distribution networks. *Networks* 16(2):113–132.

Drezner Z, Zemel E (1993) Competitive location in the plane. *Ann. Oper. Res.* 40(1):173–193.

Few L (1955) The shortest path and the shortest road through n points. *Mathematika* 2(2):141–144.

Geoffrion AM (1979) Making better use of optimization capability in distribution system planning. *AIIE Trans.* 11(2):96–108.

Grünbaum B, Shephard GC (2012) *Tilings and Patterns*, Dover Books on Mathematics Series (Dover Publications, Mineola, NY).

Haimovich M, Rinnooy Kan AHG (1985) Bounds and heuristics for capacitated routing problems. *Math. Oper. Res.* 10(4):527–542.

Hochbaum DS (1984) When are NP-hard location problems easy? *Ann. Oper. Res.* 1(3):201–214.

Langevin A, Mbaraga P, Campbell JF (1996) Continuous approximation models in freight distribution: An overview. *Transportation Res. Part B* 30(3):163–188.

Li S (2011) A 1.488 approximation algorithm for the uncapacitated facility location problem. *Proc. 38th Internat. Conf. Automata, Languages and Programming—Volume Part II, ICALP'11* (Springer-Verlag, Berlin), 77–88.

Magnanti TL, Wong RT (1984) Network design and transportation planning: Models and algorithms. *Transportation Sci.* 18(1):1–55.

Melkote S, Daskin MS (2001) An integrated model of facility location and transportation network design. *Transportation Res. Part A* 35(6):515–538.

Miranda PA, Garrido RA (2004) Incorporating inventory control decisions into a strategic distribution network design model with stochastic demand. *Transportation Res. Part E* 40(3): 183–207.

Nagy G, Salhi S (2007) Location-routing: Issues, models and methods. *Eur. J. Oper. Res.* 177(2):649–672.

Newell GF (1973) Scheduling, location, transportation, and continuum mechanics; some simple approximations to optimization problems. *SIAM J. Appl. Math.* 25(3):346–360.

Newell GF, Daganzo CF (1986) Design of multiple-vehicle delivery tours–I a ring-radial network. *Transportation Res. Part B* 20(5):345–363.

Pinedo M (2008) *Scheduling: Theory, Algorithms, and Systems* (Springer, New York).

Redmond C, Yukich JE (1994) Limit theorems and rates of convergence for Euclidean functionals. *Ann. Appl. Probab.* 4(4): 1057–1073.

Rodrigue JP, Comtois C, Slack B (2009) *The Geography of Transport Systems* (Routledge, New York).

Sheppard E (1978) Theoretical underpinnings of the gravity hypothesis. *Geographical Anal.* 10(4):386–402.

Steele JM (1981) Subadditive Euclidean functionals and nonlinear growth in geometric probability. *Ann. Probab.* 9(3):365–376.

Zemel E (1984) Probabilistic analysis of geometric location problems. *Ann. Oper. Res.* 1(3):215–238.