10. GEOMETRIC LOCATION PROBLEMS

10.1. Statement of Problem

The purpose of this chapter is to develop the asymptotics of geometric location problems on a random sample. We use the general structure developed in the earlier chapters to obtain the asymptotics for the length of the graphs defined by geometric location problems.

Given n points in $\mathbb{R}^d$, $d \geq 2$, geometric location problems essentially involve choosing a subset of size k which "best represents the set". Such problems, also termed "k-median problems", are among the oldest in combinatorial optimization and have been studied by Fermat, Steiner, and Steinhaus (1956) among many others.

Let us now define our terms. Given a set $F := \{x_1, ..., x_n\}$ of points in $\mathbb{R}^d$ and $k \in \mathbb{N}$, choose k points in F as "medians" or "centers" and join the remaining vertices to the nearest center. The cost of serving a vertex equals the Euclidean distance to the nearest center. Letting C denote the collection of medians, the total cost of serving the points in F is thus

$$\sum_{i=1}^{n} \min_{x_j \in C} \|x_i - x_j\|.$$

The k-median problem involves choosing the set C of medians so as to minimize the above sum. The *minimal cost* obtained by optimally choosing C is given by the k median functional

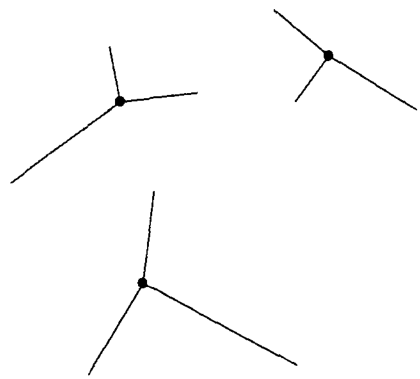
$$M(k; F) := \inf_{C \in \mathbb{C}} \sum_{i=1}^{n} \min_{x_j \in C} \|x_i - x_j\|, \tag{10.1}$$

where $\mathbb{C} := \mathbb{C}(k)$ denotes the collection of all subsets C of F of cardinality k. There is no restriction on the number of sites served by each center. See Figure 10.1.

When F consists of a three point set in $\mathbb{R}^2$, $k = 1$, and $\mathbb{C} := \mathbb{C}(1)$ ranges over all singletons which need not belong to F, then (10.1) reduces to a well-known problem of Fermat: "given three points in the plane, find a fourth point such that the sum of its distances to the three given points is a minimum". This problem has been widely popularized by Courant and Robbins (1941) and we refer to Kuhn (1974) for more on its history and applications.

The k-median problem is computationally difficult and Papadimitriou (1981) showed that it is NP-complete, resolving a conjecture of Fisher and Hochbaum (1980).

Figure 10.1. A k-median graph on 12 points with $k = 3$



In this chapter we describe the asymptotic behavior of a version of the k-median functional on random points in the unit cube. We claim that a modification of the k-median problem fits neatly into the theory of subadditive and superadditive Euclidean functionals and in this way we describe its stochastic behavior. This adds to the work of Hochbaum and Steele (1982), who made the first progress in the study of the stochastic analysis of the k-median problem.

Hochbaum and Steele (1982) showed that if $U_1, ..., U_n$ are i.i.d. random variables with the uniform distribution on $[0,1]^2$ then the functional $M(k; U_1, ..., U_n)$ behaves like a smooth subadditive Euclidean functional of order 1 when k grows linearly with n. More precisely they showed:

Theorem 10.1. *(Hochbaum and Steele, 1982) If $0 < \alpha < 1$ then*

$$\lim_{n\to\infty} M(\alpha n; U_1, ..., U_n)/n^{1/2} = C \quad a.s. \tag{10.2}$$

where $C := C(\alpha)$ is a finite non-zero constant.

We now consider a natural modification of the k-median problem. Let $|F| :=$ card$F$. Given $D \geq 2$, and F a finite set in $\mathbb{R}^d$, $d \geq 2$, consider the functional $M(\lceil |F|/(D+1) \rceil; F)$ under the additional assumption that no center serves more than D sites in F in addition to itself. Here $\lceil x \rceil$ denotes the smallest integer greater than or equal to x. Thus the degree of each center is at most D. Were we to consider $M(\lfloor |F|/(D+1) \rfloor; F)$ then there may not be enough centers of degree D to serve the points in F. Let $M(D; F)$ henceforth denote $M(\lceil |F|/(D+1) \rceil; F)$.

We claim that $M(D; F)$ is a subadditive Euclidean functional and has a canonical boundary functional $M_B(D; F)$ which is superadditive, smooth, and close in mean (3.15) to $M(D; F)$. For details we refer to McGivney and Yukich (1997). In this way we obtain rate results for $M(D; F)$ (Chapter 5), large deviations (Chapter 6), and the following asymptotics (Chapter 7):

Theorem 10.2. *(asymptotics for geometric location problems) Let X_i, $i \geq 1$, be i.i.d. random variables with values in $[0,1]^d$, $d \geq 2$. Then*

$$\lim_{n\to\infty} M(D; X_1, ..., X_n)/n^{(d-1)/d} = \alpha(M,d) \int_{[0,1]^d} f(x)^{(d-1)/d} dx \quad c.c.,$$

where $\alpha(M,d)$ is a positive constant and f denotes the density of the absolutely continuous part of the law of X_1.

It is convenient to view $M(D;F)$ as a functional over the rectangles $\mathcal{R}(d)$ in the way that has become natural in this monograph. We define for all $R \in \mathcal{R}(d)$

$$M(D; F, R) := M(D; F \cap R). \tag{10.3}$$

When R is the unit cube $[0,1]^d$ we simply write $M(D;F)$ instead of $M(D;F,[0,1]^d)$.

The proof of Theorem 10.2 depends heavily on introducing a *boundary functional* $M_B(D;F,R)$ associated with $M(D;F,R)$. Boundary functionals are a key conceptual and technical tool in the study of optimization problems and this is the case here as well. There is more than one way to define a boundary k-median functional. It turns out that the following definition insures superadditivity (3.3) and closeness in mean (3.15).

For all rectangles $R \in \mathcal{R} := \mathcal{R}(d)$ let R' denote the enlarged rectangle $\{x \in \mathbb{R}^d : d(x,R) \leq 1\}$, where $d(x,R)$ denotes the Euclidean distance between x and R. For all finite subsets S of the moat $R' - R$ consider the graph G which realizes $M(D; F \cup S, R')$ and let $G_R := G \cap R$ be its restriction to R. Denote the length of G_R by $M_R(D; F \cup S, R')$. Since the points in S are arbitrary, G_R may have overlapping edges and there may be as many as D copies of an edge joining a point in F to a point on ∂R. Let

$$M_B(D; F, R) := \inf_S M_R(D; F \cup S, R'), \tag{10.4}$$

where the infimum ranges over all finite subsets S of $R' - R$ having rational coordinates. This last requirement ensures the measurability of $M_B(D; X_1, ..., X_n, R)$ where X_i, $i \geq 1$, are random variables.

The bounded degree requirement on the centers implies that the subsets S of $R' - R$ have cardinality at most $D|F|$. The infimum (10.4) is thus realized by some set S. We call M_B the *boundary k-median functional.* Notice that $M_B \leq M$. As indicated, the boundary functional M_B is smooth of order 1, smooth, superadditive, and close in mean to the standard functional M; see McGivney and Yukich (1997b).

10.2. Additional Remarks

Rates of Convergence. It is natural to look for a rate of convergence in Theorem 10.2. To find a rate of convergence in the uniform case we apply the following lemma, essentially a modification of Theorem 5.1 when $p = 1$. This lemma is useful whenever the bound for

$$|EL(U_1, ..., U_n) - EL_B(U_1, ..., U_n)|$$

is different from $Cn^{(d-2)/d}$.

Lemma 10.3. *(rates of convergence of means) Let $U_1, ..., U_n$ be i.i.d. uniform random variables on $[0,1]^d$, $d \geq 3$. Suppose that L is a smooth subadditive Euclidean functional, L_B is a smooth superadditive Euclidean functional and*

$$|EL(U_1, ..., U_n) - EL_B(U_1, ..., U_n)| \leq \beta(n), \tag{10.5}$$

where $\beta(n)$ denotes a function of n. Then

$$|EL(U_1, ..., U_n) - \alpha(L, d)n^{(d-1)/d}| \leq \beta(n) \vee Cn^{(d-1)/2d}.$$

As shown in McGivney and Yukich (1997b), the functionals M and M_B satisfy (10.5) with $\beta(n) := Cn^{((d-1)/d)^2}$. We immediately obtain the rate result

$$|EM(D; U_1, ..., U_n) - \alpha(M, d)n^{(d-1)/d}| \leq Cn^{((d-1)/d)^2}.$$

It is not clear whether this rate estimate can be improved to give an error as small as $O(n^{(d-2)/d})$, which would be consistent with the rate results for the TSP and MST obtained in Chapter 5.

Generalizations. The methods above apply to at least one natural modification of the present problem. Consider the *Steiner* k-median problem which is a generalization of Fermat's problem and the bounded degree k-median problem. It is defined in the following way. Given points F in $[0,1]^d$, choose $k := \lceil \frac{|F|}{D+1} \rceil$ centers either from F or from Steiner points in $[0,1]^d$. Join each of the non-center points in F to its closest center under the restriction that no more than $D + 1$ points can be joined to any center. A component is said to be complete if the degree of the center is $D + 1$ and incomplete otherwise. If a center comes from F then its degree equals the number of vertices it serves, *including* itself. This will ensure that the number of incomplete components is at most D.

Define the *Steiner k-median functional* by

$$M^s(D; F, R) := \inf_S M(d; F \cup S, R),$$

where the infimum runs over all sets S of Steiner points in R. As in the definition of $M_B(D; F, R)$, observe that the infimum is realized by some set S.

Next we define the *boundary Steiner k-median functional* $M_B^s(D; F, R)$. For all rectangles $R \in \mathcal{R}$, let R' denote the enlarged rectangle $\{x \in \mathbb{R}^d : d(x, R) \leq 1\}$. For all finite sets S' of points belonging to the moat $R' - R$, consider the graph G which realizes $M^s(D; F \cup S', R')$ and let its restriction to R have length $M_R^s(D; F \cup S', R')$. Since the points in S are arbitrary, the restriction of G to R may produce as many as D copies of an edge which joins a point in F to ∂R. Let

$$M_B^s(D; F, R) := \inf_{S'} M(D; F \cup S, R),$$

where the infimum runs over all finite subsets S' of $R' - R$ having rational coordinates. We call M_B^s the *boundary Steiner k-median functional.* Notice that $M_B^s \leq M^s$.

Modifications of the proofs of simple subadditivity, subadditivity, and smoothness for $M(D; F, R)$ (see McGivney and Yukich (1997b)) show that $M^s(D; F, R)$ is a smooth subadditive Euclidean functional. We will sketch the proof of geometric subadditivity (3.4). For each $1 \leq i \leq 2$ consider the local k-median graphs G_i given by $M^s(D; F \cap R_i, R_i)$. There are at most D^2 *non-Steiner* vertices which are elements of incomplete components in each such graph. Therefore there are at most $2D^2$ *non-Steiner* vertices $\mathcal{V}$ which belong to incomplete components in $G_1 \cup G_2$. Construct a k-median graph $G_{\mathcal{V}}$ on $\mathcal{V}$. This graph together with the union of the complete components in $G_1 \cup G_2$ gives a feasible k-median graph. Since the length of $G_{\mathcal{V}}$ is bounded by a constant we have established subadditivity (3.4).

Likewise, minor changes in the proofs of superadditivity and smoothness for $M_B(D; F, R)$ show that $M_B^s(D; F, F)$ is also a smooth superadditive Euclidean functional. Since $M^s(D; F, R)$ and $M_B^s(D; F, R)$ are also close in mean, the umbrella Theorem 7.1 implies that

$$\lim_{n \to \infty} M^s(D; X_1, ..., X_n)/n^{(d-1)/d} = \alpha(M^s, d) \int_{[0,1]^d} f(x)^{(d-1)/d} dx \quad c.c.,$$

where $\alpha(M^s, d)$ is a positive constant and f denotes the density of the absolutely continuous part of the law of X_1. $\square$

Notes and References

1. There are several possibilities for extending the results of this chapter. Proving Theorem 10.2 without any assumption on the number of sites served by a center would be worthwhile. We do not address the case of power-weighted edges. Such an extension seems straightforward.

2. Most of this chapter is based on McGivney and Yukich (1997b).