Dividing a territory among several facilities

John Gunnar Carlsson*

April 7, 2011

Abstract

We consider the problem of dividing a geographic region into sub-regions so as to minimize the maximum workload of a collection of facilities over that region. We assume that the cost of servicing a demand point is a monomial function of the distance to its assigned facility and that demand points follow a continuous probability density. We show that, when our objective is to minimize the maximum workload of all facilities, the optimal partition consists of a collection of circular arcs that are induced by a multiplicatively weighted Voronoi diagram. When we require that all sub-regions have equal area, the optimal partition consists of a collection of hyperbolic or quartic curves. We show that, for both problems, the dual variables correspond to "prices" for a facility to serve a demand point, and our objective is to determine a set of prices such that the entire region is "purchased" by the facilities, i.e. that the market clears. This allows us to solve the partitioning problem quickly without discretizing the service region.

1 Introduction

Given a collection of facilities, a natural question in many disciplines is to balance the load between these facilities while minimizing the cost of providing service. Specifically, suppose that $P = \{p_1, \ldots, p_n\}$ is a collection of fixed points representing facilities in a convex region C. We would like to partition C into n sub-regions so that all clients in sub-region R_i are satisfied by facility i, while balancing the workloads of the facilities. For example, if the points P represent fire stations, we might want to minimize the maximum workload that any of the fire stations experiences over a long time horizon. On the other hand, we might want to minimize the total workload experienced by all fire stations, while imposing the constraint that all fire stations service the same amount of customers in the long run. In this paper, we consider the case where the cost of service between a demand point x and a facility i is of the form $c(x, p_i) = \alpha_i \|x - p_i\|_2^k$ (hereafter simply $\alpha_i \|x - p_i\|^k$) and we assume that demand points follow a probability density function $f(x)$ on C. Thus, the average workload assigned to facility i is given by $\iint_{R_i} \alpha_i f(x) \|x - p_i\|^k \, dA$. We prove that the optimal boundaries between sub-regions must be circular arcs that are induced by a multiplicatively weighted Voronoi diagram. Furthermore, when we require that all sub-regions service the same average number of customers (so that $\iint_{R_i} f(x) \, dA = 1/n$), the optimal partition consists of a collection of hyperbolic arcs or quartic curves called Cartesian ovals. We suggest two heuristics for enforcing shape constraints, such as connectivity and restricting the maximum distance from a point to its assigned facility. Both problems can be solved quickly by solving a convex optimization problem with no more than $2n$ variables and without discretizing C. Although our result is a simple and immediate consequence of complementary slackness in linear programming, we are unaware of its existence elsewhere in the literature.

Related work

A well-studied related problem in operations research is the *Fermat-Weber problem*, in which our objective is to place a facility p (or collection of facilities) in C so as to minimize the average distance between points in C and p. Discrete and continuous versions of this problem are discussed at length in [6], and [7] gives the first polynomial-time algorithm for various versions of the 1-norm problem. The authors also prove that the 1-norm problem with multiple facilities is NP-hard for large n.

Two other variations commonly encountered in continuous facility placement are the n-center problem [14], in which the objective is to cover C with n identical circles with the smallest possible radius, and the minimum equitable radius problem [15], in which the objective is to place n facilities whose Voronoi cells have equal area

*Industrial and Systems Engineering, University of Minnesota. The author gratefully acknowledges the support of the Boeing Company. This research is supported in part by NSF Grant GOALI #0800151.

while minimizing the maximum distance from a point to its assigned facility. In most continuous facility placement problems, the partition of C is given by the Voronoi diagram of the facilities [12]. Thus, the main contribution of this paper is to show that not-insignificant savings can be made when the partition is also an optimization variable, and in fact that it can be optimized for a given set of facilities in a tractable way.

Considerably less work is published on the problem of partitioning C optimally when the depot points are fixed. One notion of "partitioning" discussed in [2] is to allow facilities to have variable "coverage radii" r_i, where the cost $\phi(r_i)$ is a monotonically increasing function; the problem is to find the optimal number, location, and coverage radii of a collection of facilities. Another paper [1] describes a constant-factor approximation algorithm for the problem of partitioning C so as to minimize the aggregate workload over all facilities while imposing an equal-area constraint (clearly, without the equal-area constraint, the solution to this problem is a Voronoi diagram for the facilities). The authors prove that the optimal solution consists of a collection of hyperbolic arcs. The authors also describe a constant-factor approximation algorithm for dividing C into equal-area convex pieces to maximize the minimum "fatness" of any piece. This in turn gives an approximation algorithm for the problem of minimizing the aggregate workload over all facilities when facility placement is variable, as well as the sub-region boundaries.

In [4, 5], the authors consider the problem of partitioning a convex region so as to minimize the maximum workload of a fleet of vehicles originating at depots P. They give an exact algorithm for partitioning a convex polygon into n convex pieces, with each piece containing one point p_i and all pieces having equal area. This algorithm is proven to be asymptotically optimal for the multi-depot vehicle routing problem when demand is uniformly distributed.

2 Optimal partitioning

Initially, we consider the case where demand is uniformly distributed in C, in which case our problem can be formulated as

$$\begin{aligned} \underset{R_1,\ldots,R_n}{\text{minimize}}\,\max_i \iint_{R_i} \|x-p_i\|\,dA & \qquad s.t. \\ \bigcup_{i=1}^{n} R_i &= C \end{aligned} \tag{1}$$

where $\|\cdot\|$ denotes the Euclidean norm. This admits an infinite-dimensional integer programming formulation

$$\begin{aligned} \underset{I_1(\cdot),\ldots,I_n(\cdot),t}{\text{minimize}}\ t & \qquad s.t. \\ t &\geq \iint_C I_i(x)\,\|x-p_i\|\,dA \quad \forall i \\ \sum_{i=1}^{n} I_i(x) &= 1 \quad \forall x\in C \\ I_i(x) &\in \{0,1\} \quad \forall i,x \end{aligned} \tag{2}$$

where

$$I_i(x) = \begin{cases} 1 & \text{if } x \text{ is assigned to facility } i \\ 0 & \text{otherwise} \end{cases}.$$

Relaxing the integrality constraint, we obtain an infinite-dimensional linear program over a Banach space. The dual of the relaxation is

$$\begin{aligned} \underset{\lambda,\sigma(\cdot)}{\text{maximize}} \iint_C \sigma(x)\,dA & \qquad s.t. \\ \sigma(x) &\leq \lambda_i\,\|x-p_i\| \quad \forall x\in C,\,\forall i \\ \sum_{i=1}^{n}\lambda_i &\leq 1 \\ \lambda_i &\geq 0 \quad \forall i\,. \end{aligned} \tag{3}$$

which is proven in the appendix (a proof sketch can easily be obtained by discretizing the problem). We consider the optimal dual variables λ^* and $\sigma^*(\cdot)$. It is clearly true that $\lambda_i^* > 0$ for all i, since otherwise the value of the

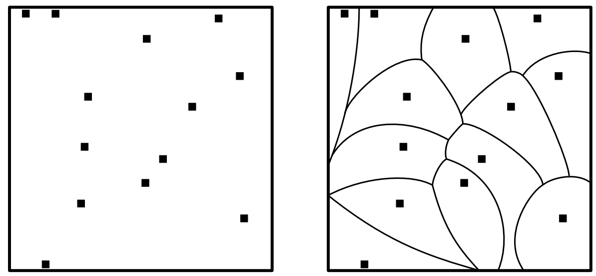


Figure 1: A load-balancing partition of the unit square with 13 facilities. Since the partitioning problem for fixed facility locations is convex, the above partition is globally optimal (within the margin of error of our ability to evaluate integrals). See section 4 for our procedure for finding such a partition.

dual program (3) is 0. It follows from complementary slackness that $t^* = \iint_C I_i^*(x)\|x - p_i\|\, dA$ for all i, i.e. that all facilities have the same workload at optimality. Next, we observe that, for any point $x \in C$, it must be the case that $\sigma^*(x) = \lambda_i^* \|x - p_i\|$ for some index i (since otherwise the solution is clearly sub-optimal). If that index i is unique, then clearly $I_i^*(x) = 1$ and $I_j^*(x) = 0$ for all other $j \neq i$, again by complementary slackness. Therefore, we find that the optimal boundaries between sub-regions are those points x for which $\lambda_i^* \|x - p_i\| = \lambda_j^* \|x - p_j\|$ for a pair of indices i and j. By the theorem of Apollonius [11], for $\lambda_i \neq \lambda_j$, these are simply arcs of a circle centered at $\frac{\lambda_i^2 p_i - \lambda_j^2 p_j}{\lambda_i^2 - \lambda_j^2}$ with radius $\frac{\lambda_i \lambda_j}{\lambda_i^2 - \lambda_j^2}\|p_i - p_j\|$ (a negative radius corresponds to p_i lying outside the circle). We define the *multiplicatively weighted Voronoi partition* of C and P, denoted $\mathcal{V}(C, P, \lambda) = \{V_1, \ldots, V_n\}$, by

$$V_i = \{x \in C | \lambda_i \|x - p_i\| \leq \lambda_j \|x - p_j\| \quad \forall j \neq i\} \ .$$

It turns out that $\mathcal{V}(C, P, \lambda^*)$ exactly describes an optimal partition.

Theorem 1. *Let λ^* be the weight vector obtained by solving the dual program (3). Then setting*

$$I_i^*(x) = \begin{cases} 1 & \text{if } x \in V_i^* \\ 0 & \text{otherwise} \end{cases},$$

with $\{V_1^*, \ldots, V_n^*\} = \mathcal{V}(C, P, \lambda^*)$, is an optimal partition of C with respect to the primal problem (9).

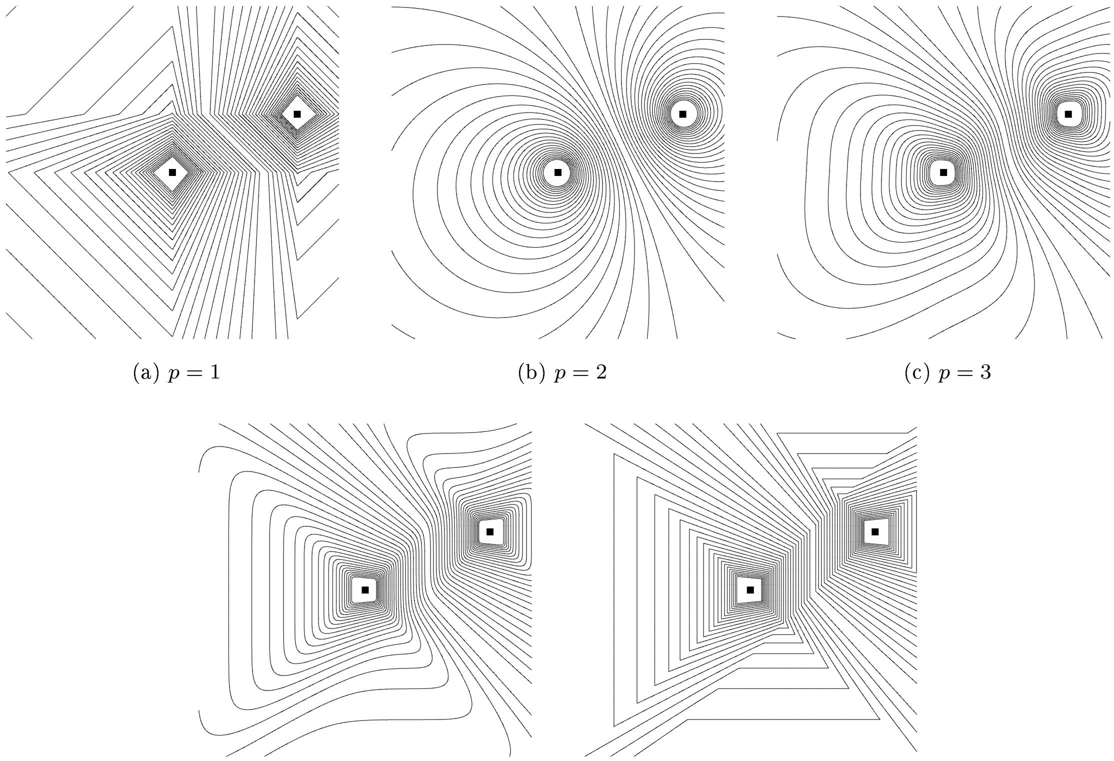
Proof. Observe that for each i we have $\iint_C I_i(x)\|x - p_i\|\, dA = \iint_C I_1(x)\|x - p_1\|\, dA$, since $\iint_C I_i^*(x)\|x - p_i\|\, dA$ is equal for all i. Plugging λ^* and the induced $\sigma^*(\cdot)$ into (**??**), we have

$$\begin{aligned}
\iint_C \sigma^*(x)\, dA &= \iint_C \min_i \lambda_i^* \|x - p_i\|\, dA \\
&= \sum_{i=1}^n \lambda_i^* \iint_{V_i^*} \|x - p_i\|\, dA \\
&= \sum_{i=1}^n \lambda_i^* \iint_{V_1^*} \|x - p_1\|\, dA \\
&= \iint_{V_1^*} \|x - p_1\|\, dA = \iint_C I_i^*(x)\|x - p_i\|\, dA
\end{aligned}$$

since $\sum_i \lambda_i^* = 1$, which completes the proof. □

See Figure 1 for an example of an optimal partition with $n = 13$ facilities. We describe our algorithm for quickly constructing these partitions in section 4.1.

Remark 2. Theorem 1 also applies to the case where the cost between a point x and its assigned depot p_i is a linear function of $\|x - p_i\|$ or any power of $\|x - p_i\|$, say $\alpha_i \|x - p_i\|^k$. In this case we find that the sub-regions of the optimal partition all have the same value of $\iint_{V_i^*} \alpha_i \|x - p_i\|^k\, dA$. In fact, the same result holds if demand is not uniformly distributed, in which case the expected cost for p_i to serve a region R is $\iint_R f(x)\|x - p_i\|\, dA$, where $f(x)$ is a probability density function. The optimal sub-region boundaries are still circular arcs because the boundaries still must satisfy $\lambda_i^* f(x)\|x - p_i\| = \lambda_j^* f(x)\|x - p_j\|$.



(a) $p = 1$ (b) $p = 2$ (c) $p = 3$

(d) $p = 10$ (e) $p = \infty$

Figure 2: Apollonian curves for various p-norms.

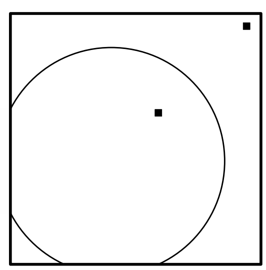


Figure 3: The partition shown above has a disconnected sub-region.

Remark 3. One interpretation of the dual variables λ_i is as follows: suppose that facilities each charge a certain *service rate*, i.e. a client at a point x must pay $\lambda_i \|x - p_i\|$ to use facility p_i. The dual problem asks us to choose rates to maximize our revenue, assuming that all clients use the cheapest facility, subject to a cap on the total rate at which we are permitted to charge them.

Remark 4. The theorem of Apollonius (that the points x for which $\lambda_i^* \|x - p_i\| = \lambda_j^* \|x - p_j\|$ are circular arcs) applies only to the Euclidean norm $\|\cdot\|_2$. Some Apollonian "curves" for other p-norms are shown in Figure 2. One might also consider this problem in a simply connected polygon, in which we define $\|x - p_i\|$ to be the geodesic distance from x to p_i. We find in this case that the optimal boundaries between sub-regions are quartic curves called *Cartesian ovals*, discussed in more detail in the next section.

Remark 5. One drawback to our formulation is that we have not imposed connectivity between regions; indeed, an optimal multiplicatively weighted Voronoi diagram need not be connected, as shown in Figure 3.

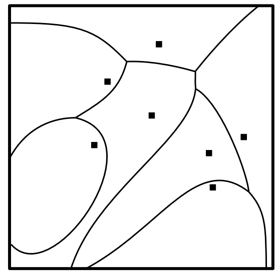


Figure 4: An equal-area load balancing partition of 7 points.

Partitioning with an equal-area constraint

One may also consider the scenario where we constrain each of our sub-regions to have equal area, in which case the integer program is

$$\begin{aligned} \underset{I_1(\cdot),\ldots,I_n(\cdot),t}{\text{minimize}}\; t \quad & s.t. \\ t &\geq \iint_C I_i(x)\,\|x-p_i\|\,dA \quad \forall i \\ \iint_C I_i(x)\,dA &\geq 1/n \quad \forall i \\ \sum_{i=1}^n I_i(x) &= 1 \quad \forall x \in C \\ I_i(x) &\in \{0,1\} \quad \forall i,x \end{aligned} \tag{4}$$

where we assume that $\text{Area}(C)=1$. The dual of the linear relaxation of this program is

$$\begin{aligned} \underset{\lambda,\gamma,\sigma(\cdot)}{\text{maximize}}\; \frac{1}{n}\sum_{i=1}^n \gamma_i + \iint_C \sigma(x)\,dA \quad & s.t. \\ \sigma(x) &\leq \lambda_i\|x-p_i\| - \gamma_i \quad \forall x \in C \\ \sum_{i=1}^n \lambda_i &\leq 1 \\ \lambda_i,\gamma_i &\geq 0 \quad \forall i\,. \end{aligned} \tag{5}$$

Again, we observe that for every x, the constraint $\sigma(x) \leq \lambda_i\|x-p_i\| - \gamma_i$ must be tight for some i, which means that we assign x to facility i. We thus find that our sub-region boundaries are curves of the form $X = \{x : \lambda_i\|x-p_i\| - \gamma_i = \lambda_j\|x-p_j\| - \gamma_j\}$. If $\lambda_i > 0$, then these boundaries are quartic curves called Cartesian ovals [8]. However, the degenerate case may occur where $\lambda_i = \lambda_j = 0$ (meaning that workloads may not be equal across all facilities) and $\gamma_i = \gamma_j$ for some pair i,j, and thus we cannot uniquely determine the assignment for those points x for which $\lambda_i\|x-p_i\| - \gamma_i = \lambda_j\|x-p_j\| - \gamma_j$ is minimal.

To address this situation, we observe that we must have at least one strictly positive entry (say λ_1) of λ. By complementary slackness, it is easy to show that we can obtain the optimal function $I_1^*(\cdot)$; first, we must have $I_1^*(x) = 1$ for all points x where $\lambda_1\|x-p_1\| - \gamma_1$ is strictly minimal and $I_1^*(x) = 0$ for all x where $\lambda_1\|x-p_1\| - \gamma_1$ is strictly non-minimal. Thus, the only points x where $I_1^*(x)$ is not defined are those where $\lambda_1\|x-p_1\| - \gamma_1 = \lambda_i\|x-p_i\| - \gamma_i$ for some i. If $\lambda_i > 0$ then this boundary is a Cartesian oval, and if $\lambda_i = 0$ then this boundary is a circle. Thus the set of all x where $I_1^*(x)$ is undefined is a finite one-dimensional set that has no impact on the value of the original problem (2), and we can set $I_1^*(x) = 0$ on these points without loss of generality.

By the preceding argument, we find an easy algorithm for finding the optimal functions $I_i^*(\cdot)$: solve the dual problem (5), and let $\mathcal{I}$ denote the set of all indices i where $\lambda_i > 0$. Then, construct $I_i^*(\cdot)$ for all $i \in \mathcal{I}$ and remove those regions from C. While $\{1,\ldots,n\} \setminus \mathcal{I}$ is non-empty, repeat the algorithm on the region $C \setminus \{x : I_i^*(x) = 1 \text{ for some } i \in \mathcal{I}\}$. An equal-area optimal partition is shown in Figure 4. Unlike the preceding problem, the case may arise where a sub-region does not contain its assigned facility, as shown in Figure 5.

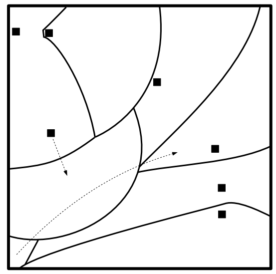


Figure 5: An equal-area load balancing partition of 7 points. Note that one sub-region does not contain its facility and one sub-region is not connected.

Remark 6. The preceding argument also holds for any arbitrary assignment of areas to sub-regions (not necessarily equal) or for soft constraints on the area of each region, in which our objective is to minimize $\max_i \iint_{R_i} \|x - p_i\| \, dA + \alpha_i \operatorname{Area}(R_i)$. Again, the result also holds if demand is not uniformly distributed, in which case we substitute the constraint that $\iint_C I_i(x) f(x) \, dA = 1/n$ for all i.

Minimizing the total workload

We may also change the objective function to $\iint_C \sum_{i=1}^n I_i(x) \|x - p_i\| \, dA$ while retaining the equal-area constraint (minimizing the aggregate workload over all facilities), whose formulation is given by

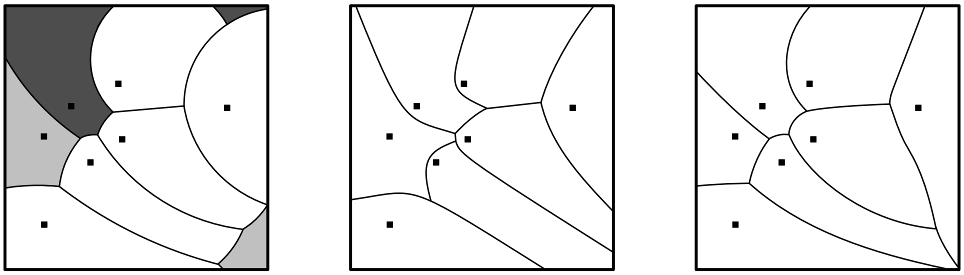
$$\begin{aligned}
\underset{I_1(\cdot),\ldots,I_n(\cdot)}{\text{minimize}} \iint_C \sum_{i=1}^n I_i(x) \|x - p_i\| \, dA & \qquad s.t. && (6)\\
\iint_C I_i(x) \, dA &\geq 1/n \quad \forall i\\
\sum_{i=1}^n I_i(x) &= 1 \quad \forall x \in C\\
I_i(x) &\in \{0,1\} \quad \forall i, x
\end{aligned}$$

which case we find the same result as the existence proof of [1] (that the optimal boundaries are *hyperbolic* arcs) using our complementary slackness argument. In addition, we are able to find these arcs efficiently using the dual program of the linear relaxation

$$\begin{aligned}
\underset{\gamma,\sigma(\cdot)}{\text{maximize}} \iint_C \sigma(x) \, dA & \qquad s.t. && (7)\\
\sigma(x) &\leq \|x - p_i\| - \gamma_i\\
\sum_i \gamma_i &= 1\,.
\end{aligned}$$

It turns out that the sub-regions are *star-convex*; that is, if point x is assigned to p_i, then so is every point x' on the segment connecting x and p_i. This is true for the same reason that an optimal (bipartite) Euclidean matching has no crossing edges. An interpretation of this dual program is as follows: suppose that clients are continuously distributed in C and suppose that each facility charges a "fee" γ_i (which may be positive or negative). The cheapest facility for a client located at a point x to use is the facility that minimizes $\|x - p_i\| - \gamma_i$. The optimal solution to (7) gives the "market-clearing" fee vector γ_i so that all facilities service the same number of clients.

Remark 7. Again, this argument applies to the non-uniform case. This result also holds for any assignment of mass to the facilities (not necessarily equal). An application of this situation is for carbon capture and sequestration, in which carbon emissions are taken from the facilities through pipelines and deposited in the surrounding forests, grasslands, and peat swamps. Here the density $f(\cdot)$ represents the amount of biomass available for accepting these emissions, which varies continuously over the terrain. Since carbon transportation is costly (estimated at $1-3 per ton of CO_2 across a pipeline), we clearly want to minimize the amount of transportation required, while still managing all emissions by the facilities.



(a) $\mu = 0$ (b) $\mu = 1$ (c) $\mu = 0.42$

Figure 6: Load-balancing partitions for $\mu = 0$ (disconnected Apollonian circles), $\mu = 1$ (connected hyperbolic arcs), and $\mu = 0.42$ (connected Cartesian ovals).

Remark 8. It is not hard to see that the primal problem (6) is a special "mixed" case of the Monge-Kantorovich transportation problem [16]: our objective is to "transport" the continuously distributed demand to the finite collection of facilities, while obeying capacity constraints and minimizing the aggregate transportation cost.

Enforcing shape properties

The algorithms given above do not always ensure that the sub-regions will be connected. We also have not put a bound on the maximum distance from a facility to a point assigned to it. In practice these are both clearly desirable properties and we show that they can be enforced using penalty functions or additional constraints.

Connectivity

We just observed that, when our objective is to minimize the *aggregate* workload over all facilities, the optimal sub-regions are connected, even when we impose an equal-area constraint. Therefore, we propose a "homotopy method" in which we minimize a weighted combination of the aggregate and maximum workloads:

$$\begin{aligned} \underset{I_1(\cdot),\ldots,I_n(\cdot),t}{\text{minimize}} \; (1-\mu)\, t + \mu \sum_{i=1}^{n} \iint_C I_i(x)\, \|x - p_i\|\, dA \qquad s.t. \\ t \;\; &\geq \;\; \iint_C I_i(x)\, \|x - p_i\| \quad \forall i \\ \iint_C I_i(x)\, \|x - p_i\| \;\; &\leq \;\; (1-\mu)\, A_0 + \mu/n \quad \forall i \\ \sum_{i=1}^{n} I_i(x) \;\; &= \;\; 1 \quad \forall x \in C \\ I_i(x) \;\; &\in \;\; \{0,1\} \quad \forall i, x\,. \end{aligned} \tag{8}$$

Here A_0 denotes the maximum area of any sub-region in the original problem (2) and we solve the above problem for $\mu \in [0,1]$ (we are guaranteed that regions will be connected at $\mu = 1$). Complementary slackness conditions imply again that the optimal sub-region boundaries are Cartesian ovals; an example is given in figure 6.

Diameter constraint

We may also impose a constraint on the maximum distance r between a point x and its assigned facility. The integer program in this case is

$$\begin{aligned} \underset{I_1(\cdot),\ldots,I_n(\cdot),t}{\text{minimize}} \; t \qquad & s.t. \\ t \;\; &\geq \;\; \iint_C I_i(x)\, \|x - p_i\|\, dA \quad \forall i \\ I_i(x) \;\; &= \;\; 0 \quad \forall i : \|x - p_i\| > r \\ \sum_{i=1}^{n} I_i(x) \;\; &= \;\; 1 \quad \forall x \in C \\ I_i(x) \;\; &\in \;\; \{0,1\} \quad \forall i, x\,. \end{aligned} \tag{9}$$

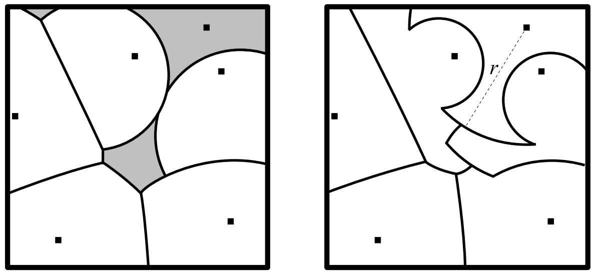


Figure 7: At left, a load-balancing partition without a distance constraint; the shaded region is disconnected. We can force our regions to be connected by imposing the distance constraint as shown at right.

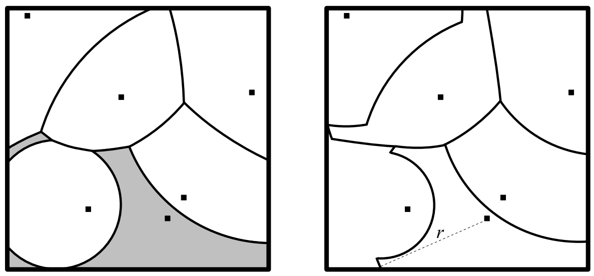


Figure 8: At left, an equal-area partition without a distance constraint; the shaded region is disconnected. Again we force our regions to be connected by imposing the distance constraint as shown at right.

The dual of the linear relaxation is

$$\underset{\lambda,\sigma(\cdot)}{\text{maximize}} \iint_C \sigma(x)\, dA \qquad s.t. \tag{10}$$

$$\begin{aligned} \sigma(x) &\leq \lambda_i \|x - p_i\|\, dA \quad \forall x \in C, \forall i : \|x - p_i\| \leq r \\ \sum_{i=1}^{n} \lambda_i &\leq 1 \\ \lambda_i &\geq 0 \quad \forall i\,. \end{aligned}$$

The optimal sub-region boundaries are a collection of circular arcs that come either from Apollonian circles or from the distance constraint. An example of such a partition is shown in Figure 7. When we impose an equal-area constraint, the optimal boundaries are either Cartesian ovals or the circles that arise from the distance constraint, as shown in Figure 8.

3 Simultaneous facility placement and partitioning

In the preceding section, we assumed that facility placement was fixed. A natural question is to consider the problem in which both the facility placement and the subdivision of territory is variable. When demand follows a uniform distribution, the n-medians algorithm described by [3] also finds an approximate (factor 5.02) solution to our min-max problem by first dividing the territory into rectangular cells, then placing the facilities. In this section we describe a local search procedure for simultaneously placing the facilities and dividing the territory. Unlike the partitioning problem in the preceding section, this problem is highly non-convex (with respect to the facility placement), and therefore the final configuration is in no way guaranteed to be globally optimal.

Let $P = \{p_1, \ldots, p_n\}$ denote the current placement of depot points and let $R = \{R_1, \ldots, R_n\}$ denote the optimal solution to (1) at P as computed via the dual program (3). We define the objective function $F(P)$ by

$$F(P) = \iint_C \sigma^*(x)\, dA = \sum_{i=1}^{n} \lambda_i^* \iint_{R_i} \|x - p_i\|\, dA$$

where $\sigma^*(x)$ and λ^* are defined via (3). Letting $p_i = (p_i^1, p_i^2)$ and $x = (x_1, x_2)$, we can approximate $\partial F / \partial p_i^k$ by

$$\begin{aligned} \frac{\partial F}{\partial p_i^k} &\approx \lambda_i^* \iint_{R_i} \left(\frac{\partial}{\partial p_i^k} \|x - p_i\| \right) dA \\ &= \lambda_i^* \iint_{R_i} \left(\frac{x_k - p_i^k}{\|x - p_i\|} \right) dA \,. \end{aligned} \tag{11}$$

The inaccuracies in this approximation arise due to the fact that the region R_i depends on p_i and λ_i^*. However, provided λ_i^* is sufficiently large, we find the above to be a useful and practical approximation. One approach for the uniform case would be to use the approximation algorithm of [3] as an initial guess, and then improve this guess with our iterative procedure; however, we believe that this is likely to converge to a local minimizer rather early. This is because the approximation algorithm divides C into rectangles, and consequently as n becomes large we expect the division to resemble a square grid (a local minimizer for our objective function). Based on the result of [13], we expect that the globally optimal configuration of P should be a "honeycomb", or hexagonal, tiling.

4 Computational experiments

In this section we report our results in a numerical experiment in which facility placement is variable. We let $C = [0,1]^2$ be the unit square and we assume that demand $f(\cdot)$ is uniformly distributed in C. We use a simple local search procedure to place facilities as described in the following section; the algorithm terminates when no further local improvement can be made.

4.1 Finding optimal partitions

A required sub-routine in placing facilities is to determine, for fixed facility locations P, the (globally) optimal solution to (1). As mentioned earlier we find it easier to solve (1) via the dual program (3) because it only depends on the n variables λ_i and because we can use numerical cubature to evaluate the objective function (as opposed to discretizing C). We performed all computations in Python and used the collapsed-square Gaussian cubature method [10] with tolerance 10^{-5} for all such evaluations. Given a configuration of facilities $P \subset C$, we find an initial guess of λ by using the Lagrange multipliers of a discretization of the linear relaxation of the primal problem (1) into a 25×25 grid. After obtaining an initial guess $\bar{\lambda}$, we then approximate the gradient (of the dual) with respect to λ using finite differencing to obtain a search direction. Since (3) is a convex problem, we use the golden section method to find the optimal λ along the gradient direction before choosing a new search direction. In practice we find that the solution converges after no more than 7 gradient evaluations are taken. We used this same procedure (possibly with more dual variables, depending on the problem and constraints) to generate the other figures in this paper. See Algorithm 1.

4.2 Local search for facility points

In our simulations we perform a gradient descent search on the points p_i. For any placement P, we approximate the gradient $\nabla F(P)$ with (11). Again, we evaluate the integrals with the collapsed-square Gaussian cubature method [10]. After determining the approximate search direction $-\nabla F(P)$, we choose the next iteration of P using a backtracking line search with parameter 0.9 (i.e. the search interval shrinks by a factor of 0.9 if the sufficient decrease condition is not met). See Algorithm 2.

4.3 Results

In our experiments, we initially place the points P uniformly at random in C. Our results are shown in Figure 9, where we compare the maximum workload and the maximum radius (distance from a facility to a point assigned to it) of our "Apollonian partitions" with the Voronoi diagrams corresponding to the best known optimal solutions to the n-center problem as reported in [14, 15]. Not surprisingly, we find that as n increases we have many values of λ_i that are the same, which causes the corresponding sub-region boundaries to be straight lines. Notably, this is not generally the case for points that are near the boundary of the square. A few examples of some locally optimal solutions are shown in Figure 10.

Input: A convex region C and a point set $P = \{p_1, \ldots, p_n\} \subset C$.

Output: A partition of C into n sub-regions R_i that minimizes $\max_i \iint_{R_i} \|x - p_i\| \, dA$ and the coefficients λ_i that define that partition.

Note: all integrals in this expression are evaluated using the collapsed-square Gaussian cubature method.

Discretize C into a collection of grid cells $\square_j$ and solve problem (13);
Let $\bar{\lambda}$ denote the Lagrange multipliers to the optimal solution of the discretization;
Using finite differencing, approximately construct a gradient $\bar{g} \in \mathbb{R}^n$ to the objective function of (13), written $H(\lambda)$, evaluated at $\bar{\lambda}$, restricted to the subspace $\{g \in \mathbb{R}^n | \sum_i g_i = 0\}$;
while $\|\bar{g}\| \geq 10^{-3}$ **do**
  Perform a golden-section search of $H(\lambda)$ on the line segment with direction $\bar{g}$ starting at $\bar{\lambda}$ and terminating at the boundary of the simplex $\{\lambda \in \mathbb{R}^n : \sum_i \lambda_i = 1, \lambda_i \geq 0\}$;
  Let $\bar{\lambda}$ denote the value of λ returned by the golden-section search;
  Using finite differencing, approximately construct a gradient $\bar{g} \in \mathbb{R}^n$ to the objective function of (13), written $H(\lambda)$, evaluated at $\bar{\lambda}$, restricted to the subspace $\{g \in \mathbb{R}^n | \sum_i g_i = 0\}$;
end
Set $R_i = \left\{x \in C \middle| \bar{\lambda}_i \|x - p_i\| \leq \bar{\lambda}_j \|x - p_j\| \quad \forall j \neq i\right\}$;
return $\{R_1, \ldots, R_n\}$ *and* $\bar{\lambda}$;

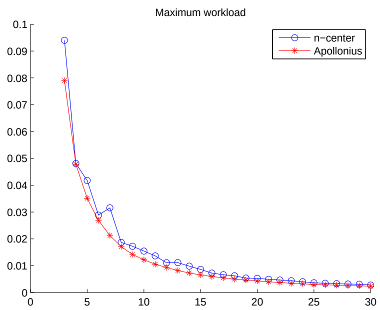
Algorithm 1: Algorithm FacilityPartition (C, P) takes a convex region C and a point set $P = \{p_1, \ldots, p_n\} \subset C$ as input and returns an optimal solution to problem (1).

Input: A convex region C and an integer n.

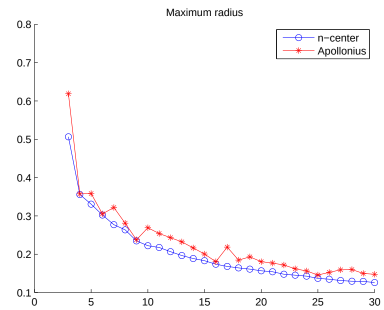
Output: A point set $P = \{p_1, \ldots, p_n\} \subset C$ and a load-balancing partition FacilityPartition (C, P) that is locally optimal with respect to P.

Let P be a set of n points distributed uniformly at random in C;
Let $\{R_1, \ldots, R_n\}$ and λ be the output of FacilityPartition (C, P);
Build an approximate gradient vector $g \in \mathbb{R}^{2n}$ of $F(P)$ by defining $\partial F / \partial p_i^k = \lambda_i \iint_{R_i} \left(\frac{x_k - p_i^k}{\|x - p_i\|} \right) dA$ for $i \in \{1, \ldots, n\}$ and $k \in \{1, 2\}$;
while $\|g\| \geq 10^{-3}$ **do**
  Perform a backtracking line search for the function $F(P)$, starting at P, in the direction $-g$ with parameter 0.9;
  Let P denote the best facility placement obtained from this line search;
  Let $\{R_1, \ldots, R_n\}$ and λ be the output of FacilityPartition (C, P);
  Build an approximate gradient vector $g \in \mathbb{R}^{2n}$ of $F(P)$ by defining $\partial F / \partial p_i^k = \lambda_i \iint_{R_i} \left(\frac{x_k - p_i^k}{\|x - p_i\|} \right) dA$ for $i \in \{1, \ldots, n\}$ and $k \in \{1, 2\}$;
end
return P *and* $\{R_1, \ldots, R_n\}$;

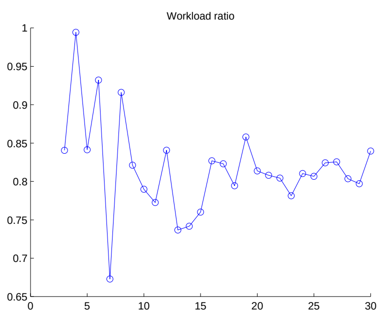
Algorithm 2: Algorithm FacilityLocation (C, n) takes a convex region C and an integer n as input and returns a point set $P = \{p_1, \ldots, p_n\} \subset C$ and a load-balancing partition FacilityPartition (C, P) that is locally optimal with respect to P.



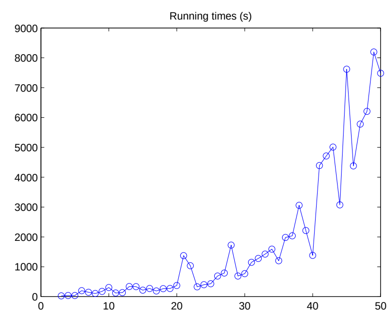
(a) The maximum workload $\max_i \iint_{R_i} \|x - p_i\| \, dA$ for $n \in \{3, \ldots, 30\}$, distributed in the unit square.



(b) The maximum radius $\max_i \max_{x \in R_i} \|x - p_i\|$ for $n \in \{3, \ldots, 30\}$, distributed in the unit square.

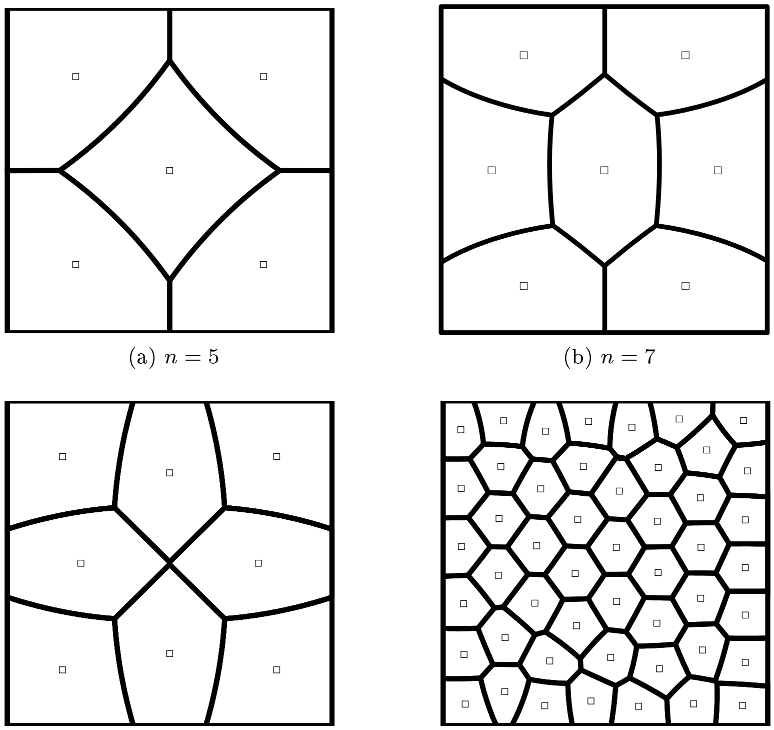


(c) The ratio between the maximum workloads of our method and the maximum workloads of the best known solutions to the n-center problem for $n \in \{3, \ldots, 30\}$, distributed in the unit square.



(d) The average running times, in seconds, for five trials each for $n \in \{3, \ldots, 50\}$, distributed in the unit square.

Figure 9: The maximum workloads and radii of our method compared to those of the best known solutions to the n-center problem.



(a) $n = 5$ (b) $n = 7$

(c) $n = 8$ (d) $n = 49$

Figure 10: The locally optimal facility placements and their associated partitions for select values of n.

5 Acknowledgments

The author thanks three anonymous referees, Zvi Drezner, Joseph Mitchell, Atsuo Suzuki, and Stan Wagon for their helpful suggestions.

References

[1] B. Aronov, P. Carmi, and M.J. Katz. Minimum-cost load-balancing partitions. *Algorithmica*, 54(3):318–336, July 2009.

[2] Oded Berman, Zvi Drezner, Dmitry Krass, and George O. Wesolowsky. The variable radius covering problem. *European Journal of Operational Research*, 196(2):516 – 525, 2009.

[3] J.G. Carlsson. An approximation algorithm for the continuous k-medians problem in a convex polygon, 2011. See http://www.tc.umn.edu/~jcarlsso/fermat-weber.pdf.

[4] J.G. Carlsson, B. Armbruster, and Y. Ye. Finding equitable convex partitions of points in a polygon efficiently. *ACM Transactions on Algorithms*, To appear, 2010.

[5] J.G. Carlsson, D. Ge, A. Subramaniam, and Y. Ye. Solving the min-max multi-depot vehicle routing problem. In *Proceedings of the FIELDS Workshop on Global Optimization*, 2007.

[6] Zvi Drezner. *Facility Location*. Springer, Berlin, 2001.

[7] Sándor P. Fekete, Joseph S. B. Mitchell, and Karin Beurer. On the continuous fermat-weber problem. *Operations Research*, 53(1):61–76, 2005.

[8] B. Lockwood. *A Book of Curves*. Cambridge University Press, Cambridge, 1961.

[9] D. G. Luenberger. *Optimization by Vector Space Methods*. John Wiley & Sons, Inc., 1st edition, 1997.

[10] J. N. Lyness and R. Cools. A survey of numerical cubature over triangles. In *Proceedings of Symposia in Applied Mathematics*, pages 127–150. American Mathematical Society, 1994.

[11] C. Ogilvy. *Excursions in Geometry.* Dover Publications, New York, 1990.

[12] A. Okabe and A. Suzuki. Locational optimization problems solved through Voronoi diagrams. *European Journal of Operational Research*, 98(3):445 – 456, 1997.

[13] C.H. Papadimitriou. Worst-case and probabilistic analysis of a geometric location problem. *SIAM Journal on Computing*, 10:542, 1981.

[14] A. Suzuki and Z. Drezner. The p-center location problem in an area. *Location Science*, 4(1-2):69 – 82, 1996.

[15] A. Suzuki and Z. Drezner. The minimum equitable radius location problem with continuous demand. *European Journal of Operational Research*, 195(1):17 – 30, 2009.

[16] C. Villani. *Topics in optimal transportation.* Graduate studies in mathematics. American Mathematical Society, 2003.

A Appendix

In this section we give a derivation of the dual problem (3) of the linear relaxation of problem (2), given by

$$\begin{aligned}\underset{I_1(\cdot),\dots,I_n(\cdot),t}{\text{minimize}}\ t \quad & s.t. \\ t &\geq \iint_C I_i(x)\,\|x-p_i\|\,dA \quad \forall i \\ \sum_{i=1}^n I_i(x) &= 1 \quad \forall x \in C \\ I_i(x) &\geq 0 \quad \forall i, x \in C\ .\end{aligned} \tag{12}$$

We can without loss of generality replace the equality constraint with the constraint $\sum_{i=1}^n I_i(x) \geq 1 \quad \forall x \in C$ (our reason for doing this will be made clear shortly). An easy proof "sketch" is to note that the discretization of (12) into grid cells $\square_j$ admits the LP

$$\begin{aligned}\underset{x,t}{\text{minimize}}\ t \quad & s.t. \\ t &\geq \epsilon \sum_j c_{ij} x_{ij} \quad \forall i \\ \sum_{i=1}^n x_{ij} &\geq 1 \quad \forall j \\ x_{ij} &\geq 0 \quad \forall i, j\end{aligned} \tag{13}$$

where ϵ is the area of grid cell $\square_j$, c_{ij} represents the Euclidean distance from p_i to the center of $\square_j$, and x_{ij} represents the fraction of grid cell j assigned to facility i. The dual of (13) is

$$\begin{aligned}\underset{q,r}{\text{maximize}} \sum_j q_j \quad & s.t. \\ q_j &\leq \epsilon c_{ij} r_i \quad \forall i, j \\ \sum_{i=1}^n r_i &\leq 1 \\ r_i, q_j &\geq 0 \quad \forall i, j\ .\end{aligned}$$

Finally, note that we can define the variable $q_j^{'} := q_j/\epsilon$ which gives

$$\begin{aligned}\underset{q^{'},r}{\text{maximize}}\ \epsilon \sum_j q_j^{'} \quad & s.t. \\ q_j^{'} &\leq c_{ij} r_i \quad \forall i, j \\ \sum_{i=1}^n r_i &\leq 1 \\ r_i, q_j &\geq 0 \quad \forall i, j\ ,\end{aligned}$$

which is precisely a discretization of (3).

A.1 Proof of optimality

For any Banach space $\mathfrak{X}$, let $\mathfrak{X}^\star$ denote its dual space. Let θ denote the zero vector for a Banach space (precisely *which* space will be clear from the context) and let $\langle \mathfrak{x}, \mathfrak{x}^\star \rangle$ denote the value of the functional $\mathfrak{x}^\star \in \mathfrak{X}^\star$ at the point $\mathfrak{x} \in \mathfrak{X}$. Theorem 1 of section 8.6 of [9] states the following:

Theorem 9. *(Lagrange Duality) Let $\mathfrak{f}$ be a real-valued convex functional defined on a convex subset Ω of a vector space $\mathfrak{X}$, and let $\mathfrak{G}$ be a convex mapping of $\mathfrak{X}$ into a normed space $\mathfrak{Z}$. Suppose there exists $\mathfrak{x}_1 \in \mathfrak{X}$ such that $\mathfrak{G}(\mathfrak{x}_1) < \theta$ and that $\mu_0 := \inf \{\mathfrak{f}(\mathfrak{x}) : \mathfrak{G}(\mathfrak{x}) \leq \theta, \mathfrak{x} \in \Omega\}$ is finite. Then*

$$\inf_{\mathfrak{x}\in\Omega,\, \mathfrak{G}(\mathfrak{x})\leq\theta} \mathfrak{f}(\mathfrak{x}) = \max_{\mathfrak{z}^*\geq\theta} \varphi(\mathfrak{z}^\star)$$

where

$$\varphi(\mathfrak{z}^\star) = \inf_{\mathfrak{x}\in\Omega} \mathfrak{f}(\mathfrak{x}) + \langle \mathfrak{G}(\mathfrak{x}), \mathfrak{z}^\star \rangle \ ,$$

and the maximum on the right is achieved by some $\mathfrak{z}_0^\star \geq \theta$.

We shall begin with the *dual* problem (3) and show that it is equivalent to the primal problem (2). The reason we perform things in this order is because the dual problem can be written as a finite-dimensional optimization problem over λ in a compact set, and thus an optimal solution λ^* must exist for (3). Theorem 9 then guarantees that an optimal solution exists for (2) as well, which will prove equivalence.

In problem (3), our variables consist of a nonnegative vector λ and a function $\sigma(\cdot)$, and consequently we have an infinite-dimensional optimization problem in the Banach space $\mathfrak{X} = \underbrace{\mathbb{R}\oplus\cdots\oplus\mathbb{R}}_{n} \oplus L_1$, where L_1 represents all functions $h(\cdot)$ defined on the convex region C such that $|h(x)|$ is Lebesgue integrable on C. Let Ω denote the positive orthant, i.e. $\lambda_i \geq 0$ and $\sigma(x) \geq 0$ (almost everywhere). Let $\mathfrak{f}(\mathfrak{x})$ be defined by

$$\mathfrak{f}(\mathfrak{x}) = \mathfrak{f}(\lambda_1, \ldots, \lambda_n, \sigma(\cdot)) = -\iint_C \sigma(x)\, dA \, .$$

We have the constraints that $\sigma(x) \leq \lambda_i \|x - p_i\|$ for all i and for all $x \in C$ and the constraint that $\sum_{i=1}^n \lambda_i \leq 1$ and therefore we define the map $\mathfrak{G} : \mathfrak{X} \to \mathfrak{Z} = \underbrace{L_1 \oplus \cdots \oplus L_1}_{n} \oplus \mathbb{R}$ by

$$\mathfrak{G} : \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \\ \sigma(\cdot) \end{pmatrix} \mapsto \begin{pmatrix} \xi_1(\cdot) \\ \vdots \\ \xi_n(\cdot) \\ \sum_{i=1}^n \lambda_i - 1 \end{pmatrix}$$

where $\xi_i(x) := \sigma(x) - \lambda_i \|x - p_i\|$. By the preceding existence argument we can replace the infimum operator in Theorem 9 with the minimum operator. Let $(I_1, \ldots, I_n, t) \in \underbrace{L_\infty \oplus \cdots \oplus L_\infty}_{n} \oplus \mathbb{R}$ denote an element of $\mathfrak{Z}^\star$. We have

$$\begin{aligned}
\min_{\mathfrak{x}\in\Omega,\, \mathfrak{G}(\mathfrak{x})\leq\theta} \mathfrak{f}(\mathfrak{x}) &= \max_{\mathfrak{z}^*\geq\theta} \varphi(\mathfrak{z}^\star) \\
\min_{\mathfrak{x}\in\Omega,\, \mathfrak{G}(\mathfrak{x})\leq\theta} -\iint_C \sigma(x)\, dA &= \max_{\mathfrak{z}^*\geq\theta} \left\{ \inf_{\lambda_i\geq0,\, \sigma(x)\geq0} -\iint_C \sigma(x)\, dA + \sum_{i=1}^n \iint_C \xi_i(x) I_i(x)\, dA + t\left(\sum_{i=1}^n \lambda_i - 1\right) \right\} \\
&= \max_{\mathfrak{z}^*\geq\theta} \left\{ \inf_{\lambda_i\geq0,\, \sigma(x)\geq0} \iint_C \left[\sum_{i=1}^n I_i(x)\sigma(x) + \lambda_i\left(t - \|x - p_i\| I_i(x)\right)\right] - \sigma(x) - t\, dA \right\} \\
&= \max_{\mathfrak{z}^*\geq\theta} \left\{ \inf_{\lambda_i\geq0,\, \sigma(x)\geq0} \iint_C \sigma(x)\left(\sum_{i=1}^n I_i(x) - 1\right) + \left[\sum_{i=1}^n \lambda_i\left(t - \|x - p_i\| I_i(x)\right)\right] - t\, dA \right\} \\
&= \max_{\mathfrak{z}^*\geq\theta} \left\{ \inf_{\lambda_i\geq0,\, \sigma(x)\geq0} \iint_C \sigma(x)\left(\sum_{i=1}^n I_i(x) - 1\right) dA + \sum_{i=1}^n \lambda_i\left(t - \iint_C \|x - p_i\| I_i(x)\, dA\right) - t \right\}
\end{aligned}$$

where we have used the fact that $\text{Area}(C) = 1$ and therefore $\iint_C t\, dA = t$, so that at optimality we know that $\sum_{i=1}^n I_i(x) \geq 1$ for all $x \in C$ and $t \geq \iint_C \|x - p_i\| I_i(x)\, dA$ for all i. Thus, the optimal solution to (3) is the same as the optimal solution to

$$\begin{aligned} \underset{I_1(\cdot),\ldots,I_n(\cdot),t}{\text{minimize}}\ t \quad & \quad s.t. \\ t \quad & \geq \quad \iint_C I_i(x)\,\|x - p_i\|\, dA \quad \forall i \\ \sum_{i=1}^n I_i(x) \quad & \geq \quad 1 \quad \forall x \in C \\ I_i(x) \quad & \geq \quad 0 \quad \forall i, x \end{aligned}$$

as desired.