8. APPLICATIONS AND EXAMPLES

We have seen that the complementary notions of subadditivity and superadditivity, together with isoperimetric methods, are well suited for proving limit theorems for the lengths of Euclidean graphs. We used these tools, together with a smoothness regularity condition, to obtain laws of large numbers, rates of convergence, and large deviations for the total edge lengths of graphs.

The previous chapters showed that our general methods describe the asymptotics of the lengths of the graphs of several of the outstanding problems of geometric probability. This includes the archetypical problems of combinatorial optimization, namely the length of the shortest tour on a random sample, the minimal length of a tree spanned by a random sample, and the length of a minimal Euclidean matching on a random sample.

The remainder of this monograph furnishes additional examples of graphs whose lengths may be studied by our general methods. We show that the general approach provides limit theorems for a wide variety of problems in combinatorial optimization, including geometric location problems on a random sample, the many traveling salesman problem on a random sample, and the semi-matching problem on a random sample.

We show that the general approach treats other classes of problems, especially the lengths of the graphs of some of the fundamental problems in computational geometry. This includes the length of the k nearest neighbors graph on a random sample. We will also show that the approach furnishes asymptotics for graphs occurring in minimal surfaces, particularly the length of the minimal triangulation of a random sample. Our examples are intended for purposes of illustration and they are not supposed to be exhaustive.

8.1. Steiner Minimal Spanning Trees

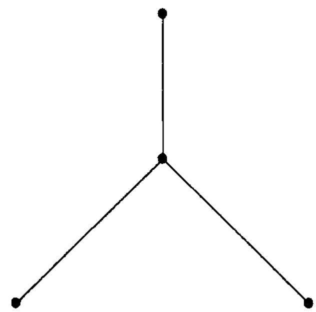
Let $F := \{x_1, x_2, ..., x_n\} \subset \mathbb{R}^d$, $d \geq 2$. A Steiner tree on F is a connected graph which contains F. The graph may include "Steiner points", that is vertices other than those in F. For example if F consists of the vertices of an equilateral triangle then the Steiner minimal spanning tree is obtained by joining all three vertices to the Steiner point lying in the center of the triangle. See Figure 8.1.

It is well-known that Steiner minimal spanning trees exist. This is a consequence of the easily proved fact that all vertices in Euclidean minimal spanning trees have bounded degree (see e.g. Melzak (1973)). More generally, the length of a Steiner minimal spanning tree on F with pth power-weighted edges is defined by

$$\hat{M}^p(F) := \min_S \sum_{e \in S} |e|^p,$$

where the minimum ranges over all Steiner trees S on F.

Figure 8.1. A Steiner minimal spanning tree (three vertices connected to a central Steiner point)



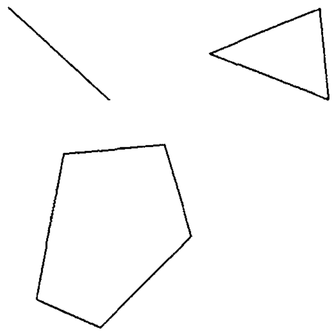
The Steiner MST functional is a smooth subadditive Euclidean functional. Indeed, if $\hat{M}^p(F,R)$ denotes the Steiner MST functional on pairs $(F,R) \in \mathcal{F} \times \mathcal{R}$, then for $F_1 \subset F_2$ we have $\hat{M}^p(F_1,R) \leq \hat{M}^p(F_2,R)$, which shows monotonicity. Using monotonicity and the simple subadditivity of $\hat{M}^p$ it is easy to verify that $\hat{M}^p$, $0 < p < d$, is a smooth subadditive Euclidean functional of order p.

The associated boundary Steiner MST functional $\hat{M}^p_B$, $1 \leq p < d$, is defined in the natural way and it is likewise easy to check that $\hat{M}^p_B$ is a superadditive Euclidean functional of order p. Checking pointwise closeness, closeness of means, and smoothness may be done exactly as in Chapter 3. The Steiner MST functional thus satisfies the conditions of the basic limit Theorem 4.1 as well as the umbrella Theorem 7.1.

8.2. Semi-Matchings

A "tour" or "cycle" is a connected graph in which all points have degree two. If we drop the requirement of connectedness then we have what is sometimes called a 2-matching or semi-matching. A minimal 2-matching on the vertex set $V := \{v_1, ..., v_n\}$ thus produces a graph on V of minimal total edge length in which all vertices have degree 2 with the understanding that an isolated edge between two vertices v_1 and v_2 actually represents two copies of the edge v_1v_2. The graph thus contains cycles with an odd number of edges ("odd cycles") as well as isolated edges. See Figure 8.2.

Figure 8.2. Semi-matchings contain isolated edges and odd cycles



Semi-matchings thus represent a relaxation of the TSP and can be described in the language of linear programming. Let $G = (V, E)$ be a graph such that for each edge $e \in E$ there is an associated weight w_e. We seek solutions to the following linear programming problem:

$$z = \min_x \sum_{e \in E} x_e w_e \tag{8.1}$$

subject to the constraints

$$\sum_{e \text{ meets } v} x_e = 1 \text{ for all } v \in V \text{ and } x_e \geq 0 \text{ for all } e \in E.$$

The Euclidean semi-matching problem seeks the solutions to (8.1) where the vertices V of G are points $v_1, ..., v_n$ in $\mathbb{R}^d$ and where the weight w_e associated with edge $e := v_i v_j$ is the Euclidean distance $\|v_i - v_j\|$. Balinski (1965) showed that the loading factor x_e can only be $0, 1/2$, or 1 in a minimal solution to (8.1). It follows that any minimal solution consists of a union of isolated edges with loading 1 and a collection of odd cycles that has all edge loading equal to $1/2$. See Figure 8.2.

We let $\hat{S}(F)$ denote the Euclidean semi-matching functional on the vertex set F; thus $\hat{S}$ equals the sum of the combined edge lengths of the minimal solution. Notice that the semi-matching functional is related to the TSP functional by $\hat{S}(F) \leq T(F)/2$. It is easy to verify that $\hat{S}$ is a subadditive functional (3.4) and in fact has no error term (that is C_1 is zero in (3.4)).

We now define the canonical boundary functional $\hat{S}_B$ associated with the semi-matching functional $\hat{S}$. This is done in a way which resembles the construction of the canonical boundary matching functional S_B. We define the boundary semi-matching functional by

$$\hat{S}_B(F, R) := \min\left(\hat{S}(F, R), \quad \inf \sum_i \hat{S}(F_i \cup \{a_i, b_i\})\right),$$

where the infimum ranges over all partitions $(F_i)_{i\geq 1}$ of F and all sequences of pairs of points $\{a_i, b_i\}$ lying on the boundary of $[0,1]^d$. The boundary is treated as a single vertex and edges belonging to the boundary are thus not counted when forming the odd cycles of the optimal semi-matching.

The boundary functional is superadditive (3.3). Indeed, the global semi-matching on points F in a rectangle $R := R_1 \cup R_2$ generates a graph in R_1 whose length exceeds that of the semi-matching on $F \cap R_1$. To see this, we modify those components of the global semi-matching which meet the boundary between R_1 and R_2. These components are odd cycles or isolated edges. If they are the latter, we will leave them alone. If an odd cycle meets the boundary, then the restriction to R_1 is either an odd or even cycle (the number of edges in the cycle is the number of non-boundary edges). If it is an odd cycle then we leave it alone. However, if the restriction is an even cycle, then we may delete every other edge in the cycle to obtain a set of isolated edges at no extra cost. In either event, the global semi-matching generates a feasible graph in R_1 which is at least as large as the optimal semi-matching of points in $F \cap R_1$. The same is of course true for rectangle R_2 and superadditivity follows.

To see that $\hat{S}$ is smooth of order 1 (3.8) we proceed as follows. Since $\hat{S}$ is subadditive we evidently have for all subsets F in $[0,1]^d$ the growth bound $\hat{S}(F) \leq C(\text{card}F)^{(d-1)/d}$ by Lemma 3.3. To show smoothness, first observe that for all sets F and G in $[0,1]^d$

$$\hat{S}(F \cup G) \leq \hat{S}(F) + \hat{S}(G) \leq \hat{S}(F) + C(\text{card}G)^{(d-1)/d}.$$

For the reverse inequality, let F_1 denote the set of points of F which are connected to points of G by the semi-matching which realizes $\hat{S}(F \cup G)$ and let $F_2 := F - F_1$. We have $\text{card}F_1 \leq 2\text{card}G$ and by considering the cycles in the optimal semi-matching of $F \cup G$ we can construct a feasible semi-matching of length $S'(F_2)$ on F_2 whose length is bounded by the length of the global optimal semi-matching:

$$\hat{S}(F) \leq S'(F_2) + \hat{S}(F_1) \leq \hat{S}(F \cup G) + C(\text{card}F_1)^{(d-1)/d}.$$

On combining the above inequalities we see that $\hat{S}$ is smooth of order 1:

$$|\hat{S}(F \cup G) - \hat{S}(F)| \leq C(\text{card}G)^{(d-1)/d}.$$

Similar arguments show that $\hat{S}_B$ is also smooth of order 1.

To see that $\hat{S}$ is pointwise close to the boundary functional $\hat{S}_B$ we need only follow the proof of Lemma 3.7 for the simple case $p = 1$. To see that $\hat{S}$ is close in mean (3.15) to $\hat{S}_B$, follow the proof of Lemma 3.10.

We thus see that the semi-matching functional $\hat{S}$ and the boundary functional $\hat{S}_B$ are respectively subadditive and superadditive Euclidean functionals of order 1 and therefore satisfy the asymptotics of the umbrella Theorem 7.1.

8.3. k Nearest Neighbors Graph

The theory of subadditive and superadditive Euclidean functionals applies to one of the fundamental constructions of computational geometry, namely the k nearest neighbors graph. Such a graph puts an edge between each point in a set F and its k nearest neighbors. The length of the graph, denoted $N(k;F)$, is the sum of the lengths of these edges. The k nearest neighbors graph receives considerable attention in computational geometry and is used in classification problems as well as in the construction of algorithms for solving geometrical problems; see Preparata and Shamos (1985). In this section we describe the asymptotic behavior of $N(k;F)$ for random sets F.

We briefly review the known results describing the behavior of $N(k;U_1,...,U_n)$, where U_i, $i \geq 1$, are i.i.d. uniform random variables on $[0,1]^d$. To simplify the notation we will write $N(k;n)$ for $N(k;U_1,...,U_n)$. Miles (1970) showed the following asymptotic result for $EN(k;n)$.

Theorem 8.1. *(Miles) The expected length of the k nearest neighbors graph $N(k;n)$ in dimension 2 satisfies*

$$\lim_{n\to\infty} \frac{EN(k;n)}{n^{1/2}} = C(k).$$

Later, Avram and Bertsimas (1993) used the technique of dependency graphs of Baldi and Rinott (1989) to show that $N(k;n)$ exhibits asymptotic normality. Their result adds to a similar result of Bickel and Breiman (1983), who show through long and complex arguments that $N(1;n)$ satisfies a central limit theorem (CLT). They were motivated to study the distribution of $N(1;n)$ in an attempt to derive the limiting distribution of a goodness of fit test for multidimensional densities based on nearest neighbor distances. Here $N(0,1)$ denotes the standard normal random variable.

Theorem 8.2. *(CLT for k nearest neighbors) In dimension 2 $N(k;n)$ satisfies*

$$\lim_{n\to\infty} \frac{N(k;n) - EN(k;n)}{(\mathrm{Var}N(k;n))^{1/2}} \stackrel{\mathrm{d}}{=} N(0,1).$$

This section will show that $N(k;n)$ is a smooth subadditive Euclidean functional of order 1 and in this way establish a law of large numbers which goes beyond Theorem 8.1. We are also able to treat the non-uniform case. Following McGivney (1997) we prove:

Theorem 8.3. *Let $X_1, ..., X_n$ be i.i.d. random variables with values in $[0,1]^d$, $d \geq 2$. Then*

$$\lim_{n\to\infty} N(k; X_1, ..., X_n)/n^{(d-1)/d} = \alpha(k,d) \int_{[0,1]^d} f(x)^{(d-1)/d} \quad c.c., \tag{8.2}$$

where f is the density of the absolutely continuous part of the law of X_1 and where $\alpha(k,d)$ is a constant depending only on k and d.

Proof of Theorem 8.3. We first show that $N(k;\cdot)$ is a smooth subadditive Euclidean functional. Given a rectangle R and a set F, we write $N(k; F, R)$ for $N(k, F \cap R)$ and in this way we turn $N(k,\cdot,\cdot)$ into a functional defined on the parameter set of rectangles. Here and throughout we adopt the convention that if $\text{card}F \leq k$, then the k nearest neighbors functional $N(k; F, R)$ allows points to have multiple matches to the same neighbor.

The k nearest neighbors functional is translation invariant and homogeneous of order 1. Moreover, we clearly have $N(k; F \cup G) \leq N(k; F) + N(k; G)$, that is $N(k;\cdot)$ is simply subadditive with no error term. It follows that the functional $N(k;\cdot,\cdot)$ satisfies geometric subadditivity (3.4) with no error term. Thus the nearest neighbors functional is a subadditive Euclidean functional of order 1 and thus satisfies the growth bounds of Lemma 3.3 with $p = 1$ there. We now show that it is smooth of order 1, that is we show

$$|N(k; F \cup G) - N(k; F)| \leq C(\text{card}G)^{(d-1)/d}. \tag{8.3}$$

To establish (8.3) we call upon a simple lemma which shows that the vertices in the graph of the k nearest neighbors functional have bounded degree. A related lemma appears in Bickel and Breiman (1983).

Lemma 8.4. *Vertices in the k nearest neighbors graph have bounded degree.*

Proof. Consider the k nearest neighbors graph on $\{x_i\}_{i=1}^n$. We will show that the degree of x_n is finite and independent of n.

It is well-known (see e.g. Bickel and Breiman (1983)) that $\mathbb{R}^d$ can be expressed as the union of $C(d)$ disjoint cones having x_n as a common vertex. Were the degree of x_n to exceed $k \cdot C(d)$, then the pigeonhole principle would imply that at least one cone contains $k+1$ points having x_n as a k nearest neighbor. Without loss of generality we may label these points $x_1, x_2, ..., x_{k+1}$ and we may assume that x_1 is farthest from x_n. However, by the definition of a cone, the distance between x_1 and the k points $x_2, ..., x_{k+1}$ is less than the distance between x_1 and x_n, showing that x_n is not one of the k nearest neighbors of x_1. Thus the degree of x_n is at most $k \cdot C(d)$. □

We now show smoothness (8.3). By simple subadditivity we have

$$N(k; F \cup G) \leq N(k; F) + N(k; G) \leq N(k; F) + C(\text{card}G)^{(d-1)/d},$$

where the last inequality follows from the growth bounds. To complete the proof of smoothness (8.3) we need therefore only show

$$N(k;F) \le N(k;F \cup G) + C(\mathrm{card} G)^{(d-1)/d}. \tag{8.4}$$

Given G, let F_G denote those points in F which have at least one of their k nearest neighbors in G. For any $H \subset F$, let $N'(k;H,F)$ denote the length of the k nearest neighbor graph on H with matching to points in F allowed. Note that $N'(k;H,F) \le N'(k;F,F) = N(k,F)$ and $N'(k;H,F) \le N(k;H)$. Thus

$$\begin{aligned} N(k;F) &= N(k;F,F) \\ &\le N'(k;F - F_G, F) + N'(k;F_G,F) \\ &\le N(k;F \cup G) + N(k;F_G). \end{aligned}$$

By Lemma 8.4, the cardinality of F_G is bounded by $C\mathrm{card}G$, where $C := C(k,d)$. Thus by the growth bounds for $N(k;\cdot)$ we obtain

$$N(k;F_G) \le C(\mathrm{card}F_G)^{(d-1)/d},$$

proving (8.4) and establishing that the k nearest neighbors functional is smooth of order 1.

Having shown that the k nearest neighbors functional is a smooth subadditive Euclidean functional of order 1, we now consider the canonical *boundary* k nearest neighbors functional. Given a rectangle R and a finite subset $F \subset R$, the boundary k nearest neighbors functional is the length of the graph connecting each vertex $V \in F$ to its k nearest neighbors, where now points on the boundary of R are potential neighbors and eligible for consideration. V may be joined to a boundary point P up to k times and in general V is matched to P exactly j times, $1 \le j \le k$, if there are exactly $k - j$ points in F whose distance to V is less than the distance between P and V. Let $N_B(k;F,R)$ designate the boundary k nearest neighbors functional on $\mathcal{F} \times \mathcal{R}$.

It is straightforward to check that $N_B(k;F,R)$ is a smooth Euclidean functional of order 1, simply subadditive (2.2), and superadditive (3.3).

We now show that the k nearest neighbors functional is pointwise close (3.10) to its respective boundary functional, that is we show

$$|N(k;F) - N_B(k;F)| = o\left((\mathrm{card}F)^{(d-1)/d}\right). \tag{8.5}$$

This will be achieved by showing the more refined estimate

$$N(k;F) \le N_B(k;F) + C\left((\mathrm{card}F)^{(d-2)/(d-1)} \vee \log(\mathrm{card}F)\right). \tag{8.6}$$

To show (8.6) we need the analog of Lemma 3.8 and a little notation.

Lemma 8.5. *Let F be a subset of $[0,1]^d$ of cardinality n and consider the boundary k nearest neighbors graph on F. The sum S of the lengths of the edges connecting points in F with $\partial[0,1]^d$ is bounded by $C(k,d)\left(n^{(d-2)/(d-1)} \vee \log n\right)$.*

Proof. Follow the proof of Lemma 3.8 *verbatim* with $p = 1$ and note that in each subcube Q of the partition $\mathcal{P}$ there are at most k points in $F \cap Q$ which are joined to the boundary. $\square$

To show (8.6) we proceed as follows. Let $F_1 \subset F$ denote the subset of points which are linked to the boundary of $[0,1]^d$ by the boundary k nearest neighbors graph G_B on F. Let $\mathcal{B}$ denote the points on the boundary which are connected to F_1 by G_B. We note that

$$N(k;F) \le N(k;F_1) + N'(k;F - F_1, F),$$

where the second term on the right side is bounded by $N_B(k;F)$. To prove (8.5) it will suffice to show that if $n := \mathrm{card}F$ then

$$N(k;F_1) \le C(n^{(d-2)/(d-1)} \vee \log n). \tag{8.7}$$

To show (8.7), consider the optimal tour on $\mathcal{B}$. Clearly $\mathrm{card}\mathcal{B} \le n$, and so

$$T(\mathcal{B}) \le Cn^{(d-2)/(d-1)},$$

since $\mathcal{B}$ lies in a set of dimension $d-1$. We use the graph T given by $T(\mathcal{B})$ to construct a k neighbors graph on F_1. Let V_i, $i \ge 1$, be an enumeration of the points in F_1. Without loss of generality we may assume that $\mathrm{card}F_1 \ge k+1$. We will assume that $N_B(k;F)$ connects V_i to $B_i \in \mathcal{B}$ for all choices of i. Relabeling if necessary we may without loss of generality let $B_1, B_2, ..., B_k, B_{k+1}, ...$ be the successive points in $\mathcal{B}$ visited by the tour T. Consider V_1. Choose the k neighbors of V_1 to be $V_2, ..., V_{k+1}$. The edges joining V_1 to these newly defined neighbors $V_2, ..., V_{k+1}$ come at a cost which is bounded by the sum of $\sum_{i=2}^{k+1} \|B_i - V_i\|$, $k\|B_1 - V_1\|$, and $\sum_{i=2}^{k+1} \|B_1 - B_i\|$. This latter sum is in turn bounded by a constant multiple $C(k)$ of the sum $\sum_{i=1}^{k} \|B_i - B_{i+1}\|$.

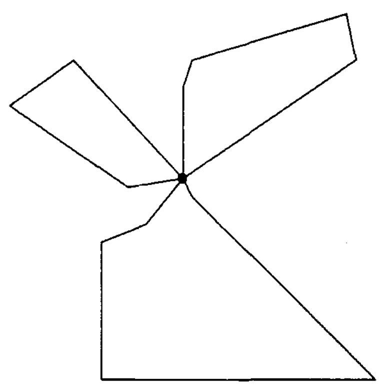
Now repeat the above construction for the remaining points in F_1. For each point V_i in F_1 we find thus k neighbors in F_1; they are not necessarily the nearest neighbors. The total cost of performing this operation is bounded by $C(k)(S + T(\mathcal{B}))$. The previously announced bounds on S and $T(\mathcal{B})$ now give the required result (8.7). This completes the proof of pointwise closeness and proves Theorem 8.3. $\square$

Remark. We may similarly establish that the k nearest neighbors functional is close in mean (3.16) to the boundary functional. Take $F := \{U_1, ..., U_n\}$ in the above analysis and show that the sum $S(n)$ of the lengths of the edges connecting points in F with the boundary satisfies $E(S(n)) \le Cn^{(d-2)/d}$. We also apply the bound $E\mathrm{card}\mathcal{B} \le Cn^{(d-1)/d}$, which may be established using the methods for proving Lemma 3.10.

8.4. The Many Traveling Salesman Problem

We consider a generalization of the TSP in which more than one salesman (and thus more than one tour) is allowed. In the k-traveling salesman problem on F (the k-TSP) we find a collection $\mathcal{C}$ of k subtours, each containing a distinguished vertex V, such that each point in F is in one subtour. $T(k;\mathcal{C},F)$ denotes the sum of the combined lengths of the k subtours in $\mathcal{C}$ and we define the k-TSP functional as the infimum $T(k;F) := \inf_{\mathcal{C}} T(k;\mathcal{C},F)$. This problem models the situation in which k salesmen work for a company with a home office at V and between them they need to visit each city once.

Figure 8.3. The many traveling salesman problem ($k = 3$)



A similar version of the k-TSP involves minimizing the maximum length subtour in the collection of subtours on F. This problem has received great attention but we will not consider it here. We refer to Christofides et al. (1979, Chapter 11) for a general discussion.

The k-TSP resembles the distribution problem in which k vehicles based at a central facility or depot are required to visit geographically dispersed customers in order to distribute or collect commodities. The collection and delivery of mail from mail boxes is one example of what is generally termed the vehicle routing problem (VRP). The k-TSP on F involves finding the shortest overall length of k subtours which originate at the depot and pass through all points of F. There is however no distribution or collection of commodities.

Observe that the k-TSP is not in general homogeneous nor does it satisfy translation invariance $T(F+y) = T(F)$ if $V \notin F$. These limitations are easily overcome if we view the k-TSP as a functional defined over d-dimensional rectangles, a point of view which has been the common thread throughout this monograph.

We now describe the k-TSP *functional* over the d-dimensional rectangles. This is a natural context for the k-TSP and brings out its intrinsic subadditivity. Given a rectangle R in $(\mathbb{R}^+)^d$, let $V := V(R)$ denote the vertex which is closest to the origin.

If $F \subset R$ is a finite set, then we let $T(k; F, R)$ denote the k-TSP tour length $T(k; F)$, where each of the k subtours must contain the distinguished vertex V. $T(k; F, R)$ is a functional indexed by the d-dimensional rectangles and is subadditive in the sense of (3.4), i.e.,

$$T(k; F, R) \le T(k; F \cap R_1, R_1) + T(k; F \cap R_2, R_2) + C\mathrm{diam}R.$$

To see this, consider the $2k$ subtours on R_1 and R_2. These subtours pass through the vertices $V(R_1)$ and $V(R_2)$. By the triangle inequality, the $2k$ tours on R_1 and R_2 may be tied together to generate k subtours on R at an extra cost of at most $Ck(\mathrm{diam}R)$. These k subtours each pass through $V(R)$ and generate a feasible solution of the k-TSP on R. Thus subadditivity follows.

We now verify that the k-TSP is smooth of order 1, that is

$$|T(k; F) - T(k; F \cup G)| \le C(\mathrm{card}G)^{(d-1)/d}. \tag{8.8}$$

As is the case with the standard TSP, the triangle inequality shows that $T(k; \cdot)$ is monotone so that

$$T(k; F) \le T(k; F \cup G).$$

Smoothness of order 1 follows once we establish

$$T(k; F \cup G) \le T(k; F) + C(\mathrm{card}G)^{(d-1)/d}. \tag{8.9}$$

However, the k-TSP is simply subadditive in the usual sense

$$T(k; F \cup G) \le T(k; F) + T(k; G) + Ck$$

since we may obtain a k tour graph on $F \cup G$ by tying together the k tour graph on F with the k tour graph on G at an additional cost of Ck. Since $T(k; G) \le C(\mathrm{card}G)^{(d-1)/d}$, we get (8.9) and thus smoothness (8.8).

The boundary k-TSP functional on rectangles R is defined in a natural way. Given $F \subset R$, the graph of a feasible boundary k-TSP tour consists of at most k subtours joined to the distinguished vertex $V(R)$. Moreover, subtours may exit to the boundary of R and subsequently re-enter. Travel along the boundary is free. The minimal total edge length of such a graph is called the boundary k-TSP functional and is denoted by $T_B(k; F, R)$; it is the natural analog of the canonical boundary TSP functional $T_B(F, R)$ studied in earlier chapters.

$T_B(k; F, R)$ is superadditive in the sense of (3.3), that is

$$T_B(k; F, R) \ge T_B(k; F \cap R_1, R_1) + T_B(k; F \cap R_2, R_2),$$

where $R := R_1 \cup R_2$ is the union of rectangles. To see this, observe that the restriction of the graph given by $T_B(k; F, R)$ to the subrectangle R_i, $1 \le i \le 2$, consists of a collection of paths with endpoints lying on the boundary of R_i. These paths may be linked with edges lying on the boundary of R_i, $1 \le i \le 2$, to form a collection of at most k closed loops, each of which passes through the distinguished vertex $V := V(R_i), 1 \le i \le 2$. Since the extra edges used in this operation all lie on the boundary of R_i, this operation comes at no additional cost. The resulting collection of at most k subtours can only exceed the minimal such collection, showing that the restriction of the graph given by $T_B(k; F, R)$ to R_i must exceed $T_B(k; F \cap R_i, R_i)$, $1 \le i \le 2$. This establishes superadditivity.

We may prove that $T_B(k; F, R)$ is smooth of order 1 by following the smoothness proof for $T(k; F, R)$.

Finally, the k-TSP functional is both pointwise close (3.10) and close in mean (3.15) to the canonical boundary k-TSP functional. This is seen by following the proofs of Lemmas 3.7 and 3.10; no new ideas are needed.

We have thus established that the k-TSP problem is a subadditive Euclidean functional which is smooth of order 1 and that the canonical k-TSP boundary functional is superadditive and smooth of order 1. Since they are pointwise close and close in mean they satisfy the basic limit Theorem 4.1, the rate results of Chapter 5, and the umbrella Theorem 7.1.

8.5. The Greedy Matching Heuristic

The theory of subadditive and superadditive Euclidean functionals evidently covers a wide range of problems in geometric probability. In the sequel we will see that the triangulation problem and the k-median problem of Steinhaus also fit neatly into the present theory. Yet the theory doesn't cover all problems in combinatorial optimization and operations research. Indeed, some heuristic solutions unfortunately do not seem to fit neatly into the theory. In this section we will see that the greedy heuristic for minimal matching does not quite fit into our theory.

The greedy heuristic provides an algorithm for obtaining a Euclidean matching which closely approximates the optimal matching. The greedy heuristic $G(F)$ on a point set F successively matches the closest unmatched pairs of points. Much work has been done to show that the heuristic G is a good approximation to the minimal matching S. Reingold and Tarjan (1981) have shown that for point sets $F \in \mathbb{R}^2$ of cardinality n, one has the bound

$$\frac{G(F)}{S(F)} \le \frac{4}{3} n^{\log_2 1.5}.$$

The greedy heuristic G is clearly translation invariant and homogeneous. Less obvious is the fact that G is smooth of order 1, a fact proved by Avis, Davis, and Steele (1988). The proof of smoothness depends in part on a growth estimate $G(F) \le C(\text{card}F)^{(d-1)/d}$. Thus the heuristic G is a smooth Euclidean functional of order 1. However, examples show that G is apparently not subadditive in the usual sense (3.4). The canonically defined boundary functional G_B is apparently not superadditive either.

Despite the fact that the greedy heuristic and its canonically defined boundary functional are not subadditive and superadditive, respectively, Avis, Davis, and

Steele (1988) showed that the heuristic can be approximately localized by the sum of its values on the m^d subcubes of $[0,1]^d$. In this way they showed that the greedy heuristic exhibits the asymptotic behavior (7.2) given by the umbrella Theorem 7.1.

This raises the following question: can the theory of smooth subadditive Euclidean functionals be appropriately modified and enlarged to accommodate functionals such as the greedy matching heuristic which are Euclidean and smooth but not subadditive in the sense of (3.4)? We believe that this question can be resolved positively. Ideally we would like to enlarge the theory so that it also encompasses other heuristics such as the greedy heuristic for the TSP. It is not clear that this can be done, however.

8.6. The Directed TSP

Consider the random directed graph G_n whose vertices are independent and uniformly distributed random variables $U_1, ..., U_n$ on the unit square. For $1 \leq i < j \leq n$, the orientation of the edge $X_i X_j$ is selected at random, independently for each edge and independently of the U_i, $i \geq 1$. Thus the edges in G_n are given a direction just by flipping fair coins. The directed TSP involves finding the shortest directed path through the random vertex set. While it is clear that there may not be a directed cycle through the vertex set, a classic result of Rédei (1934) nonetheless guarantees that a path exists. We write $D(n) := D(U_1, ..., U_n)$ for the length of the directed TSP path through the sample $U_1, ..., U_n$.

While the directed TSP is Euclidean and subadditive, it is not clear that it is smooth. Nonetheless, Steele (1986) has shown that $D(n)$ behaves as though it were a subadditive Euclidean functional which is smooth of order 1:

$$\lim_{n \to \infty} ED(n)/n^{1/2} = \alpha.$$

Steele's (1986) methods do not yield the complete convergence of the directed TSP and it was left to Talagrand (1991) to make this improvement. Using martingale difference sequences and an inequality related to Azuma's inequality, Talagrand (1991) actually shows the following stronger result, which implies the complete convergence of the directed TSP:

$$\sum_{n=1}^{\infty} P\{|D(n) - ED(n)| > C \log n\} < \infty.$$

As with any application of Azuma's inequality, the hard part is to obtain sharp bounds on the martingale difference sequence associated to $D(n)$. Talagrand (1991) handles this by showing that for t in the range $C(\log n/n)^{1/2} \leq t \leq 1$ we have

$$P\{|D(n) - D(n+1)| \geq t\} \leq C\exp(-t^2 n/C).$$

It is unresolved whether the directed TSP satisfies the asymptotics (7.2) of the umbrella Theorem 7.1.

Notes and References

1. In addition to the k nearest neighbors functional, there are other problems of computational geometry which have some or all of the properties of smooth subadditive Euclidean functionals. We mention the total edge length of the Voronoi tessellation, Gabriel graph, and Delaunay triangulation. Concerning the first two, Talagrand (1995, Chapter 11) shows that these constructions satisfy a smoothness condition.

Let $V(X_1, ..., X_n)$ denote the total edge length of the Voronoi tessellation on random variables $X_1, ..., X_n$. If $X_1, X_2, ...$ are independent and have a common density $f(x)$ on the unit square which satisfies $0 < \alpha \leq f(x) \leq \gamma$ for constants α and γ then McGivney and Yukich (1997a) show that $\lim_{n\to\infty} V(X_1, ..., X_n)/n^{1/2} = 2\int_{[0,1]^2} f(x)^{1/2} dx$ *c.c.* This extends upon Miles (1970).

2. We expect that the work of Jaillet (1993c) will also fit into the theory of subadditive Euclidean functionals. Jaillet considers a probabilistic version of the TSP in the following sense: find the shortest tour T through a point set V of cardinality n. For any given instance of the problem, only a random subset V' of points from V has to be visited. The subset V' is visited in the *same order* as they appear in the tour T; this gives a tour T' through V'. The length of the tour T' is a Euclidean functional and it seems likely that it may satisfy the umbrella Theorem 7.1.

3. We also expect that the degree-K minimum spanning tree (MST) problem will also fit into our theory. The degree-K MST problem asks for the minimum length spanning tree that has no vertex of degree greater than K. Such problems have been studied by Papadimitriou and Vazirani (1984) among others.

4. There are several well-known heuristics for the TSP and in his survey Steele (1990b) asks whether they satisfy the asymptotics (7.2). Goemans and Bertsimas (1991) show that the Held-Karp (1970,1971) heuristic enjoys the limit result (7.2). Although they do not use methods involving boundary functionals it seems likely that their results could be handled (and therefore extended) by the approach discussed in this monograph. There are other heuristics for the TSP and we refer to Lawler et al. (1985), Rosenkrantz et al. (1977), and Papadimitriou and Steiglitz (1982). Some of these heuristics may satisfy the umbrella Theorem 7.1. In particular, it is not yet clear whether the 2-OPT heuristic and the Christofides algorithm for the TSP satisfy the asymptotics (7.2).

5. Semi-matchings. One could easily define a power-weighted edge version of $\hat{S}$ but we do not consider this generalization here. We note that the solution to the semi-matching problem runs in polynomial time.

6. For additional limit results for the k nearest neighbors graph with power-weighted edges we refer to McElroy (1997) and McGivney (1997).