Probabilistic Analysis of Euclidean Multi Depot Vehicle Routing and related Problems

Dissertation

zur Erlangung des akademischen Grades

Doktor der Naturwissenschaften

(Dr. rer. nat.)

der Technischen Fakultät

der Christian-Albrechts-Universität zu Kiel

vorgelegt von

Dipl. Math. Sören Werth

Kiel

2006

1. Gutachter: Prof. Dr. Anand Srivastav

2. Gutachter: Prof. Dr. Klaus Jansen

Datum der mündlichen Prüfung: 20.12.2006

Acknowledgments

I owe sincere thanks to Prof. Dr. Anand Srivastav for introducing me to the probabilistic analysis of Euclidean optimization problems and supervising this thesis. I would like to thank all the people I had the opportunity to work with [BDS+06, DHW06, CDT+06], especially Andreas Baltz.

I am thankful to the Deutsche Forschungsgemeinschaft (DFG) for granting me a scholarship of the Graduiertenkolleg 357 "Effiziente Algorithmen und Multiskalenmethoden". Last but not least I would like to thank Prof. Dr. Peter Damaschke for the research stay at Chalmers University in Göteborg.

Contents

Chapter 1

Introduction

Euclidean optimization problems are naturally associated to graphs: we are given one or more sets of points in $[0,1]^d$ and we are looking for a certain graph on the points so that the total edge length with respect to the Euclidean distance is minimal. The most famous example is the Traveling Salesman Problem (TSP), where n points in $[0,1]^d$ are given and we are searching for the shortest cycle that contains each point exactly once. The total edge lengths of those graphs representing the solutions to Euclidean optimization problems of a "typical" instance are studied. From the mathematical point of view, in a "typical" instance the points are given by random variables.

The central theme of this thesis is the investigation of the asymptotic behavior of the total edge length of typical instances if the number of points tends to infinity. We consider the following Euclidean optimization problems: the multi depot vehicle routing problem (MDVRP), the all nearest neighbor problem (ANNP) and the b-degree constrained minimal spanning tree problem (bMST).

The first result concerning the asymptotics of a Euclidean optimization problem is the celebrated paper of Beardwood, Halton and Hammersley [BHH59] from 1959 where they proved that for n independently and identically distributed random variables $P_1, \ldots, P_n$ in $[0,1]^d$, $d \geq 2$, the optimal TSP tour length $L_{TSP}(P_1, \ldots, P_n)$ is asymptotically $n^{(d-1)/d}$, more precisely there is a constant $\alpha(L_{TSP}, d) > 0$ such that $\lim_{n\to\infty} L_{TSP}(P_1, \ldots, P_n)/n^{(d-1)/d} = \alpha(L_{TSP}, d) \int_{[0,1]^d} f(x)^{(d-1)/d} dx$ almost surely, where f is the density of the absolutely continuous part of the law of P_1. Papadimitriou [Pap78] modified in 1978 the proof and showed a similar result for the minimal matching in two dimensions. This is the first general approach, he determined four conditions so that all problems satisfying these conditions have the same asymptotic behavior. In 1981 Steele[Ste81] also presented a general approach and showed that a large class of problems has the same $n^{(d-1)/d}$ asymptotics as the TSP. Twelve years later, 1993, Rhee [Rhe93] brought isoperimetric inequalities into play and showed that Steele's results hold in the sense of complete convergence, which is stronger than almost sure convergence. In 1994 Redmond and Yukich [RY94] extended Steele's and Rhee's results to an even broader class of problems.

So Beardwood, Halton and Hammersley [BHH59] motivated a large body of research on the probabilistic analysis of Euclidean optimization problems as minimal spanning tree, minimum perfect matching, etc. Today, there is a good understanding of the general structure that underlies the asymptotic behavior of these problems. A good overview on the history and main developments in this area is given in the books of Yukich [Yuk98] and Steele [Ste97].

The first problem studied in this thesis is an important generalization of the classical traveling salesman problem: the multi depot vehicle routing problem (MDVRP). In its simplest variant, several depots and a set of customers who must be served from these depots are given. A multi depot vehicle routing tour is a set of disjoint cycles such that all customers are covered and each cycle contains exactly one depot. Note that not all depots have to be used. The problem consists of two subproblems: first, assign the customers to a depot, and secondly find an optimal routing of all customers assigned to a given depot. In the following the set of depots is denoted by D, $|D| = k$, and the set of customers by P, $|P| = n$.

In Chapter 2 we introduce the basics of the general approaches by Steele [Ste81], Rhee [Rhe93] and Redmond and Yukich [RY94]. The class of problems they can be applied to is determined via properties of the associated graphs. These properties are defined in Section 2.1. Rhee [Rhe93] uses a concentration inequality in her approach, that is introduced in Section 2.2. In Section 2.3 we sketch the approach by Redmond and Yukich [RY94]. The main tool they use is the boundary functional, that is a general modification of Euclidean optimization problems.
The MDVRP is a Euclidean optimization problem that is defined on two point sets in contrary to the classic problems that are defined on a single point set, e.g. traveling salesman problem, minimal matching etc. So we extend the classic definitions of the properties in order to treat problems defined on two point sets in this chapter.

In Chapter 3 we show that the MDVRP does not have the same properties as most Euclidean optimization problems studied in the literature, so the general approach of Chapter 2 cannot be applied directly. But in Section 3.2 we show that the boundary modification of the problem has sufficient properties to analyze its asymptotics and that it approximates the original functional. Consequently we can analyze the asymptotics of the MDVRP via the modification.

We study in Chapter 4 the asymptotic behavior of the optimal MDVRP tour length for random depot and customer sets in $[0,1]^d$ given by iid random variables with uniform distribution. We show that the asymptotic behavior depends on the customer-depot ratio n/k, and we distinguish three cases: $k = o(n)$, $k = \lambda n + o(n)$ and $k = \Omega(n^{1+\varepsilon})$. The last case is studied in Chapter 5 and first two cases in Chapter 4:
First we show: if $k = o(n)$, then $\lim_{n\to\infty} L(D,P)/n^{(d-1)/d} = \alpha(L_{TSP}, d)$ c.c. Thus, the MDVRP for "small" numbers of depots behaves asymptotically exactly like the TSP.
Secondly, we show: if $k = \lambda n + o(n)$ for a constant $\lambda > 0$, then $\lim_{n\to\infty} L(D,P)/n^{(d-1)/d} = \alpha(L_{MDVRP}, \lambda, d)$ c.c., where $\alpha(L_{MDVRP}, \lambda, d)$ is a positive constant. Note that the constant only depends on the value of λ and it does not depend on the $o(n)$ term. In this case the MDVRP shows the $n^{(d-1)/d}$ asymptotics of the TSP, but the constant differs from $\alpha(L_{TSP}, d)$. constant. Since the exact values of the constants are unknown for almost all studied Euclidean optimization problems, we do not expect to determine the exact value of the MDVRP constant. In Section 4.2.1 we give upper and lower bounds for the MDVRP constant.

This is the first analysis of the MDVRP and it is noteworthy that it considers the numbers of customers and depots. Bompadre, Dror and Orlin [BDO06] analyze a related problem, where the number of depots is constant. They give a probabilistic analysis for the capacitated multi depot vehicle routing problem, where each customer has unit demand and at each depot there are an unlimited number of vehicles with capacity Q. Each vehicle starts and ends at a depot and cannot visit more customers than its capacity allows. In this problem the k depots are fixed in advance. They show for a customer set P, $|P| = n$, of iid uniform random points in $[0,1]^2$ with expected distance ν to a closest depot that if $\lim_{n\to\infty} Q/\sqrt{n} = 0$ then $\lim_{n\to\infty} L_{MDVRP}(P)Q/n = 2\nu$ a.s., and if $\lim_{n\to\infty} Q/\sqrt{n} = \infty$ then $\lim_{n\to\infty} L_{MDVRP}(P)/\sqrt{n} = \alpha(L_{TSP}, d)$ a.s.

In Chapter 5 we consider the second problem of this thesis, the all nearest neighbor problem(ANNP). In the ANNP two point sets $D, P \subset [0,1]^d$ are given. The aim is to connect each point in P to a closest neighbor in D. So each point of P is connected to a point a in D, but the points in D may be connected to more than one point of P or none at all. We show in Chapter 5 for the total edge length $L_{ANN}(D,P)$ for iid random variables with uniform distribution that $\lim_{k\to\infty} \frac{\mathbb{E}[L_{ANN}(D,P)]}{nk^{-1/d}} = \frac{\Gamma(\frac{1}{d})\Gamma(\frac{d}{2}+1)^{1/d}}{d\sqrt{\pi}}$.

In Section 5.2 we consider the MDVRP in case $k = \Omega(n^{1+\varepsilon})$ for an arbitrary $\varepsilon > 0$. In this case most customers are connected to a depot by two edges. So the structure is very similar to the all nearest neighbor problem and the analysis of the MDVRP is based on the results for the ANNP. We prove that $\lim_{n\to\infty} \frac{\mathbb{E}[L(D,P)]}{nk^{-1/d}} = \frac{2\Gamma(\frac{1}{d})\Gamma(\frac{d}{2}+1)^{1/d}}{d\sqrt{\pi}}$ if $k = \Omega(n^{1+\varepsilon})$ for an arbitrary $\varepsilon > 0$. The concentration inequality by Rhee [Rhe93] can not be applied to this problem, since it does not have the needed properties. But in two dimensions in the case $k = n^{1+\varepsilon}$ for $0 \leq \varepsilon < 1$, we can show via an isoperimetry inequality that $\lim_{n\to\infty} \frac{L(D,P)}{nk^{-1/2}} = \frac{2\Gamma(\frac{1}{d})\Gamma(\frac{d}{2}+1)^{1/d}}{d\sqrt{\pi}} c.c.$

In Chapter 6 we treat the MDVRP in the cases $k = o(n)$ and $k = \lambda n + o(n)$ with more general distributions of the customers and depots, i.e. all distributions with an absolutely continuous part of the density, denoted by f, are considered:
As in Chapter 4 the MDVRP behaves like the TSP in case $k = o(n)$. We show: $\lim_{n\to\infty} L(D,P)/n^{(d-1)/d} = \alpha(L_{TSP}, d) \int_{[0,1]^d} f(x)^{(d-1)/d} dx$ c.c.
If $k = \lambda n + o(n)$ for a constant $\lambda > 0$, then we have $\lim_{n\to\infty} L(D,P)/n^{(d-1)/d} = \alpha(L_{MDVRP}, \lambda, d) \int_{[0,1]^d} f(x)^{(d-1)/d} dx$ c.c., where $\alpha(L_{MDVRP}, \lambda, d)$ is a positive constant. As for the uniform distribution the MDVRP shows the $n^{(d-1)/d}$ asymptotics of the TSP in this case.

In Chapter 7 we consider MDVRP heuristics. It is known that vehicle routing problems are very hard to solve - in the current state of knowledge, they can rarely be solved to optimality for sizes in excess of 50 customers, [TV02]. Usually one resorts to heuristics instead and as for other vehicle routing problems, there is a wide body of literature consisting of the application of various heuristics to the MDVRP, tested with various benchmark problems [CGL97, CGW93, RLB96, GTV02]. We give probabilistic analyses for two heuristics in Chapter 7. First, we focus on the very natural class of heuristics that follow the two-step scheme [FJ81, GTV02] *cluster first route second*, i.e., first assign each customer to a depot to form clusters of customers centered at each depot and then find an optimal routing tour connecting all customers to the corresponding depots. Let $\mathcal{C}$ denote a clustering, i.e., an assignment of customers to depots. Let $\mathcal{C}^N$ denote the clustering produced by applying the nearest neighbor rule and let $\mathcal{C}^*$ denote the clustering in an optimal tour. Let $\mathcal{T}(\mathcal{C})$ denote the total length of an optimal tour for the clustering $\mathcal{C}$. We show that $\limsup_{k\to\infty} \frac{\mathcal{T}(\mathcal{C}^N)-\mathcal{T}(\mathcal{C}^*)}{\sqrt{k}} \leq 6 \int_{[0,1]^2} (f(x))^{1/2} dx$ c.c., where f is a continuous density of the customers and depots which is bounded away from 0 and ∞.

In the landmark articles [K76, K77] Karp introduced polynomial time partitioning heuristics which for every $\varepsilon > 0$ construct a TSP tour of length at most $(1 + \varepsilon)L_{TSP}(P_1, \ldots, P_n)$ for n independently and identically distributed random points in $[0,1]^d$. For the first time this result showed that the stochastic version of an NP-hard optimization problem allows a tight approximation in polynomial time almost surely. Later, it was shown that the heuristic can be extended to other classic Euclidean optimization problems, see [Yuk98, Ste97].

In Section 7.2 we show that the heuristics can also be extended for the MDVRP in the case $k = o(n)$ and $k = \lambda n$, although the problem has not the needed properties. As before, the analysis relies on the boundary modification of the problem, which has the needed properties and has the same asymptotic behavior.

In Chapter 8 we analyze the b-degree constrained minimal spanning tree(bMST): we are given n points in $[0,1]^d$ and a degree constraint $b \geq 2$, the aim is to find the minimal weight spanning tree, where each vertex has at most degree b. This is a generalization of the path version of the TSP. We verify the conjecture by Yukich [Yuk98] that the asymptotics of the problem may be determined via the approach by Redmond and Yukich [RY94]. We show: the optimal length $L_{bMST}(P)$ of a b-degree constrained minimal spanning tree on $P_1, \ldots, P_n$ given by iid random variables with values in $[0,1]^d$ satisfies $\lim_{n\to\infty} \frac{L_{bMST}(P_1,\ldots,P_n)}{n^{(d-1)/d}} = \alpha(L_{bMST}, d) \int_{[0,1]^d} f(x)^{(d-1)/d} dx$ c.c., where $\alpha(L_{bMST}, d)$ is a positive constant and f is the density of the absolutely continuous part of the law of P_1. In the case $b = 2$, the b-degree constrained MST has the same behavior like the TSP, we have $\alpha(L_{bMST}, d) = \alpha(L_{TSP}, d)$.

Chapter 2

The General Approach

The general approach used in this thesis in order to describe the asymptotics of the total edge length of graphs associated to combinatorial optimization problems has been developed in the last 50 years, as noted in the introduction. During these years the approach has been refined in order to determine the asymptotics of a growing class of problems. It turns out that properties of the graphs, e.g. subadditivity and smoothness, form the key conditions that determine the asymptotic behavior of the Euclidean optimization problems. In the next section these properties are defined and the first unifying result determining the asymptotics of a large class of problems, published in 1981 by Steele [Ste81], is stated. The definitions of the properties are generalized in order to treat the multi depot vehicle problem that is defined on two point sets in contrary to the classic problems that are defined on a single set.

Rhee [Rhe93] generalized Steele's approach in 1993, her approach can be applied to a larger class of problems. Additionally, Steele's limit theorem shows only almost sure convergence and Rhee's limit theorem yields complete convergence. Rhee discovered that isoperimetric inequalities yield a general concentration inequality for Euclidean optimization problems, and the complete convergence is shown via the concentration inequality. In the proof of her limit theorem, the properties of the graphs are used to determine the asymptotics of the mean of the total edge length. Then, concentration inequalities are applied to show that the total edge length has the same behavior as its mean. Rhee's limit theorem and the used concentration inequality is stated in Section 2.2.

In 1994 Redmond and Yukich [RY94] generalized Rhee's approach. Their limit theorem ensures complete convergence not only for independent and uniformly distributed random variables as considered by Rhee, but for any law of the random variables(with an absolute continuous part of the density) giving the points. Furthermore they were able to give rates of convergence. The main tool of their approach is the 'quasiadditive' structure of many problems, that is defined via modifications of the considered problems that approximate the original problem. This approach is sketched in Section 2.3.

2.1 Euclidean Functionals and Steele's Theorem

We give definitions of the necessary properties of a total edge length function F of a Euclidean optimization problem defined on two finite subsets of $\mathbb{R}^d$, $d \geq 2$. In this thesis the dimension is always at least 2. Naturally, the definitions are based on the definitions used for problems on a single vertex set. We begin with the definition of Euclidean functionals.

Let F be a function $F : \mathcal{S} \times \mathcal{S} \to \mathbb{R}^+$, where $\mathcal{S}$ is the set of finite subsets of $\mathbb{R}^d$. F has the *translation invariance* property if for all $y \in \mathbb{R}^d$ and finite subsets $D, P \subset \mathbb{R}^d$:

$$F(D, P) = F(D + y, P + y),$$

the *homogeneity* property, if for all $\alpha > 0$ and finite subsets $D, P \subset \mathbb{R}^d$:

$$F(\alpha D, \alpha P) = \alpha F(D, P),$$

and the *normalization* property, if

$$F(\emptyset, \emptyset) = 0.$$

From now on all functionals we are speaking of are Euclidean functionals. In the following $|P|$ denotes the cardinality of a point set P.

As mentioned before, the Euclidean functionals were originally defined for functions on a single finite point set in $\mathbb{R}^d$. We are considering the MDVRP, where the point set contains two different kinds of points(customers/depots), so we extend the definition of Euclidean functionals. We demand that the conditions that hold originally for a single set, are satisfied for two point sets, e.g. the original homogeneity property is $F(\alpha P) = \alpha F(P)$. In the following we present other properties extended in the same way to functionals defined on two point sets. We will compare the properties of the MDVRP with classic problems, although the properties are not explicitly defined for single set problems in this thesis. If we speak of properties of functions on a single point set, we demand that the property is satisfied for the single set.

Now we consider additional features of functionals associated to Euclidean optimization problems. Each property alone may seem weak, but a combination together with the basic features of Euclidean functionals is a strong composition. A very simple and powerful one is the subadditivity.

F is called *subadditive* if for all d-dimensional rectangles $R \subset \mathbb{R}^d$, all finite subsets $D, P \subset R$ and all partitions of R into rectangles R_1 and R_2,

$$F(D,P) \le F(D \cap R_1, P \cap R_1) + F(D \cap R_2, P \cap R_2) + C,$$

with a constant C that may depend on d and the diameter of R. In the following the value of the constant C may change from line to line and it may depend on d and the diameters of the considered rectangles.

It is easy to see that most Euclidean optimization problems are subadditive, e.g. traveling salesman problem, minimal spanning tree, minimal matching etc. We give a reasoning for the traveling salesman problem, and similar argumentations can be applied to most problems: we are given a finite set P and a d-dimensional rectangle $R = R_1 \cup R_2$ with diameter $\mathrm{diam}(R)$. Let T_1 and T_2 be optimal tours in R_1 respectively R_2. In each tour an arbitrary edge is deleted, and by two edges of length at most the diameter of R the tours are merged into a single tour T. The optimal tour in R is at most as long as T, and the length of T is at most the sum of the lengths of T_1 and T_2 and the connecting edges, so we have

$$L_{TSP}(P) \le L_{TSP}(P \cap R_1) + L_{TSP}(P \cap R_2) + 2\,\mathrm{diam}(R).$$

This property is used over and over in the study of the asymptotic behavior of these problems, first of all it has been used in the seminal paper of Bearwood, Halton and Hammersley [BHH59]. The subadditive structure of the graphs associated to optimization problems is used to approximately express the global graph length as a sum of the lengths of local components, this is a crucial method in the probability theory of Euclidean optimization problems.

Rhee [Rhe93] was the first to show that subadditivity has useful non-trivial consequences. She showed the following growth bound for subadditive Euclidean functionals which are defined on a single point set P.

Lemma 2.1. *[Rhe93] Let F be a subadditive Euclidean functional. Then there exists a constant C such that for all cubes $R \subset [0,1]^d$ and all point sets $P \subset R$ we have*

$$F(P) \le C|P|^{(d-1)/d},$$

with a constant C.

Instead of using subadditivity in order to express the global graph length as a sum of local components, one can use superadditivity. The difference is that we obtain by superadditivity a lower bound of the global length, while subadditivity gives an upper bound.

A functional F is called *superadditive*, if for all rectangles $R \subset \mathbb{R}^d$, all finite subsets $D, P \subset R$ and all partitions of R into rectangles R_1 and R_2

$$F(D,P) \ge F(D \cap R_1, P \cap R_1) + F(D \cap R_2, P \cap R_2).$$

Another very strong and useful property in the analysis of classic functionals is monotonicity. The lengths of some functionals are monotone increasing with an increasing number of points.

A functional F is *monotone* if for all finite sets $D, P \subset \mathbb{R}^d$ and $y \in \mathbb{R}^d$

$$F(D,P) \leq F(D \cup \{y\}, P) \text{ and } F(D,P) \leq F(D, P \cup \{y\}).$$

For example, the traveling salesman functional is monotone, since adding a point can not shorten the tour. But this property is not shared by all classic problems. The minimal spanning tree is obviously not monotone, having four points in the corners of a square and adding a point in the center shortens the total length of a minimal spanning tree.

By using subadditivity and monotonicity only, Steele [Ste81] characterized the asymptotic behavior for the large class of Euclidean functionals with these properties:

Theorem 2.2. *[Ste81] Suppose F is a monotone subadditive Euclidean functional. If the random Variables $X_1, \ldots, X_n$ are independent with the uniform distribution on $[0,1]^d$, then as $n \to \infty$ we have with probability one that*

$$\frac{F(X_1, \ldots, X_n)}{n^{(d-1)/d}} = \beta_F,$$

where $\beta_F \geq 0$ is a constant.

We only mention that Steele [Ste81] also showed the asymptotics of Euclidean functionals having these and some additional properties for more general distributions of the random variables giving the points. Theorem 2.2 can be applied to many functionals, an example is the traveling salesman functional. It is subadditive and monotone as shown above.

Rhee [Rhe93] introduced with the smoothness of Euclidean functionals a tool to handle subadditive functionals that are not monotone.

A Euclidean functional F is *smooth* if there is a constant $C > 0$ such that for all finite sets $D_1, D_2, P_1, P_2 \subset \mathbb{R}^d$,

$$|F(D_1 \cup D_2, P_1 \cup P_2) - F(D_1, P_1)| \leq C(|D_2| + |P_2|)^{(d-1)/d}.$$

So smoothness bounds the variation of F when points and depots are added and deleted. Rhee [Rhe93] proved that the smoothness condition yields inequalities that show that a smooth Euclidean functional is sharply concentrated around its mean. This is done via an isoperimetry inequality, her approach is the topic of the next chapter.

2.2 Isoperimetry and Rhee's Theorem

Rhee [Rhe93] was able to improve two points in Steele's Theorem 2.2 [Ste81]. The first drawback of Steele's approach is that it is only applicable to monotone functionals, so it does not determine the asymptotics of the minimal spanning tree or the minimal matching. Rhee was able to show that smoothness can be used as a substitute for monotonicity, so the theorem can be applied to a larger class of functionals, e.g. the above mentioned minimal matching and the minimal spanning tree. The second point is that Steele's theorem states almost sure convergence, where Rhee's theorem covers complete convergence.

We remind the reader of the different notions of stochastic convergence. Let $(Y_n)_{n\geq 1}$ be a sequence of random variables and let Y be a random variable on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We say Y_n *converges almost surely* to Y, written $\lim_{n\to\infty} Y_n = Y$ a.s., if

$$\mathbb{P}[\omega \in \Omega : \lim_{n\to\infty} Y_n(\omega) = Y(\omega)] = 1,$$

and we say Y_n *converges completely* to Y, written $\lim_{n\to\infty} Y_n = Y$ c.c., if for all $\varepsilon > 0$

$$\sum_{n=1}^{\infty} \mathbb{P}[|Y_n - Y| > \varepsilon] < \infty.$$

Note that the Borel-Cantelli lemma shows that the condition for complete convergence is sufficient for almost sure convergence. But the main benefit of complete convergence is that it yields convergence results for two different random problem models that differ in the transition from $F(X_1, \ldots, X_n)$ to $F(X_1, \ldots, X_n, X_{n+1})$, where F is a functional and $X_1, \ldots, X_{n+1}$ are random variables giving the points. In the *incrementing problem model* an additional sample point is given by X_{n+1} in order to get $F(X_1, \ldots, X_n, X_{n+1})$, while in the *independent problem model* a completely new sample of points $\{x_1, \ldots, x_{n+1}\}$ is given. The important point is that almost sure convergence results for the independent model imply almost sure convergence for the incrementing model, but the converse is false in general; where complete convergence covers both models. Weide [Wei78] was the first to distinguish the models in the probabilistic analysis of algorithms. In his thesis he defines that an algorithm *succeeds strongly* if the random variable X_n giving the error of an algorithm of a problem of size n converges almost surely to 0. The difference of the models is that in the independent model the variables $X_1, \ldots, X_n, \ldots$ are independent, which is obviously false for the incrementing model. We state the vital part of Weide's Theorem.

Theorem 2.3. *[Wei78] If an algorithm succeeds strongly in the independent problem model, then it succeeds strongly in the incremental model, but not necessarily vice versa.*

Rhee's tool in order to prove complete convergence is the following inequality. It shows that, except for a small set with polynomially small probability, smooth Euclidean functionals are close to their means. By the inequality it is sufficient to determine the asymptotics of the mean in order to show complete convergence of the functional. This simplifies the analysis of many problems and we will use the fact for all considered problems.

Theorem 2.4. *[Rhe93] Let $X_1, \ldots, X_n$ be independent uniformly distributed random variables with values in $[0,1]^d$, $d \geq 2$, and let $F(X_1, \ldots, X_n)$ be a smooth Euclidean functional. Then there are constants C and C' such that for all $t > 0$:*

$$\mathbb{P}\left[|F(X_1, \ldots, X_n) - \mathbb{E}\left[F(X_1, \ldots, X_n)\right]| > t\right] \leq C \exp\left(-\frac{1}{Cn}\left(\frac{t}{C'}\right)^{2d/(d-1)}\right).$$

The theorem is proven via an isoperimetric inequality. We do not state a proof of the theorem, but refer to Talagrand [Tal95] and Ledoux [Led96] for a complete treatment of isoperimetry, and introduce the isoperimetric inequality used in the proof of Theorem 2.4. The inequality origins in work of Milman and Schechtman [MS86] that will be applied in a proof in Chapter 5, where we can not apply Theorem 2.4 directly. We need the following notation: the Hamming distance H on $\Omega^n = ([0,1]^d)^n$ measures the distance of x and y by the number of coordinates in which x and y do not match:

$$H(x, y) := |\{1 \leq i \leq n : x_i \neq y_i\}|.$$

The Hamming distance between $y \in \Omega^n$ and a set $A \subset \Omega^n$ is

$$\Phi_A(y) = \min\{H(x, y) : x \in A\}.$$

Let μ^n denote the product measure of Ω^n. Now we are ready to give the isoperimetric inequality for the Hamming distance H.

Lemma 2.5. *If $A \subset \Omega^n$ satisfies $\mu^n(A) \geq \frac{1}{3}$ then*

$$\mu^n\left(\{y \in \Omega^n : \Phi_A(y) \geq t\}\right) \leq 6 \exp\left(-t^2/8n\right).$$

To conceive the connection to Theorem 2.4 consider the t-enlargement A_t of a set $A \subset \Omega^n$

$$A_t := \{x \in \Omega^n : \exists y \in A \text{ s.t. } H(x, y) \leq t\},$$

that contains A and all points of Ω^n that are within Hamming distance t of A. The Lemma 2.5 states that for a set $A \subset \Omega^n$ with $\mu^n(A) \geq 1/3$ and $t \gg \sqrt{n}$ the measure of A_t is almost 1, more precisely $\mu^n(A_t) \geq 1 - 6\exp\left(-t^2/8n\right)$. The smoothness of a Euclidean functional says that modifying some points does not change the total edge length too much. Since the number of modified points equals the Hamming distance, Lemma 2.5 can be used to prove Theorem 2.4.

At last we state the limit theorem by Rhee [Rhe93]:

Theorem 2.6. *[Rhe93] Let $X_1, \ldots, X_n$ be independent uniformly distributed random variables with values in $[0,1]^d$. If F is a smooth subadditive Euclidean functional on $\mathbb{R}^d$, then*

$$\lim_{n\to\infty} \frac{\mathbb{E}[F(X_1,\ldots,X_n)]}{n^{(d-1)/d}} = \alpha(F,d)$$

where $\alpha(F,d)$ is a constant. For some universal constants C and C', we have, for all $n \geq 1$, and $t \geq 0$,

$$\mathbb{P}\left[|F(X_1,\ldots,X_n) - \mathbb{E}[F(X_1,\ldots,X_n)]| \geq t\right] \leq C \exp\left(-\frac{1}{Cn}\left(\frac{t}{C'}\right)^{2d/(d-1)}\right).$$

2.3 The Boundary Functional by Redmond and Yukich

Redmond and Yukich [RY94] developed a general unifying approach describing the stochastic behavior of the total edge length of a variety of graphs associated to combinatorial optimization problems, e.g. the traveling salesman problem, minimal spanning tree, minimal Steiner tree and so on. The core of this approach is the combined use of subadditivity and superadditivity. With a certain "quasiadditive" property they gave an approach to the limit theory of Euclidean functionals that is more general than Steele's approach [Ste81], since it is not limited to monotone increasing functionals. The approach also extended the work of Rhee [Rhe93] by showing complete convergence for any density f of the law of the random variables giving the points. Furthermore they were able to show rate of convergence results with the help of the quasiadditive structure. Redmond and Yukich get this structure in the following way. They show that the edge length function of many graph problems is not only subadditive, but it admits a simple and natural modification, called boundary functional, that is smooth and superadditive and approximates the original functional. Roughly speaking, the boundary modification of a problem is the same problem, but edges that lie on the boundary are assigned zero length. The boundary functional has to be defined for each problem explicitly, so that it displays the desired features. We state the definition of the boundary TSP functional in order to give an intuition of boundary functionals: for all rectangles $R \subset \mathbb{R}^d$, finite point sets $P \subset R$ and points a, b on the boundary of R let $L'_{TSP}(P, \{a,b\})$ denote the length of the shortest path through all points of P with endpoints a and b. The *boundary functional* L^B_{TSP} is defined by

$$L^B_{TSP}(P) := \min\left\{L_{TSP}(P),\ \inf\left\{\sum_{i\geq 1} L'_{TSP}(P_i, \{a_i, b_i\})\right\}\right\},$$

where the infimum ranges over all sequences $(a_i, b_i)_{i\geq 1}$ of points on the boundary of R and all partitions $(P_i)_{i\geq 1}$ of P.

The boundary functional is only useful for the analysis of the asymptotics of a functional if it approximates the functional. A functional F defined on two point sets and its boundary modification F^B are called *pointwise close* if for all $D, P \subset [0,1]^d$

$$|F(D,P) - F^B(D,P)| = o(|P|^{(d-1)/d}).$$

Redmond and Yukich call a smooth subadditive functional that is pointwise close to its superadditive boundary functional *quasiadditive*. We state the limit theorem by Redmond and Yukich and recommend the monograph [Yuk98] by Yukich for more information.

Theorem 2.7. *[RY94] Let $X_1, \ldots, X_n$ be independent identically distributed random variables with values in $[0,1]^d$, $d \geq 2$, and let $F(X_1, \ldots, X_n)$ be a quasiadditive smooth Euclidean functional, then*

$$\lim_{n\to\infty} F(X_1, \ldots, X_n)/n^{(d-1)/d} = \alpha(F) \int f(x)^{(d-1)/d} dx \ \ c.c.,$$

where f is the absolutely continuous part of the law of X_1.

Note that in this thesis we first consider uniformly distributed variables. Furthermore we mainly consider functionals that are not quasiadditive, but functionals that are close to their smooth superadditive boundary functional. As we will see, instead of attacking a problem directly, it is often easier to determine the stochastic behavior of the superadditive boundary functional and by the closeness to the original functional we get the behavior of the original problem. So we take a careful look the following central general asymptotic result for independent and uniformly distributed random variables, formulated by Yukich [Yuk98] in the spirit of Rhee:

Theorem 2.8. *If F^B is smooth superadditive Euclidean functional on $\mathbb{R}^d$, then*

$$\lim_{n\to\infty} \frac{F^B(P)}{n^{(d-1)/d}} = \alpha(F^B, d) \ \ c.c.,$$

where $\alpha(F^B, d)$ is a positive constant. If F is a Euclidean functional on $\mathbb{R}^d$, which is pointwise close to F^B, then

$$\lim_{n\to\infty} \frac{F(P)}{n^{(d-1)/d}} = \alpha(F^B, d) \ \ c.c.$$

Applying the theorem is a quite easy way to determine the asymptotics of a functional: we have to define a proper boundary functional that is close to the original functional and satisfies superadditivity and smoothness. In the next chapters we modify the proof of the theorem in order to apply the theorem to the MDVRP functional that is defined on two point sets. In Chapter 6 and Chapter 8 we consider more general distributions and we modify the proof of Theorem 2.7. The proof is based on the tools and methods used for uniformly distributed random variables.

Chapter 3

The Multi Depot Vehicle Routing Problem

In 1959 Dantzig and Ramsey [DR59] presented a paper concerning the delivery of gasoline to gas stations and introduced the vehicle routing problem. Nowadays there is a large variety of models that take a lot of real-world constraints into considerations. They cover the number and capacity of the delivering vehicles, different regional provenance of the goods, as well as the demand and availability of the customers and so forth. The book "The Vehicle Routing Problem" edited by P. Toth and D. Vigo [TV02] gives an extensive survey of the models and algorithms developed in the last decades.

As mentioned before, we consider the following very simple variant of the multi depot vehicle routing problem (MDVRP): Let $k, n \in \mathbb{N}$ and $D = \{D_1, \ldots, D_k\}$ respectively $P = \{P_1, \ldots, P_n\}$ be sets of points in $[0,1]^d$ with the usual Euclidean metric. The D_i's are called depots and the P_i's points or customers. A multi depot vehicle routing tour is a set of disjoint cycles such that all points are covered and each cycle contains exactly one depot. Note that not all depots have to be used. If $P = \emptyset$ or $D = \emptyset$ the length of an optimal tour is 0, since there is no cycle containing points and a depot. The goal is to find a tour of minimum length. $L(D, P)$ denotes the length of an optimal MDVRP tour for depot set D and point set P.

In order to prepare the asymptotic analysis, we examine the MDVRP functional in Section 3.1 and show that it does not have the same properties as most Euclidean functionals studied in the literature, it is neither sub- nor superadditive. But in Section 3.2 we introduce the boundary functional and prove that it approximates the original functional. Furthermore, it is shown that the boundary functional is a smooth superadditive functional. These properties suffice to determine its asymptotics, which is done in Chapter 4.

3.1 Properties of the MDVRP Functional

The main difference between the MDVRP functional L and most classic functionals is: it has neither a subadditive nor a superadditive structure. We will see that this behavior is caused by the two set structure of the MDVRP.

First we show that L is not subadditive. Note that a functional F is called subadditive if for all d-dimensional rectangles $R \subset \mathbb{R}^d$, all finite subsets $D, P \subset R$ and all partitions of R into rectangles R_1 and R_2, $F(D, P) \leq F(D \cap R_1, P \cap R_1) + F(D \cap R_2, P \cap R_2) + C$, with a constant C that may depend on d and the diameter of R. A counterexample showing that the MDVRP is not subadditive is given in Figure 3.1 and 3.2. As shown in the figures, the problem is that a rectangle containing only points does not produce a depot tour, since a feasible depot tour contains a depot. In this way it is possible to diminish the instance to an instance with only a few customers. This is not possible for problems defined on a single set, since they yield a graph with edges as soon as we have two points in a rectangle.

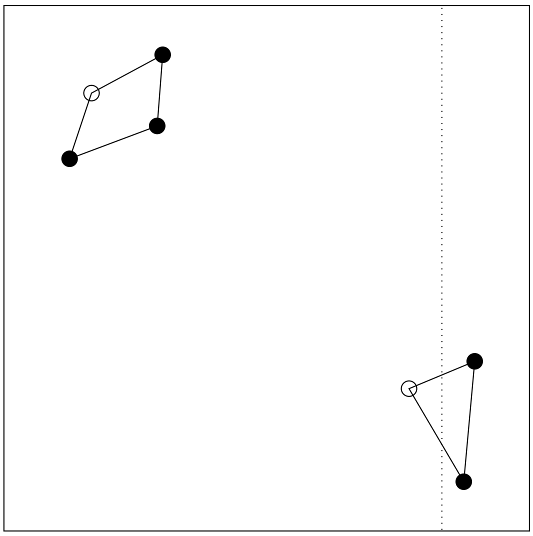

Figure 3.1: The MDVRP tour in R, the partition is indicated by the dotted line.

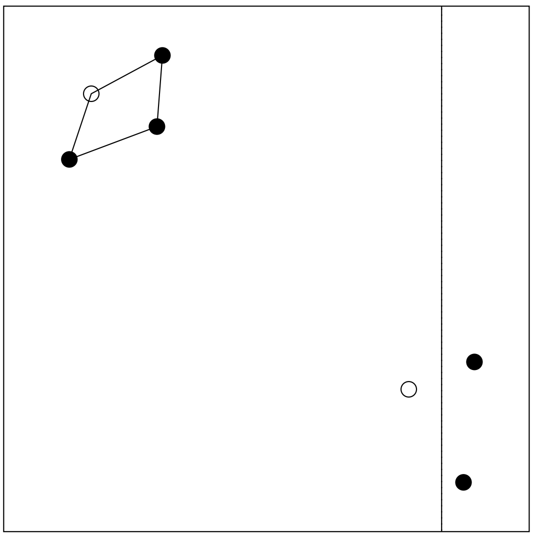

Figure 3.2: The MDVRP tours in R_1 and R_2.

Recall that a functional F is called superadditive, if for all rectangles $R \subset \mathbb{R}^d$, all finite subsets $D, P \subset R$ and all partitions of R into rectangles R_1 and R_2 $F(D, P) \geq F(D \cap R_1, P \cap R_1) + F(D \cap R_2, P \cap R_2)$. As in the counterexample for the subadditivity, the two set structure of the MDVRP functional is also exploited to show that the functional is not superadditive, a counterexample is given in Figure 3.3 and 3.4. In the example a subtour in an optimal solution is enlarged by cutting out the used depot, so the subtour has to be connected to a different depot that is far away.

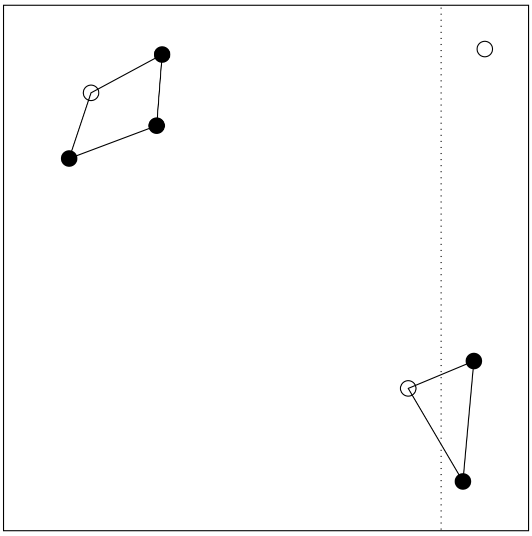

Figure 3.3: The MDVRP tour in R, the partition is indicated by the dotted line.

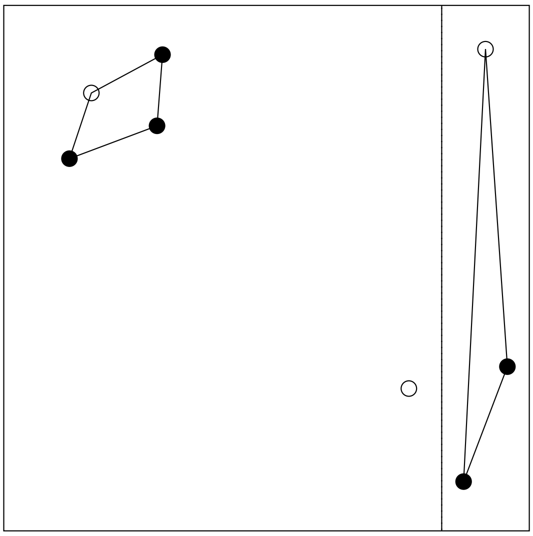

Figure 3.4: The MDVRP tours in R_1 and R_2.

Another feature that the MDVRP functional does not share with other problems is monotonicity. This is obvious, since adding a depot may shorten the MDVRP tour.

Altogether, the multi depot problem is not monotone and it is neither subadditive nor superadditive. As noted in Chapter 2, these properties are used a lot in the study of Euclidean optimization problems. So there is a very significant difference between the multi depot problem and most Euclidean optimization problems, since we can not apply the ideas used in the classic results from Chapter 2 for the stochastic behavior of optimization problems that rely on the subadditivity or superadditivity directly. Thus, we state this fact in the following lemma.

Lemma 3.1. *The MDVRP functional is not monotone and it is neither subadditive nor superadditive.*

Thus, in the next section we consider a modification of the MDVRP functional, that has the same asymptotic behavior and that is easier to analyze: the boundary MDVRP functional.

3.2 Properties of the Boundary MDVRP Functional

For the MDVRP functional, the boundary modification of Euclidean functionals introduced by Redmond and Yukich [RY94] helps to overcome the lack of sub- and superadditivity. The boundary modification of the total edge length function of the MDVRP is the total edge length function of the least expensive depot tour, where the cost of traveling along any path on the boundary is zero, and the paths connected to the boundary do not have to contain a depot, see Figure 3.5 and 3.6. The formal definition of the boundary functional of the MDVRP follows: for all rectangles $R \subset \mathbb{R}^d$, finite point sets $D, P \subset R$ and points a, b on the boundary of R let $L'(\emptyset, P, \{a, b\})$ denote the length of the shortest path through all points of P with endpoints a and b. The *boundary functional* L^B is defined by

$$L^B(D,P) := \min \left\{ L(D,P),\ \ \inf \left\{ L(D,P_1) + \sum_{i>1} L'(\emptyset, P_i, \{a_i, b_i\}) \right\} \right\},$$

where the infimum ranges over all sequences $(a_i, b_i)_{i>1}$ of points on the boundary of R and all partitions $(P_i)_{i\geq 1}$ of P.

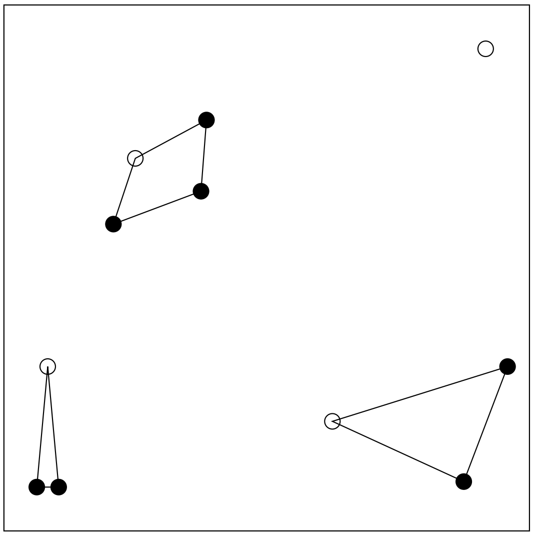

Figure 3.5: A MDVRP tour.

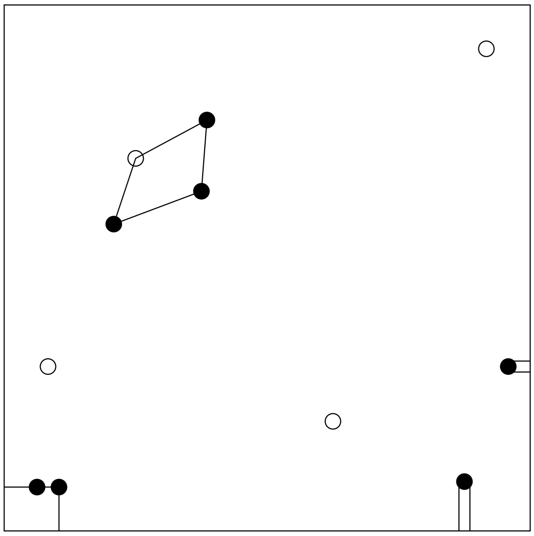

Figure 3.6: A boundary MDVRP tour.

From the definition of the boundary functional it is easy to see that $L^B(D,P) \leq L(D,P)$. Furthermore, the boundary functional of the MDVRP is superadditive: Consider a rectangle $R \subset \mathbb{R}^d$, finite subsets $D, P \subset R$ and a partition of R into rectangles R_1 and R_2. The restriction of the boundary tour in R to the rectangles R_i for $i \in \{1, 2\}$ defines boundary tours on $P \cap R_i$ and $D \cap R_i$ that are at least as large as $L^B(P \cap R_i, D \cap R_i)$, so we get $L^B(D, P) \geq L^B(P \cap R_1, D \cap R_1) + L^B(P \cap R_2, D \cap R_2)$, see Figure 3.7 and 3.8.

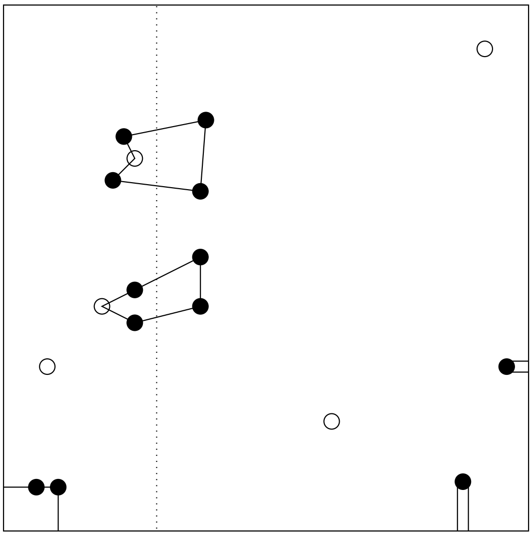

Figure 3.7: The optimal boundary MD-VRP tour in R.

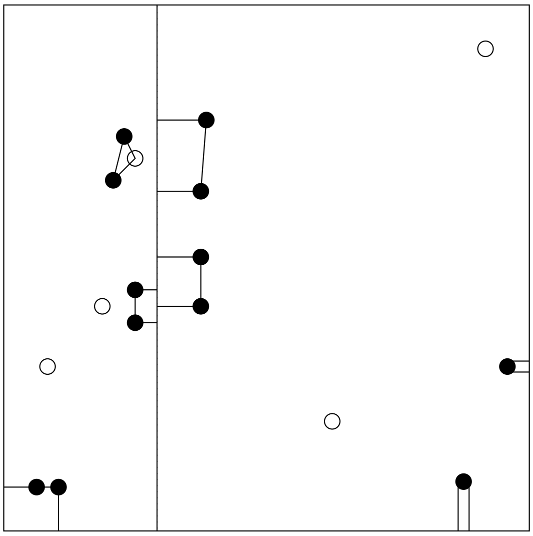

Figure 3.8: The optimal boundary MD-VRP tours in R_1 and R_2.

We state this fact in the following lemma.

Lemma 3.2. *The boundary functional of the MDVRP is superadditive.*

The boundary functional satisfies the inequality in the next lemma that is used as a substitute for subadditivity.

Lemma 3.3. *Let D and P be sets of depots respectively customers in $[0,1]^d$. If $[0,1]^d$ is partitioned into m^d congruent subcubes Q_j, then*

$$L^B(D, P) \leq \sum_{i=1}^{m^d} L^B(D \cap Q_i, P \cap Q_i) + m^{d-1} C \cdot \left(\frac{n}{m^d}\right)^{(d-2)/(d-1)} .$$

Proof. Let $|P| = n$ and $|D| = k$. Optimal tours in the subcubes Q_i are merged into a boundary tour in $[0,1]^d$ in the following way. Consider two neighboring subcubes Q_i and Q_j. There are two types of paths connected to the separating face between Q_i and Q_j: paths that have their start and endpoint on the face and paths that are only connected with one point of the face. We denote those points on the separating face where paths of the first type meet by B_1 and those points where paths of the second type meet by B_2. If the cardinality of B_2 is odd, a point of the boundary of the separating face is added to B_2. Now, the edges of a minimal matching of B_2 are adjoined to the graph. The total length of the added edges is $\frac{C}{m}|B_2|^{(d-2)/(d-1)}$ by Lemma 2.1, since the face is a $d-1$ dimensional unit cube stretched by a factor of $\frac{1}{m}$. Furthermore, we add a minimal perfect matching of B_1 and a TSP tour through B_1 to the graph. In the connected component containing B_1 all vertices have even degree. So there is a Eulerian tour through the component. We turn the Eulerian tour into a TSP tour by shortcuts, delete an edge and connect both endpoints to the boundary of $Q_i \cup Q_j$. The total length of all edges used to connect the first type paths to the boundary of $Q_i \cup Q_j$ is at most $\frac{C}{m}|B_1|^{(d-2)/(d-1)}$ by Lemma 2.1. The number of faces that have to be considered in this procedure is bounded by dm^d. In every face i let B_1^i and B_2^i denote the number of points on the face where paths of the first respectively second type meet. The total length of all added edges is

$$\sum_{i=1}^{dm^d} \frac{C}{m} \left(|B_1^i|^{(d-2)/(d-1)} + |B_2^i|^{(d-2)/(d-1)} \right). \tag{3.1}$$

Each point may be connected to the boundary of its subcube by two edges, so $\sum_{i=1}^{dm^d} |B_1^i| + |B_2^i| \leq 2n$. The sum (3.1) is maximized for $|B_j^i| = \frac{2n}{2dm^d}$ for $i = 1, \ldots, dm^d$ and $j = 1, 2$. We have

$$L^B(D, P) \leq \sum_{i=1}^{m^d} L^B(D \cap Q_i, P \cap Q_i) + Cm^{d-1} d \left(\frac{n}{dm^d} \right)^{(d-2)/(d-1)}.$$

□

As indicated in section 2.3, the asymptotic behavior of Euclidean functionals can often be determined via the boundary modification. This is the case if the boundary functional approximates the original functional, i.e. if the functional and its boundary functional are pointwise close. We show that this is true for the MDVRP functional in the following lemma.

Lemma 3.4. *For the tour length L of MDVRP and the boundary functional L^B of MDVRP we have*

$$|L(D, P) - L^B(D, P)| \leq C|P|^{(d-2)/(d-1)},$$

for a constant $C > 0$.

Proof. Since $L(D, P) \geq L^B(D, P)$, we only have to show that $L(D, P) - L^B(D, P) \leq C|P|^{(d-2)/(d-1)}$. Let G be the graph associated to $L^B(D, P)$. We modify the paths in G that are connected to the boundary in order to construct an MDVRP tour. Let B denote the set of points where the graph G meets the boundary of $[0, 1]^d$. Since B is a set of endpoints of paths, $|B|$ is even and a perfect matching of B exists. We adjoin to G the edges of a perfect minimal matching of B and a traveling salesman tour through B, Figure 3.9. By Lemma 2.1, the total length of the added edges is at most $C|B|^{(d-2)/(d-1)}$, since all points of B are on the boundary of $[0,1]^d$, which has dimension $d-1$.

So all paths of G are connected via the TSP tour, see Figure 3.9. All points connected by the tour have even degree, so there is a Eulerian tour through this connected component. Shortcutting the Eulerian tour yields a traveling salesman tour. Furthermore, we eliminate in this tour the points of B by shortcuts. Connecting the resulting tour to a depot with two edges of length at most $\sqrt{d}$ we get, together with the unmodified cycles in G, a feasible tour of length $L^B(D,P) + C|B|^{(d-2)/(d-1)} + 2\sqrt{d}$, Figure 3.10. Since $|B| \leq 2|P|$, we have

$$L(D,P) \leq L^B(D,P) + C|P|^{(d-2)/(d-1)}.$$

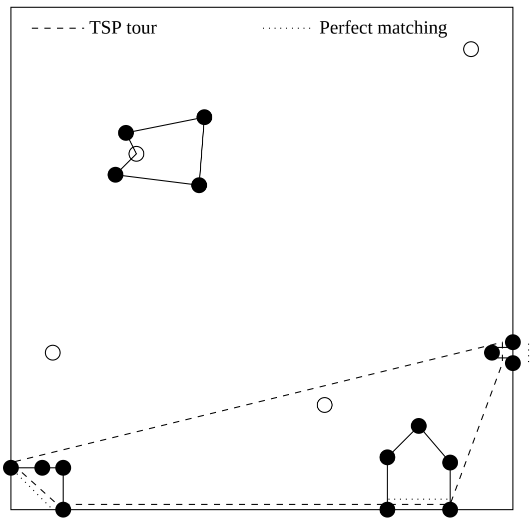

Figure 3.9: The boundary MDVRP tour with added TSP tour and matching.

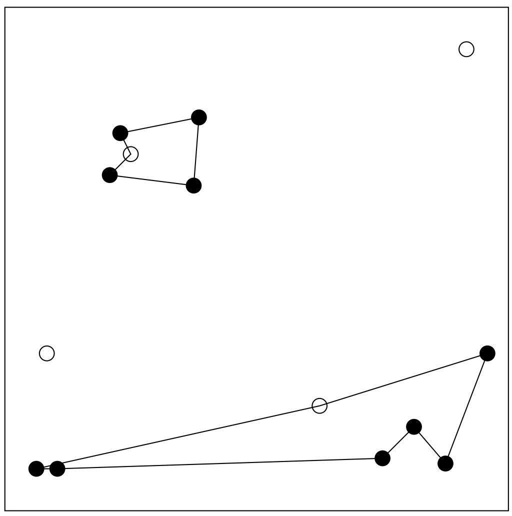

Figure 3.10: The resulting tour.

□

Since we are trying to modify Theorem 2.8 in order to handle the MDVRP functional, a functional defined on two point sets, we need in addition to superadditivity the smoothness of the boundary functional.

Lemma 3.5. *The boundary functional L^B of the MDVRP is smooth.*

Proof. Let $D_1, D_2, P_1, P_2 \subset [0,1]^d$ be sets containing depots respectively points. To see that

$$L^B(D_1 \cup D_2, P_1 \cup P_2) - L^B(D_1, P_1) \leq C(|D_2| + |P_2|)^{(d-1)/d},$$

consider an optimal boundary MDVRP tour for D_1 and P_1 and form a traveling salesman tour through all elements of P_2 and a depot of D_2, see Figure 3.11 and 3.12. This is a feasible solution for the boundary MDVRP on $D_1 \cup D_2$ and $P_1 \cup P_2$ if $D_2 \neq \emptyset$. The length of the tour through D_2 and P_2 is at most $C(|D_2|+|P_2|)^{(d-1)/d}$ for a $C > 0$ by Lemma 2.1. If $D_2 = \emptyset$, we transform the cycle into a path starting and ending at the boundary of $[0,1]^d$. The tour/path together with the graph associated to $L^B(D_1, P_1)$ yields a feasible solution for the problem on $D_1 \cup D_2$ and $P_1 \cup P_2$ of length at most $L^B(D_1, P_1) + C(|D_2| + |P_2|)^{(d-1)/d}$.

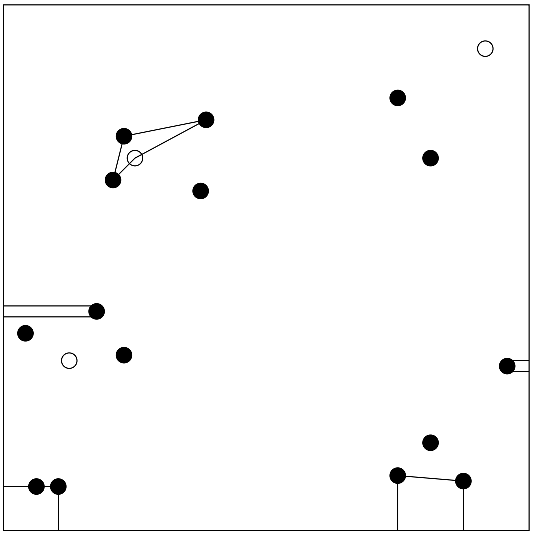

Figure 3.11: The boundary MDVRP tour on P_1 and D_1.

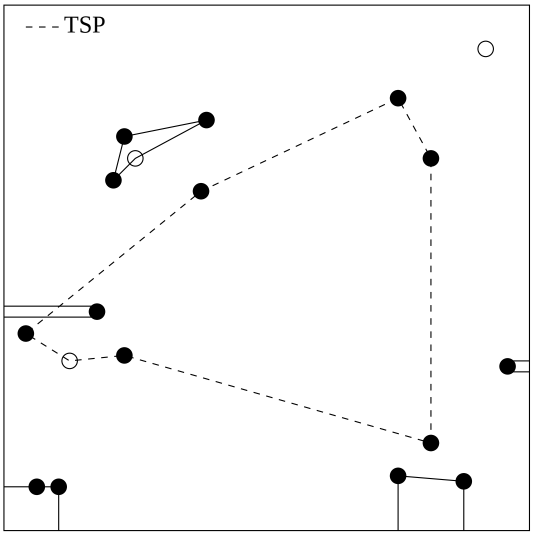

Figure 3.12: The boundary MDVRP tour on P_1 and D_1 with dashed TSP tour through P_2 and a depot of D_2.

Once we show

$$L^B(D_1, P_1) \leq L^B(D_1 \cup D_2, P_1 \cup P_2) + C(|D_2| + |P_2|)^{(d-1)/d},$$

the smoothness of L^B follows. We start with a graph G associated to $L^B(D_1 \cup D_2, P_1 \cup P_2)$, see Figure 3.13, and we remove all elements of D_2 and P_2. Let B denote the set of points where the graph meets the boundary of $[0,1]^d$, we add B to the graph, Figure 3.14. In the following, we do not consider unmodified components in G.

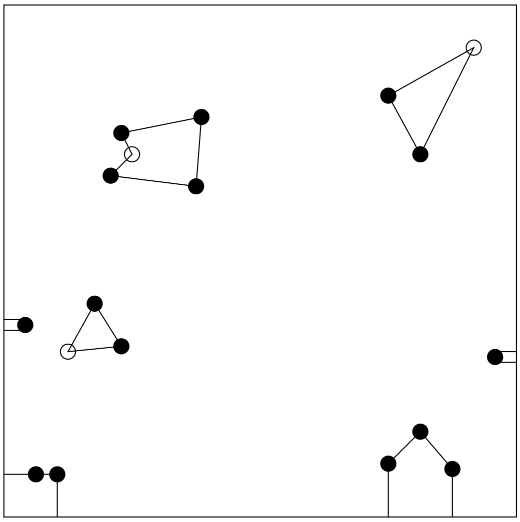

Figure 3.13: The boundary MDVRP tour.

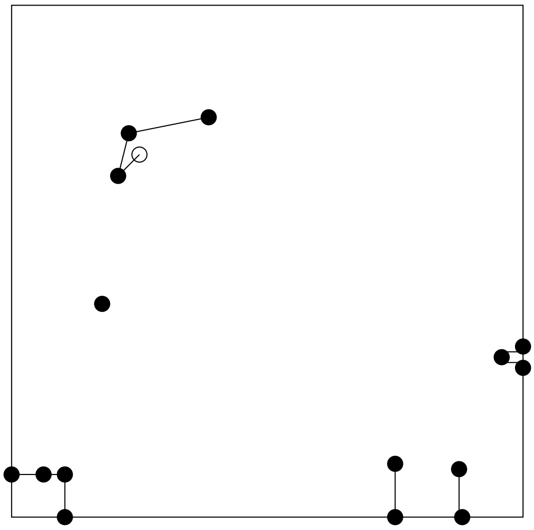

Figure 3.14: P_2 and D_2 deleted and B added.

The deletion of D_2 and P_2 generates at most $2(|D_2|+|P_2|)$ connected components, since the elements of D_2 and P_2 are either in a path connected to the boundary or they are in a closed cycle. Deleting the first element in a path results in at most two connected components and deleting further elements in this path generates at most an additional connected component for each element. Deleting elements in a closed cycle results in at most one connected component for each element.

The resulting components are either paths or isolated vertices. The endpoints of the paths are vertices with degree 1. So the total number of vertices with degree 1 is even, and the overall number of vertices with degree 1 and 0 is at most $4(|D_2|+|P_2|)$. We adjoin the edges of a traveling salesman tour through the points of degree 1 and 0 and a minimal perfect matching of the points with degree 1 to the graph, see Figure 3.15. The length of the tour is at most $C(|D_2|+|P_2|)^{(d-1)/d}$ with a constant C. The same holds for the matching, see Lemma 2.1.

Adding the TSP tour and the matching yields a connected graph. Every vertex has an even degree so that there exists a Eulerian tour. We shortcut the Eulerian tour into a traveling salesman tour. If the tour contains a depot, it is a feasible cycle, otherwise we delete an edge and connect the remaining path to the boundary. Together with the unmodified cycles in G we get a feasible tour for D_1 and P_1, Figure 3.16, so it follows that

$$L^B(D_1, P_1) \leq L^B(D_1 \cup D_2, P_1 \cup P_2) + C(|D_2| + |P_2|)^{(d-1)/d}.$$

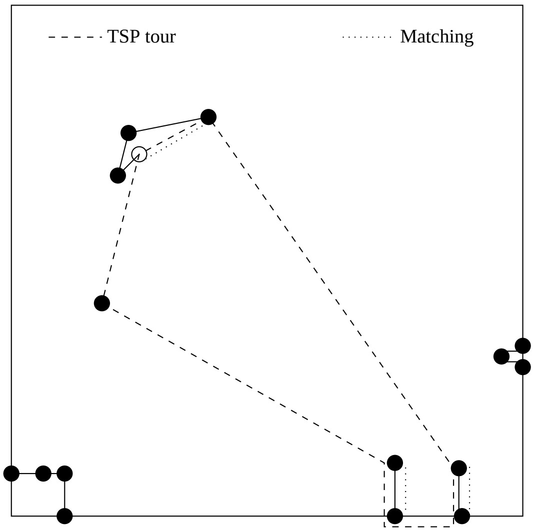

Figure 3.15: Boundary MDVRP tour, D_2 and P_2 deleted and the TSP tour and matching added.

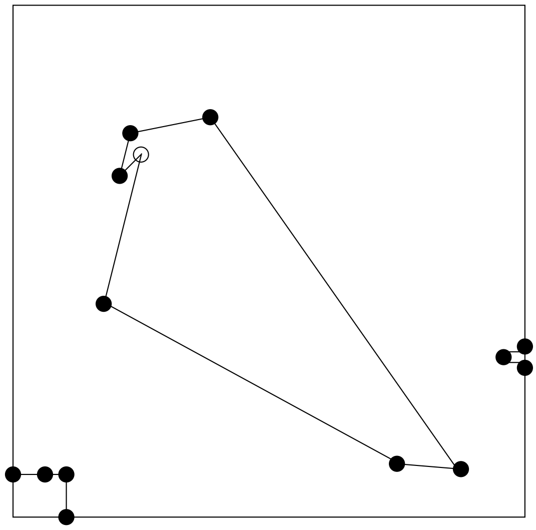

Figure 3.16: The final tour.

□

Chapter 4

Analysis of the Multi Depot Vehicle Routing Problem

We consider the asymptotic behavior of the multi depot vehicle routing functional on random instances in $[0,1]^d$, where the depots D and the points P are given by independent uniformly distributed random variables. Intuitively, we expect each of the k tours to occupy (roughly) a cell with sidelength $k^{-1/d}$ of the d-dimensional grid. Hence the TSP result applied to the (roughly) n/k points in each tour yields k times $(n/k)^{(d-1)/d}/k^{1/d}$, which is $n^{(d-1)/d}$. It is evident that the asymptotic behavior of $L(D,P)$ depends on the point-depot ratio n/k. All in all, we study three cases. In this chapter we consider the cases $k = o(n)$ and $k = \lambda n + o(n)$ for a constant $\lambda > 0$, in Chapter 5 we analyze the case $k = \Omega(n^{1+\varepsilon})$ for $\varepsilon > 0$. Our main result of this chapter covers the first two cases:

Theorem 4.1. *Let $D = \{D_1, \ldots, D_k\}$ and $P = \{P_1, \ldots, P_n\}$ be depots and points in $[0,1]^d$ given by independent uniformly distributed random variables. The optimal length $L(D,P)$ of an MDVRP tour through D and P satisfies*

(i) $\lim_{n\to\infty} \frac{L(D,P)}{n^{(d-1)/d}} = \alpha(L_{TSP}, d)$ *c.c., if* $k = o(n)$,

(ii) $\lim_{n\to\infty} \frac{L(D,P)}{n^{(d-1)/d}} = \alpha(L_{MDVRP}, \lambda, d)$ *c.c., if* $k = \lambda n + o(n)$ *for a constant* $\lambda > 0$,

where $\alpha(L_{TSP}, d)$ is the constant for the TSP and $\alpha(L_{MDVRP}, \lambda, d)$ is a positive constant.

Note that the constant $\alpha(L_{MDVRP}, \lambda, d)$ does not depend on the $o(n)$ term, it depends only on λ and the dimension d.

The theorem is proven in the following two sections. First we show in Section 4.1 that the behavior of the MDVRP and the TSP is the same up to the limiting constant for $k = o(n)$. In Section 4.2 we consider the case $k = \lambda n + o(n)$, $\lambda > 0$. The superadditive structure of the boundary MDVRP functional is used to show that the mean of the MDVRP has the same $n^{(d-1)/d}$ asymptotics like most Euclidean functionals, e.g. TSP, MST and so on. By the concentration inequality of Rhee, Theorem 2.4, we have that the boundary functional has the same asymptotics as its mean. Since the boundary functional L^B approximates the original functional L, the asymptotics of the MDVRP are determined for $k = \lambda n + o(n)$, $\lambda > 0$. Naturally, the value of the limiting constant depends on the value of λ. In Section 4.2.1 we give upper and lower bounds for the MDVRP constant. The values for the limiting constants of most problems are unknown, so we can only compare the values with bounds for constants of other problems.

4.1 The case $k = o(n)$

For $k = o(n)$ the MDVRP behaves asymptotically exactly like the TSP, even the constant is the same. The reason is that the number of depots is too small to affect the asymptotics. In the following lemma we show that the lengths of optimal solutions for the MDVRP and the TSP are very close if $k = o(n)$, afterwards we prove that the multi depot tour length divided by $n^{(d-1)/d}$ converges completely to the same constant as the TSP tour length divided by $n^{(d-1)/d}$.

Let $L_{TSP}(P)$ denote the length of an optimal TSP tour through a point set $P \subset [0,1]^d$.

Lemma 4.2. *Let $D, P \subset [0,1]^d$. If $|D| = o(|P|)$, then*

$$|L_{TSP}(P) - L(D,P)| = o(|P|^{(d-1)/d}).$$

Proof. Let D, P be finite point sets in $[0,1]^d$ with $|D| = o(|P|)$. We have $L(D,P) \leq L_{TSP}(P) + 2\sqrt{d}$, because we get a feasible multi depot tour by inserting a depot into an optimal TSP tour. It remains to show that $L_{TSP}(P) - L(D,P) = O(|D|^{(d-1)/d})$. We are given an optimal multi depot tour of length $L(D,P)$, we add a traveling salesman tour through D of length $O(|D|^{(d-1)/d})$, see the growth-bound Lemma 2.1. The resulting graph is connected and all vertices have even degree, so there is a Eulerian tour. In the tour we remove all depots by short-cuts, and this procedure yields a traveling salesman tour through P of length at most $L(D,P) + O(|D|^{(d-1)/d})$. □

So the influence of the depots on the length of an optimal MDVRP tour is small in this case and we have Theorem 4.1 (i) straightaway:

Proof of Theorem 4.1 (i). Let $|P| = n$ and $|D| = o(n)$. We show that $\frac{L(D,P)}{n^{(d-1)/d}}$ converges completely to $\alpha(L_{TSP}, d)$. Let $\varepsilon > 0$,

$$\begin{aligned}
&\sum_{n=1}^{\infty} \mathbb{P}\left[\left|\frac{L(D,P)}{n^{(d-1)/d}} - \alpha(L_{TSP}, d)\right| > \varepsilon\right] \\
&= \sum_{n=1}^{\infty} \mathbb{P}\left[\left|\frac{L(D,P) - L_{TSP}(P) + L_{TSP}(P)}{n^{(d-1)/d}} - \alpha(L_{TSP}, d)\right| > \varepsilon\right] \\
&\leq \sum_{n=1}^{\infty} \mathbb{P}\left[\left|\frac{L(D,P) - L_{TSP}(P)}{n^{(d-1)/d}}\right| + \left|\frac{L_{TSP}(P)}{n^{(d-1)/d}} - \alpha(L_{TSP}, d)\right| > \varepsilon\right] \\
&\leq \sum_{n=1}^{\infty} \mathbb{P}\left[\left|\frac{L(D,P) - L_{TSP}(P)}{n^{(d-1)/d}}\right| > \frac{\varepsilon}{2}\right] + \mathbb{P}\left[\left|\frac{L_{TSP}(P)}{n^{(d-1)/d}} - \alpha(L_{TSP}, d)\right| > \frac{\varepsilon}{2}\right].
\end{aligned}$$

With Lemma 4.2 we have $|L_{TSP}(P) - L(D,P)| = o(n^{(d-1)/d})$, so

$$\sum_{n=1}^{\infty} \mathbb{P}\left[\left|\frac{L(D,P) - L_{TSP}(P)}{n^{(d-1)/d}}\right| > \frac{\varepsilon}{2}\right] < \infty.$$

The length of the TSP functional converges completely to $\alpha(L_{TSP}, d)$, refer Rhee [Rhe93],

$$\sum_{n=1}^{\infty} \mathbb{P}\left[\left|\frac{L_{TSP}(P)}{n^{(d-1)/d}} - \alpha(L_{TSP}, d)\right| > \frac{\varepsilon}{2}\right] < \infty,$$

thus, we have complete convergence of the MDVRP functional in the case $k = o(n)$:

$$\sum_{n=1}^{\infty} \mathbb{P}\left[\left|\frac{L(D,P)}{n^{(d-1)/d}} - \alpha(L_{TSP}, d)\right| > \varepsilon\right] < \infty.$$

□

4.2 The case $k = \lambda n + o(n)$ for a constant $\lambda > 0$

In this case the depots influence the asymptotic tour length. We show that the asymptotic multi depot tour length divided by $n^{(d-1)/d}$ is $\alpha(L_{MDVRP}, \lambda, d)$, where the constant $\alpha(L_{MDVRP}, \lambda, d)$ depends on the value of λ and the dimension d. First we consider the case $k = \lambda n$ and after that we show that adding $o(n)$ depots does not change the asymptotics.

The proof of the asymptotic behavior is based on the multi depot tour boundary functional: the superadditive and smooth structure of the boundary functional is used to determine the asymptotics of its expectation in Lemma 4.4. In Lemma 4.3

we apply Theorem 2.4 directly to show that the asymptotic tour length of the boundary functional has the same behavior as its expectation. Since the multi depot tour functional and its boundary functional are pointwise close, Lemma 3.4, they have the same asymptotic behavior. Therefore, it is sufficient to show Theorem 4.1 (ii) for the expectation of L^B.

Lemma 4.3. *If*

$$\lim_{n\to\infty} \frac{\mathbb{E}\left[L^B(D_1 \ldots, D_{\lambda n}, P_1 \ldots, P_n)\right]}{n^{(d-1)/d}} = \alpha,$$

then

$$\lim_{n\to\infty} \frac{L^B(D_1 \ldots, D_{\lambda n}, P_1 \ldots, P_n)}{n^{(d-1)/d}} = \alpha \ c.c.$$

Proof. Suppose that $\lim_{n\to\infty} \mathbb{E}\left[L^B(D_1 \ldots, D_{\lambda n}, P_1 \ldots, P_n)\right] / n^{(d-1)/d} = \alpha$. By Theorem 2.4 for all $\varepsilon > 0$,

$$\begin{aligned}
&\sum_{n=1}^{\infty} \mathbb{P}\left[\left|\frac{L^B(D_1,\ldots,D_{\lambda n},P_1,\ldots,P_n) - \mathbb{E}\left[L^B(D_1,\ldots,D_{\lambda n},P_1,\ldots,P_n)\right]}{n^{(d-1)/d}}\right| > \varepsilon\right] \\
&\quad = \sum_{n=1}^{\infty} \mathbb{P}\left[\left|L^B(D_1,\ldots,D_{\lambda n},P_1,\ldots,P_n)\right.\right. \\
&\qquad\qquad \left.\left. - \mathbb{E}\left[L^B(D_1,\ldots,D_{\lambda n},P_1,\ldots,P_n)\right]\right| > \varepsilon n^{(d-1)/d}\right] \\
&\quad \leq C \sum_{n=1}^{\infty} \exp\left(-\left(\frac{\varepsilon}{C'}\right)^{2d/(d-1)} \frac{n}{C}\right) < \infty,
\end{aligned}$$

for constants $C, C' > 0$. Thus, $\left|\frac{L^B(D_1,\ldots,D_{\lambda n},P_1,\ldots,P_n) - \mathbb{E}\left[L^B(D_1,\ldots,D_{\lambda n},P_1,\ldots,P_n)\right]}{n^{(d-1)/d}}\right|$ converges completely to zero. ☐

The next lemma is the main lemma of this chapter, we show convergence of the expectation of the boundary functional in case $k = \lambda n$. The lemma together with the previous lemma yields the proof of Theorem 4.1 (ii) in case $k = \lambda n$.

The asymptotic behavior of the expectation of L^B is shown via typical methods, they are used since the seminal paper of Beardwood, Halton and Hammersley [BHH59] and exploit the superadditive and smooth structure of the functional, refer to Chapter 3 in [Yuk98] or Chapter 3 in [Ste97].

Lemma 4.4. *Let $k = \lambda n$, $\lambda > 0$ constant. Then there exists a constant $\alpha(L_{MDVRP}, \lambda, d) > 0$ such that*

$$\lim_{n\to\infty} \frac{\mathbb{E}\left[L^B(D_1 \ldots, D_{\lambda n}, P_1 \ldots, P_n)\right]}{n^{(d-1)/d}} = \alpha(L_{MDVRP}, \lambda, d).$$

Proof. Set $\Phi(k,n) := \mathbb{E}\left[L^B(D_1,\dots,D_k,P_1,\dots,P_n)\right]$. The number of depots respectively points that fall into a given subcube of $[0,1]^d$ of volume m^{-d} are given by binomial random variables X respectively Y with distribution $B(k,m^{-d})$ and $B(n,m^{-d})$. Set $\Phi(X,Y) := \mathbb{E}\left[L^B(D_1,\dots,D_X,P_1,\dots,P_Y)\right]$. We divide $[0,1]^d$ into m^d identical cubes R_i of volume m^{-d}. With the superadditivity we have that the global tour length in $[0,1]^d$ is at least as large as the sum of the tour lengths in the m^d cubes R_i:

$$\Phi(k,n) \geq \sum_{i=1}^{m^d} \mathbb{E}\left[L^B(\{D_1,\dots,D_k\}\cap R_i,\{P_1,\dots,P_n\}\cap R_i)\right],$$

We scale the R_i and the contained tours by a factor m up to the unit cube $[0,1]^d$ and exploit the homogeneity of L^B,

$$\Phi(k,n) \geq \frac{1}{m}\sum_{i=1}^{m^d}\Phi(X,Y).$$

Via the smoothness of L^B and Jensen's inequality one can show that

$$\begin{aligned}
&\Phi(k,n)\\
&\geq \frac{1}{m}\sum_{i=1}^{m^d}\Big[\Phi(km^{-d},nm^{-d}) - C\,\mathbb{E}\left[|X-km^{-d}|\right]^{(d-1)/d}\\
&\quad -C\,\mathbb{E}\left[|Y-nm^{-d}|\right]^{(d-1)/d}\Big]\\
&\geq \frac{1}{m}\sum_{i=1}^{m^d}\Big[\Phi(km^{-d},nm^{-d}) - C(\mathbb{E}\left[X-km^{-d}\right]^2)^{(d-1)/2d}\\
&\quad -C(\mathbb{E}\left[Y-nm^{-d}\right]^2)^{(d-1)/2d}\Big]\\
&\geq \frac{1}{m}\sum_{i=1}^{m^d}\left[\Phi(km^{-d},nm^{-d}) - C(km^{-d}-km^{-2d})^{(d-1)/2d}\right.\\
&\quad \left.-C(nm^{-d}-nm^{-2d})^{(d-1)/2d}\right]\\
&\geq \frac{1}{m}\sum_{i=1}^{m^d}\left[\Phi(km^{-d},nm^{-d}) - C(km^{-d})^{(d-1)/2d}\right.\\
&\quad \left.-C(nm^{-d})^{(d-1)/2d}\right]\\
&\geq m^{d-1}\Phi(km^{-d},nm^{-d}) - Ck^{(d-1)/2d}m^{(d-1)/2} - Cn^{(d-1)/2d}m^{(d-1)/2}.
\end{aligned}$$

Dividing by $k^{(d-1)/2d}n^{(d-1)/2d}$, we get

$$\frac{\Phi(k,n)}{k^{(d-1)/2d}n^{(d-1)/2d}} \geq m^{d-1}\frac{\Phi(km^{-d},nm^{-d})}{k^{(d-1)/2d}n^{(d-1)/2d}} - \frac{Cm^{(d-1)/2}}{n^{(d-1)/2d}} - \frac{Cm^{(d-1)/2}}{k^{(d-1)/2d}}.$$

Now, n is replaced by nm^d and k by km^d:

$$\frac{\Phi(km^d, nm^d)}{(km^d)^{(d-1)/2d}(nm^d)^{(d-1)/2d}} \geq \frac{\Phi(k,n)}{k^{(d-1)/2d}n^{(d-1)/2d}} - \frac{Cm^{(d-1)/2}}{n^{(d-1)/2d}m^{(d-1)/2}} - \frac{Cm^{(d-1)/2}}{k^{(d-1)/2d}m^{(d-1)/2}}.$$

Simplified,

$$\frac{\Phi(km^d, nm^d)}{(km^d)^{(d-1)/2d}(nm^d)^{(d-1)/2d}} \geq \frac{\Phi(k,n)}{k^{(d-1)/2d}n^{(d-1)/2d}} - \frac{C}{n^{(d-1)/2d}} - \frac{C}{k^{(d-1)/2d}}.$$

For the case $k = \lambda n$, this yields

$$\frac{\Phi(\lambda nm^d, nm^d)}{\lambda^{(d-1)/2d}(nm^d)^{(d-1)/d}} \geq \frac{\Phi(\lambda n, n)}{\lambda^{(d-1)/2d}n^{(d-1)/d}} - \frac{C}{n^{(d-1)/2d}} - \frac{C}{\lambda^{(d-1)/2d}n^{(d-1)/2d}}.$$

Therefore we have

$$\frac{\Phi(\lambda nm^d, nm^d)}{(nm^d)^{(d-1)/d}} \geq \frac{\Phi(\lambda n, n)}{n^{(d-1)/d}} - \frac{C\lambda^{(d-1)/2d}}{n^{(d-1)/2d}} - \frac{C}{n^{(d-1)/2d}}.$$

Set $\alpha(L_{MDVRP}, \lambda, d) := \alpha := \limsup_{n\to\infty} \frac{\Phi(\lambda n,n)}{n^{(d-1)/d}}$. The smoothness of L^B guarantees $\alpha < \infty$. For all $\varepsilon > 0$, choose n_0 such that for all $n \geq n_0$ we have $\frac{C}{n^{(d-1)/2d}} < \varepsilon$, $\frac{C\lambda^{(d-1)/2d}}{n^{(d-1)/2d}} < \varepsilon$ and $\frac{\Phi(\lambda n_0,n_0)}{n_0^{(d-1)/d}} > \alpha - \varepsilon$. Thus, for all $m \geq 1$ it follows that

$$\frac{\Phi(\lambda n_0 m^d, n_0 m^d)}{(n_0 m^d)^{(d-1)/d}} > \alpha - 3\varepsilon.$$

Now we use an interpolation argument and the smoothness of the functional. For an arbitrary integer t we choose m as the unique integer m such that

$$n_0 m^d < t \leq n_0(m+1)^d.$$

Then $|n_0 m^d - t| \leq C' n_0 m^{d-1}$ and by smoothness we obtain

$$\begin{aligned} \frac{\Phi(\lambda t, t)}{t^{(d-1)/d}} &\geq \frac{\Phi(\lambda n_0 m^d, n_0 m^d)}{(n_0(m+1)^d)^{(d-1)/d}} - \frac{C'(n_0 m^{d-1})^{(d-1)/d}}{(n_0(m+1)^d)^{(d-1)/d}} - \frac{C'(\lambda n_0 m^{d-1})^{(d-1)/d}}{(n_0(m+1)^d)^{(d-1)/d}} \\ &\geq (\alpha - 3\varepsilon)\left(\frac{m}{m+1}\right)^{d-1} - \frac{C' m^{(d-1)^2/d}}{(m+1)^{d-1}} - \frac{C'(\lambda m)^{(d-1)^2/d}}{(m+1)^{d-1}}. \end{aligned}$$

Since the last two terms go to zero as m goes to infinity,

$$\liminf_{t\to\infty} \frac{\Phi(\lambda t,t)}{t^{(d-1)/d}} \geq \alpha - 3\varepsilon.$$

For ε tending to zero we see that the lim inf and the lim sup of the sequence $\frac{\Phi(\lambda t,t)}{t^{(d-1)/d}}, t \geq 1$, coincide, and we may define

$$\alpha(L_{MDVRP}, \lambda, d) := \lim_{t\to\infty} \frac{\Phi(\lambda t, t)}{t^{(d-1)/d}}.$$

It remains to show that $\alpha(L_{MDVRP}, \lambda, d) > 0$. For a set of independent random variables $\mathcal{X} := \{X_1, \ldots, X_{n+\lambda n}\}$ with uniform distribution in $[0,1]^d$, there is a $c > 0$ such that $\mathbb{E}[\min\{|X_i - X_j| : X_i, X_j \in \mathcal{X}\}] > \frac{c}{(n+\lambda n)^{1/d}}$. Since $\frac{c}{(n+\lambda n)^{1/d}} = \frac{c'}{n^{1/d}}$ for a $c' > 0$, and a depot tour through n points contains at least $n+1$ edges, we have $\Phi(\lambda n, n) > c' n^{(d-1)/d}$. Consequently, $\alpha(L_{MDVRP}, \lambda, d)$ is positive. □

Note that the proof is not valid in the case $k = n^{1+\varepsilon}$ for $\varepsilon > 0$. In this case, the graph looses its self-similarity: for example, dividing $[0,1]^2$ into four subcubes, we have $\frac{n}{4}$ points and $\frac{k}{4}$ depots in each cube, but if e.g. $k = n^2$, we have $\frac{n}{4}$ points and $\frac{n^2}{4}$ depots instead of $(\frac{n}{4})^2$ depots in each cube. So we do not have small instances of the original type in the subcubes.

It remains to show that adding $o(n)$ depots does not change the asymptotics of the multi depot tour functional and this completes the proof of Theorem 4.1 (ii).

Corollary 4.5. *Let $D = \{D_1, \ldots, D_k\}$ and $P = \{P_1, \ldots, P_n\}$ be depots and points in $[0,1]^d$ given by independent uniformly distributed random variables. The optimal length $L(D,P)$ of an MDVRP tour through D and P satisfies*

$$\lim_{n\to\infty} \frac{L(D,P)}{n^{(d-1)/d}} = \alpha(L_{MDVRP}, \lambda, d) \ c.c.$$

, if $k = \lambda n + o(n)$ for a constant $\lambda > 0$ and where $\alpha(L_{MDVRP}, \lambda, d)$ is the constant from Lemma 4.4.

Proof. Let P be a set of n points, D a set of λn depots and D' a set of $o(n)$ depots. First we prove $|L(D,P) - L(D \cup D', P)| = o(n^{(d-1)/d})$. We begin with a MDVRP tour on $D \cup D'$ and P and delete all depots of D'. In the remaining graph we have at most $2|D'|$ vertices with degree 1 and 0. We denote the set of vertices with degree 1 by B. The cardinality of B is even, since the vertices with degree 1 are endpoints of paths. We adjoin the edges of a perfect minimal matching of B and a TSP tour on all vertices with degree 1 and 0 to the graph. All vertices in the resulting connected component without a depot have even degree so there exists an Eulerian tour. We turn the Eulerian tour into a TSP tour by shortcuts. The resulting cycle is connected to a depot of D or merged into a cycle with a depot of D. By Lemma 2.1 the total length of all added edges is at most $C|D'|^{(d-1)/d}$. Since $|D'| = o(n)$ and $L(D \cup D', P) \leq L(D, P)$, we have $|L(D, P) - L(D \cup D', P)| = o(n^{(d-1)/d})$.

By Lemma 4.4 and Lemma 4.3 we have $\lim_{n\to\infty} \frac{L(D,P)}{n^{(d-1)/d}} = \alpha(L_{MDVRP}, \lambda, d)$ c.c. Analogous to the proof of Theorem 4.1 (i), we can show that $\lim_{n\to\infty} \frac{L(D\cup D',P)}{n^{(d-1)/d}} = \alpha(L_{MDVRP}, \lambda, d)$ c.c. □

4.2.1 Bounds for $\alpha(L_{MDVRP}, \lambda, d)$ for $k = \lambda n$

The exact values for the limiting constants of optimization problems such as the traveling salesman problem, minimal spanning tree, minimal matching, etc. are unknown. The best bounds for the TSP constant are given by Beardwood, Halton and Hammersley in [BHH59]. For $d = 2$, they show $0.62 < \alpha(L_{TSP}, d) < 0.93$. For the minimal matching, Papadimitriou [Pap78] showed that for $d = 2$, the constant is in the interval $[0.25, 0.40]$.

We consider bounds for the MDVRP constant $\alpha(L_{MDVRP}, \lambda, d)$ for $k = \lambda n$. It is expected that the constant decreases with increasing λ. We give upper and lower bounds that decrease as $(1 + \lambda)^{-1/d}$ in the following lemma. For the analysis we need the following notions. Let $B(x, r)$ denote the ball around x with radius r and let $v_d := \frac{\pi^{d/2}}{\Gamma(\frac{d}{2}+1)}$. The volume of the d-dimensional ball with radius r is $r^d v_d$.

Theorem 4.6. *In the case $k = \lambda n$, $\lambda > 0$, the bounds for the limiting constant $\alpha(L_{MDVRP}, \lambda, d)$ for the multi depot tour functional are:*

(i) $\min\left\{\alpha(L_{TSP}, d), \frac{2\alpha(L_{TSP},d)}{(1+\lambda)^{1/d}}\right\} \geq \alpha(L_{MDVRP}, \lambda, d)$

(ii) $\frac{\Gamma(\frac{1}{d})\Gamma(\frac{d}{2}+1)}{\pi^{1/2}d(1+\lambda)^{1/d}}\left(1 + \frac{1}{2d(1+\lambda)}\right) \leq \alpha(L_{MDVRP}, \lambda, d)$.

Proof. For the lower bound we consider an arbitrary point $p \in P$. Let $O = P \setminus \{p\} \cup D$. We consider the following equivalent random setting. We have $n + k - 1$ objects given by random variables with uniform distribution in $[0, 1]^d$. After the objects are placed, we choose a random subset of k elements which we consider as depots. Let δ_i denote the distance from p to the i-nearest object of O. In an optimal tour p is either connected with two edges to a depot or it is connected to two different objects in O, two points respectively a point and a depot. We bound the lengths of the edges connecting p in an optimal tour: the length of one edge is bounded from below by δ_1. The length of the second edge is bounded by δ_1, too, if the nearest neighbor is a depot or by δ_2 if the nearest neighbor is a point. Let $z_1 = \delta_1$ and $z_2 = \delta_1$ if the nearest object is a depot and

$z_2 = \delta_2$ if the nearest object is a point. So z_1 is a bound for the shorter and z_2 for the longer edge connecting p in an optimal tour. Thus, $\mathbb{P}[z_2 = \delta_1] = \frac{k}{n+k-1}$ and $\mathbb{P}[z_2 = \delta_2] = \frac{n-1}{n+k-1}$.

We have $2\,\mathbb{E}[L(D,P)] \geq n\,\mathbb{E}[z_1 + z_2]$. To determine the probability that the nearest neighbor has a distance of at least r, we consider the probability that there is no point in a circle around p with radius r. Then,

$$\begin{aligned}
&\mathbb{E}[L(D,P)] \geq \frac{n}{2}\,\mathbb{E}[z_1 + z_2] \\
= \;& \frac{n}{2}\left(\mathbb{E}(\delta_1) + \frac{k}{n+k-1}\,\mathbb{E}(\delta_1) + \frac{n-1}{n+k-1}\,\mathbb{E}(\delta_2)\right) \\
\geq \;& \frac{n}{2}\left(\left(1 + \frac{k}{n+k-1}\right)\int_0^\infty \mathbb{P}[\delta_1 > r]dr + \frac{n-1}{n+k-1}\int_0^\infty \mathbb{P}[\delta_2 > r]dr\right) \\
\geq \;& \frac{n}{2}\left(\left(1 + \frac{k}{n+k-1}\right)\int_0^{v_d^{-\frac{1}{d}}} (1 - v_d r^d)^{n+k-1} dr\right. \\
& \left. + \frac{n-1}{n+k-1}\int_0^{v_d^{-\frac{1}{d}}} (1 - v_d r^d)^{n+k-1} + (n+k-1) v_d r^d (1 - v_d r^d)^{n+k-2} dr\right) \\
= \;& \frac{n}{2}\left(2\int_0^{v_d^{-\frac{1}{d}}} (1 - v_d r^d)^{n+k-1} dr + (n-1)\int_0^{v_d^{-\frac{1}{d}}} v_d r^d (1 - v_d r^d)^{n+k-2} dr\right).
\end{aligned}$$

We substitute $z = v_d r^d$,

$$\begin{aligned}
&\mathbb{E}[L(D,P)] \\
&\geq \frac{n}{2}\left(2\int_0^1 \frac{(1-z)^{n+k-1}}{d v_d^{1/d} z^{(d-1)/d}} dz + (n-1)\int_0^1 \frac{z(1-z)^{n+k-2}}{d v_d^{1/d} z^{(d-1)/d}} dz\right) \\
&\geq \frac{n}{2 d v_d^{1/d}}\left(2\int_0^1 z^{-(d-1)/d}(1-z)^{n+k-1} dz + (n-1)\int_0^1 z^{\frac{1}{d}}(1-z)^{n+k-2} dz\right).
\end{aligned}$$

Using $\int_0^1 t^{x-1}(1-t)^{y-1} dt = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}$ and $x\Gamma(x) = \Gamma(x+1)$, one can calculate

$$\begin{aligned}
\mathbb{E}[L(D,P)] &\geq \frac{n}{2 d v_d^{1/d}}\left(2\frac{\Gamma(n+k)\Gamma(\frac{1}{d})}{\Gamma(n+k+\frac{1}{d})} + (n-1)\frac{\Gamma(n+k-1)\Gamma(1+\frac{1}{d})}{\Gamma(n+k+\frac{1}{d})}\right) \\
&= \frac{n}{d v_d^{1/d}}\left(\frac{\Gamma(n+k)\Gamma(\frac{1}{d})}{\Gamma(n+k+\frac{1}{d})} + \frac{n-1}{2d(n+k-1)}\frac{\Gamma(n+k)\Gamma(\frac{1}{d})}{\Gamma(n+k+\frac{1}{d})}\right) \\
&\geq \frac{n}{d v_d^{1/d}}\frac{\Gamma(n+k)\Gamma(\frac{1}{d})}{\Gamma(n+k+\frac{1}{d})}\left(1 + \frac{n-1}{2d(n+k-1)}\right).
\end{aligned}$$

Now consider $a_l = \frac{\Gamma(l) l^{1/d}}{\Gamma(l+\frac{1}{2})}$. Since $\frac{a_l}{a_{l+1}} = \left(1+\frac{1}{l}\right)^{-\frac{1}{d}}\left(1+\frac{1}{dl}\right) \geq 1$ and $a_l \to 1$ for $l \to \infty$, we get the following inequality with $l = n + k$.

$$\mathbb{E}[z_1 + z_2] \geq \frac{\Gamma(\frac{1}{d})}{dv_d^{1/d}} \frac{n}{(n+k)^{1/d}} \left(1 + \frac{n-1}{2d(n+k-1)}\right).$$

Thus, for $k = \lambda n$, we have

$$\lim_{n\to\infty} \frac{\mathbb{E}[L(D,P)]}{n^{(d-1)/d}} \geq \frac{\Gamma(\frac{1}{d})}{dv_d^{1/d}(1+\lambda)^{1/d}} \left(1 + \frac{1}{2d(1+\lambda)}\right).$$

For the upper bound, we consider a TSP tour through all points and depots. We give the tour an arbitrary direction, start with an arbitrary point and follow the tour, duplicating all edges starting at a point and deleting all edges starting at a depot. In this procedure we delete k random edges of the tour, considering the model where $n + k$ objects are placed and a random subset with k elements is chosen afterwards to determine the depots. So we use a fraction of $\frac{n}{n+k}$ of the $(n+k)$ edges of the TSP tour and the expected length of the used edges is $\frac{n}{n+k}$ times the length of the TSP tour. Each of the resulting connected components is turned into a feasible depot tour by shortcuts. Thus, for the length T of the constructed depot tour we have $\frac{\mathbb{E}[T]}{n^{(d-1)/d}} \leq \frac{2\,\mathbb{E}[\text{TSP tour}](n+k)^{(d-1)/d} n}{(n+k)^{(d-1)/d} n^{(d-1)/d}(n+k)} = \frac{2\,\mathbb{E}[\text{TSP tour}]}{(n+k)^{(d-1)/d}(1+\lambda)^{1/d}}$.

So we get

$$\limsup_{n\to\infty} \frac{\mathbb{E}[T]}{n^{(d-1)/d}} \leq \frac{2\alpha(L_{TSP}, d)}{(1+\lambda)^{1/d}}.$$

□

In Figure 4.1 we have a plot of the bounds for $d = 2$ that shows the behavior of the bounds for increasing λ. These are the first results for the MDVRP constants, so we can only compare the values with the bounds for other Euclidean optimization problems. We choose the TSP and the minimal matching due to their close relation to the MDVRP. The relation to the TSP is obvious, and in the case $k \geq n^{1+\varepsilon}$ a lot of points are connected by two edges to their nearest depot, so the graph is similar to a matching graph. This is analyzed in the next section. As noted in the introduction of this section, the lower bound of the TSP constant is 0.62 and the upper 0.93, the Matching constant is in $[0.25, 0.40106]$. So for $0 \leq \lambda \leq 8$ the MDVRP upper bound is in the range of the TSP constant, and for $20.6 \leq \lambda 54.4$ it is in the range of the minimal matching constant. The lower bound of $\alpha(L_{MDVRP}, \lambda, d)$ is smaller than the lower bound of the minimal matching constant for $\lambda \geq 4.1$. In comparison to the minimal matching, a point in the MDVRP may be connected to a depot or other points, so there are λn additional possible neighbors for each point compared to the minimal matching with n points. This might indicate that there is some space to improve our upper bound.

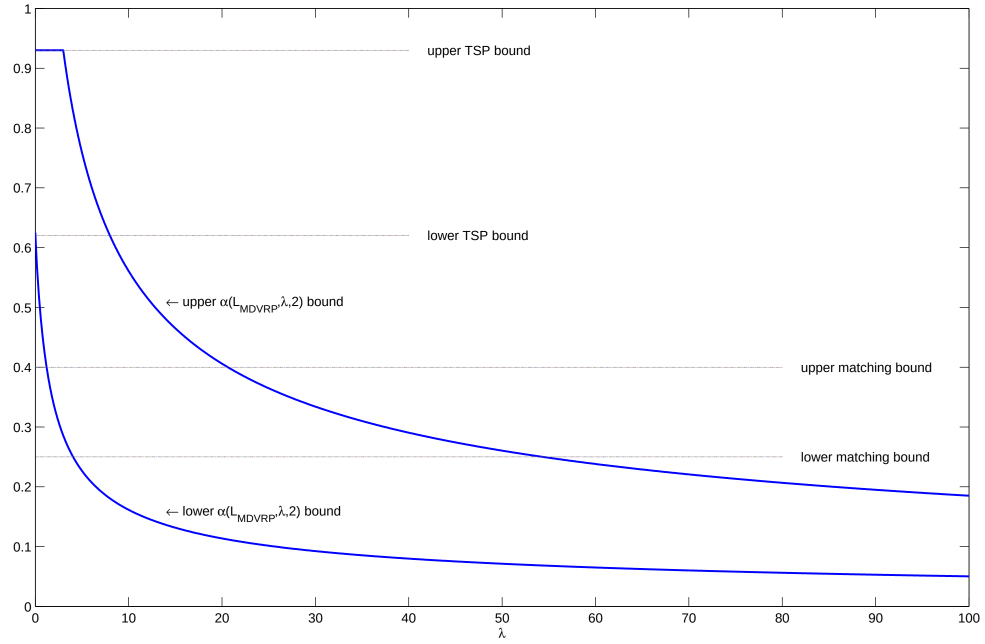

Figure 4.1: Plot of the bounds for increasing λ (for $d = 2$).

Chapter 5

The All Nearest Neighbor Problem and the MDVRP in case $k \geq n^{1+\varepsilon}$

There is a large variety of nearest neighbor problems, a detailed overview is given by Smid [S00], in a chapter of the 'Handbook on Computational Geometry' edited by Sack and Urrutia. An elementary construction in computational geometry is the k nearest neighbor graph: we are given a set of n points, and we have to connect each points to its k nearest neighbors. The length function of this problem fits directly into the theory of smooth subadditive Euclidean functionals, and McGivney [M97] showed the asymptotic behavior for the total graph length $L(k;n)$.

Theorem 5.1. *[M97] Let $X_1, \ldots, X_n$ be iid random variables with values in $[0,1]^d$, $d \geq 2$. Then*

$$\lim_{n\to\infty} \frac{L(k;n)}{n^{(d-1)/d}} = \alpha(k,d) \int_{[0,1]^d} f(x)^{(d-1)/d} \ c.c.,$$

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

Agarwal et al. [AHR+92] introduced the multi-chromatic closest pair problem. Here, we are given a collection of point sets and each point p has an edge to its nearest neighbor that does not belong to p's set.

We consider a slight modification, the so-called all nearest neighbor problem (ANNP): we are given two point sets P and D, $|P| = n$ and $|D| = k$, in $[0,1]^d$, and each point of P has to be connected to its nearest neighbor in D. Note that an element of D may be connected to several elements of P or not a single one at all.

The total edge length is denoted by $L_{ANN}(D,P)$,

$$L_{ANN}(D,P) = \min_{\sigma} \sum_{i=1}^{n} \left\| p_i - d_{\sigma(p_i)} \right\|_2$$

where $p_i \in P$ and $d_i \in D$ and the minimum is taken over all mappings $\sigma : P \to D$.

This problem occurs frequently in several database applications. In Geographical Informations Systems example queries include 'find the nearest warehouse for each supermarket' or 'find the nearest parking lot for each subway station', so all nearest neighbor queries are common in urban planning and resource allocation problems. In Data Analysis all nearest neighbor queries have been considered as a core module of clustering [JMF99] and outlier detection [AY01]. In Computer Architecture/VLSI design, the operability and speed of very large circuits depends on the relative distance between the various components. The ANNP is solved to detect abnormalities and guide relocation of components [NO97].

In this chapter we analyze the asymptotic behavior of the expectation of L_{ANN}, we prove the following theorem in Section 5.1.

Theorem 5.2. *Let $P = \{X_1, \ldots, X_n\}$ and $D = \{X_{n+1}, \ldots, X_{n+k}\}$ be sets of independent random variables with uniform distribution and values in $[0,1]^d$. Then*

$$\lim_{k\to\infty} \frac{\mathbb{E}[L_{ANN}(D,P)]}{nk^{-1/d}} = \frac{\Gamma(\frac{1}{d})\Gamma(\frac{d}{2}+1)^{1/d}}{d\sqrt{\pi}}.$$

Unfortunately, we can not apply the concentration inequality of Theorem 2.4 in order to show the behavior of the functional, because the ANNP functional is obviously not smooth. Furthermore the functional does not satisfy an appropriate smoothness so that the theorem may be modified to meet our needs, as far as we know.

We use Theorem 5.2 for the analysis of the MDVRP in the case $k \geq n^{1+\varepsilon}$ for $\varepsilon > 0$. The connection between the MDVRP and the ANNP is shown in Section 5.2. We will see that in the case $k \geq n^{1+\varepsilon}$ for $\varepsilon > 0$ almost all points are connected to the nearest depot with two edges. Thus, the analysis of the MDVRP is based on the analysis of the ANNP. Again, Theorem 2.4 can not be directly applied to show the asymptotics and complete convergence for the MDVRP functional. But the isoperimetric inequality used in the proof of Theorem 2.4 gives complete convergence for the MDVRP for the two dimensional case with $k = n^{1+\varepsilon}$ for $0 < \varepsilon < 1$. It remains an open problem to show the asymptotics of the functional for other cases. As our main result we prove the following Theorem 5.3 in Section 5.2 and 5.2.1

Theorem 5.3. *Let $P = \{X_1, \ldots, X_n\}$ and $D = \{X_{n+1}, \ldots, X_{n+k}\}$ be sets of independent random variables with uniform distribution and values in $[0,1]^d$. Then*

$$\lim_{n\to\infty} \frac{\mathbb{E}[L(D,P)]}{nk^{-1/d}} = \frac{2\Gamma(\frac{1}{d})\Gamma(\frac{d}{2}+1)^{1/d}}{d\sqrt{\pi}} \text{ if } k = \Omega(n^{1+\varepsilon}) \text{ for an arbitrary } \varepsilon > 0.$$

Let $d = 2$ and $k = n^{1+\varepsilon}$ for $0 < \varepsilon < 1$. Then

$$\lim_{n\to\infty} \frac{L(D,P)}{nk^{-1/2}} = 1 \text{ c.c.}$$

5.1 The All Nearest Neighbor Problem

In this section we prove Theorem 5.2. The proof is based on the following lemma, which gives the expected minimal distance between a point and its nearest depot.

Lemma 5.4. *Let P_j be a point and D be a point set in $[0,1]^d$, $|D| = k$, given by independent random variables with uniform distribution. Then*

$$\lim_{k\to\infty} k^{1/d}\, \mathbb{E}\left[\min_{1\le i\le k} \|P_j - D_i\|_2\right] = \frac{\Gamma(\frac{1}{d})\Gamma(\frac{d}{2}+1)^{1/d}}{d\sqrt{\pi}}.$$

Proof. For the upper bound we have to bound the probability that a depot D_i is not contained in a ball around P_j with radius r from below. So we have to consider the case that the ball around P_j is not entirely contained in $[0,1]^d$ and therefore we distinguish if P_j is in $Q := \left[\frac{1}{k^{1/2d}}, 1 - \frac{1}{k^{1/2d}}\right]^d$ or not. The volume of $B(P_j, r) \cap [0,1]^d$ is minimized if P_j coincides with one of the corners of $[0,1]^d$, and in that case $B(P_j, r)$ has volume $2^{-d} v_d r^d$. We have

$$\begin{aligned}
&\mathbb{E}\left[\min_{1\le i\le k} \|P_j - D_i\|_2\right] \\
&= \int_0^\infty \mathbb{P}\left[\min_{1\le i\le k} \|P_j - D_i\|_2 \ge r\right] dr \\
&= \int_0^\infty \mathbb{P}\left[\min_{1\le i\le k} \|P_j - D_i\|_2 \ge r \mid P_j \in Q\right] \cdot \mathbb{P}\left[P_j \in Q\right] dr \\
&\quad + \int_0^\infty \mathbb{P}\left[\min_{1\le i\le k} \|P_j - D_i\|_2 \ge r \mid P_j \notin Q\right] \cdot \mathbb{P}\left[P_j \notin Q\right] dr \\
&\le \int_0^{k^{-1/2d}} (1 - v_d r^d)^k dr + \int_{k^{-1/2d}}^{2v_d^{-1/d}} (1 - v_d r^d 2^{-d})^k dr \\
&\quad + \int_0^\infty \mathbb{P}\left[\min_{1\le i\le k} \|P_j - D_i\|_2 \ge r \mid P_j \notin Q\right] \frac{2^d}{k^{1/2d}} dr
\end{aligned}$$

$$
\begin{aligned}
&\le \int_0^{k^{-1/2d}} (1-v_d r^d)^k dr + \int_{k^{-1/2d}}^{2v_d^{-1/d}} (1-v_d r^d 2^{-d})^k dr \\
&\quad + \int_0^{2v_d^{-1/d}} (1-v_d r^d 2^{-d})^k \frac{2^d}{k^{1/2d}} dr \\
&\le \int_0^{\infty} \exp(-v_d r^d k) dr + \int_{k^{-1/2d}}^{2v_d^{-1/d}} \exp(-v_d 2^{-d} k^{1/2}) dr \\
&\quad + \int_0^{\infty} \exp(-v_d r^d k 2^{-d}) \frac{2^d}{k^{1/2d}} dr.
\end{aligned}
$$

Substituting $y = v_d r^d k$ in the first and $y = v_d r^d k 2^{-d}$ in the last integral yields,

$$
\begin{aligned}
&\mathbb{E}\left[\min_{1\le i\le k} \|P_j - D_i\|_2\right] \\
&\le \frac{1}{d(v_d k)^{1/d}} \int_0^{\infty} y^{\frac{1-d}{d}} \exp(-y) dy + \int_0^{2v_d^{-1/d}} \exp(-v_d 2^{-d} k^{1/2}) dr \\
&\quad + \frac{2^d}{k^{1/2d} d(v_d k 2^{-d})^{1/d}} \int_0^{\infty} y^{\frac{1-d}{d}} \exp(-y) dy.
\end{aligned}
$$

By definition of the Gamma function $\int_0^\infty t^{x-1} \exp(-t) dt = \Gamma(x)$,

$$
\begin{aligned}
&\mathbb{E}\left[\min_{1\le i\le k} \|P_j - D_i\|_2\right] \\
&\le \frac{\Gamma(\frac{1}{d})}{d(v_d k)^{1/d}} + 2v_d^{-1/d} \exp(-v_d 2^{-d} k^{1/2}) + \frac{2^d \Gamma(\frac{1}{d})}{k^{1/2d} d(v_d k 2^{-d})^{1/d}}
\end{aligned}
$$

We multiply the inequality with $k^{1/d}$ and let k tend to infinity and get

$$
\limsup_{k\to\infty} k^{1/d}\, \mathbb{E}\left[\min_{1\le i\le k} \|P_j - D_i\|_2\right] \le \frac{\Gamma(\frac{1}{d})}{d v_d^{1/d}} = \frac{\Gamma(\frac{1}{d})\Gamma(\frac{d}{2}+1)^{1/d}}{d\sqrt{\pi}}.
$$

We consider the lower bound: the probability that a depot D_i is not contained in a ball around P_j with radius r is $1 - \mathrm{Vol}(B(P_j, r) \cap [0,1]^d)$, so it is at least $1 - v_d r^d$. Using $1 - x \ge \exp(-(1 + \frac{1}{k^{1/4d}})x)$, $0 \le x \le \frac{2k^{1/4}}{k^{1/4}+1}$,

$$
\begin{aligned}
&\mathbb{E}\left[\min_{1\le i\le k} \|P_j - D_i\|_2\right] \\
&\ge \int_0^{k^{-1/2d}} (1 - v_d r^d)^k dr
\end{aligned}
$$

$$
\begin{aligned}
&\geq \int_0^{k^{-1/2d}} \exp\left(-v_d r^d k(1+\frac{1}{k^{1/4}})\right) dr \\
&\geq \int_0^{\infty} \exp\left(-v_d r^d k(1+\frac{1}{k^{1/4}})\right) dr - \int_{k^{-1/2d}}^{1} \exp\left(-v_d k^{1/2}\right) dr \\
&\quad - \int_1^{\infty} \exp(-v_d r k) dr.
\end{aligned}
$$

The substitution $y = v_d r^d k(1+\frac{1}{k^{1/4}})$ yields

$$
\begin{aligned}
&\mathbb{E}\left[\min_{1\leq i\leq k} \|P_j - D_i\|_2\right] \\
&\geq \frac{1}{d(v_d k(1+\frac{1}{k^{1/4}}))^{1/d}} \int_0^{\infty} y^{\frac{1-d}{d}} \exp(-y) dy - \int_0^1 \exp\left(-v_d k^{1/2}\right) dr \\
&\quad - \int_1^{\infty} \exp(-v_d r k) dr \\
&\geq \frac{\Gamma(\frac{1}{d})}{d(v_d k(1+\frac{1}{k^{1/4}}))^{1/d}} - \exp\left(-v_d k^{1/2}\right) - \frac{1}{v_d k \exp(v_d k)} \\
&= \frac{\Gamma(\frac{1}{d})}{d(v_d k(1+\frac{1}{k^{1/4}}))^{1/d}} \left(1 - \frac{d(v_d k(1+\frac{1}{k^{1/4}}))^{1/d}}{\Gamma(\frac{1}{d}) \exp(v_d k^{1/2})} - \frac{d(v_d k(1+\frac{1}{k^{1/4}}))^{1/d}}{\Gamma(\frac{1}{d}) v_d k \exp(v_d k)}\right).
\end{aligned}
$$

We multiply by $k^{1/d}$ and let k tend to infinity and get

$$
\liminf_{k\to\infty} k^{1/d}\, \mathbb{E}\left[\min_{1\leq i\leq k} \|P_j - D_i\|_2\right] \geq \frac{\Gamma(\frac{1}{d})}{d v_d^{1/d}} = \frac{\Gamma(\frac{1}{d})\Gamma(\frac{d}{2}+1)^{1/d}}{d\sqrt{\pi}}.
$$

□

Theorem 5.2 follows directly from Lemma 5.4 by linearity of expectation.

5.2 The MDVRP in case $k \geq n^{1+\varepsilon}$

In this section we show that the analysis of the MDVRP in case $k \geq n^{1+\varepsilon}$ reduces to the analysis of a nearest neighbor problem, where all but a negligible amount of points will be connected to the nearest depots via two edges. By applying the Lemma 5.4, we get the first part of Theorem 5.3:

Lemma 5.5. *Let $D = \{D_1, \ldots, D_k\}$ and $P = \{P_1, \ldots, P_n\}$ be sets of depots and points in $[0,1]^d$ given by independent uniformly distributed random variables. The* *expected value of the optimal length $L(D,P)$ of an MDVRP tour through D and P satisfies*

$$\lim_{n\to\infty} \frac{\mathbb{E}[L(D,P)]}{nk^{-1/d}} = \frac{2\Gamma(\frac{1}{d})\Gamma(\frac{d}{2}+1)^{1/d}}{d\sqrt{\pi}}, \text{ if } k=\Omega(n^{1+\varepsilon}) \text{ for an arbitrary } \varepsilon>0.$$

Proof. First we show that MDVRP reduces to a nearest neighbor problem: for the lower bound we convince ourselves that in an optimal depot tour (for $k = \Omega(n^{1+\varepsilon})$) all but at most $o(n)$ points are connected via two edges to distinct depots.

Let $\varepsilon > 0$ and $\varepsilon' < \frac{1}{d}\varepsilon$. For each $P_j \in P$, we have

$$\begin{aligned}
&\mathbb{P}\left[\min_{1\le i\le k}\|P_j-D_i\|_2 > \frac{1}{2}\min_{\substack{1\le i\le n\\ i\ne j}}\|P_j-P_i\|_2\right]\\
&= \mathbb{P}\left[\min_{1\le i\le k}\|P_j-D_i\|_2 > \frac{1}{2}\min_{\substack{1\le i\le n\\ i\ne j}}\|P_j-P_i\|_2 \text{ and } \min_{1\le i\le k}\|P_j-D_i\|_2 \le \frac{1}{n^{\frac{1}{d}+\varepsilon'}}\right]\\
&\quad +\mathbb{P}\left[\min_{1\le i\le k}\|P_j-D_i\|_2 > \frac{1}{2}\min_{\substack{1\le i\le n\\ i\ne j}}\|P_j-P_i\|_2 \text{ and } \min_{1\le i\le k}\|P_j-D_i\|_2 > \frac{1}{n^{\frac{1}{d}+\varepsilon'}}\right]\\
&\le \mathbb{P}\left[\frac{1}{2}\min_{\substack{1\le i\le n\\ i\ne j}}\|P_j-P_i\|_2 \le \frac{1}{n^{\frac{1}{d}+\varepsilon'}}\right] + \mathbb{P}\left[\min_{1\le i\le k}\|P_j-D_i\|_2 > \frac{1}{n^{\frac{1}{d}+\varepsilon'}}\right]\\
&\le \mathbb{P}\left[B(P_j,\frac{2}{n^{\frac{1}{d}+\varepsilon'}})\cap P \ne \{P_j\}\right] + \mathbb{E}\left[\min_{1\le i\le k}\|P_j-D_i\|_2\right]\cdot n^{\frac{1}{d}+\varepsilon'}.
\end{aligned}$$

Since $\mathbb{P}\left[B(P_j,\frac{2}{n^{\frac{1}{d}+\varepsilon'}})\cap P = \{P_j\}\right] \ge \left(1-v_d\left(\frac{2}{n^{\frac{1}{d}+\varepsilon'}}\right)^d\right)^n$ and

$\mathbb{E}\left[\min_{1\le i\le k}\|P_j-D_i\|_2\right] \le \frac{\Gamma(\frac{1}{d})}{d(v_dk)^{1/d}} + 2v_d^{-1/d}\exp(-v_d2^{-d}k^{1/2}) + \frac{2^d\Gamma(\frac{1}{d})}{k^{1/2d}d(v_dk2^{-d})^{1/d}}$, see proof of Lemma 5.4,

$$\begin{aligned}
&\mathbb{P}\left[\min_{1\le i\le k}\|P_j-D_i\|_2 > \frac{1}{2}\min_{\substack{1\le i\le n\\ i\ne j}}\|P_j-P_i\|_2\right]\\
&\le 1-\left(1-v_d\left(\frac{2}{n^{\frac{1}{d}+\varepsilon'}}\right)^d\right)^n + \mathbb{E}\left[\min_{1\le i\le k}\|P_j-D_i\|_2\right]\cdot n^{\frac{1}{d}+\varepsilon'}\\
&\le 1-\exp(-\frac{2^{d+1}v_d}{n^{d\varepsilon'}})
\end{aligned}$$

$$+\frac{n^{\frac{1}{d}+\varepsilon'}\Gamma(\frac{1}{d})}{d(v_d k)^{1/d}} + 2v_d^{-1/d}\exp(-v_d 2^{-d}k^{1/2}) + \frac{2^d\Gamma(\frac{1}{d})}{k^{1/2d}d(v_d k 2^{-d})^{1/d}}$$

since $1 - x \geq \exp(-2x)$ for small x. So the expected number of points whose nearest neighbors are distinct depots at a distance at most half of the distance to the nearest point is at least $n(\exp(-\frac{2^{d+1}v_d}{n^{d\varepsilon'}}) - \frac{n^{\frac{1}{d}+\varepsilon'}\Gamma(\frac{1}{d})}{d(v_d k)^{1/d}} - 2v_d^{-1/d}\exp(-v_d 2^{-d}k^{1/2}) - \frac{2^d\Gamma(\frac{1}{d})}{k^{1/2d}d(v_d k 2^{-d})^{1/d}}) = n(1 - o(1))$ for $k = \Omega(n^{1+\varepsilon})$.

Using Lemma 5.4 and taking into account that each point is connected by two edges to its nearest depot, we conclude that $\frac{\mathbb{E}[L(D,P)]}{nk^{-1/d}}$ tends to $\frac{2\Gamma(\frac{1}{d})\Gamma(\frac{d}{2}+1)^{1/d}}{d\sqrt{\pi}}$ for $n \to \infty$. □

We leave open the problem to exhibit the asymptotics not only for the expectation, but also for the functional. But we can not apply the concentration inequality by Rhee to the functional, since we do not have the usual $n^{(d-1)/d)}$ asymptotics here. In order to show the $nk^{-1/d}$ asymptotics, we would need a different concentration inequality or a stronger smoothness property of the functional.

5.2.1 The two dimensional MDVRP in case $k = n^{1+\varepsilon}$ for $0 < \varepsilon < 1$

In two dimensions, we can analyze the total length of the MDVRP in case $k = n^{1+\varepsilon}$ for $0 < \varepsilon < 1$. In two dimensions and in the given range of values for ε, we can modify the proof of Rhee's theorem 2.4 and get a concentration result. We use the notation introduced in Section 2.2. Note that the following lemma covers the second part of Theorem 5.3.

Lemma 5.6. *Let $P = \{X_1, \ldots, X_n\}$ and $D = \{X_{n+1}, \ldots, X_{n+k}\}$ be sets of independent random variables with uniform distribution and values in $[0,1]^2$, and let $k = n^{1+\varepsilon}$ for $0 < \varepsilon < 1$. Then*

$$\lim_{n\to\infty} \frac{L(D,P)}{nk^{-1/2}} = 1 \ c.c.$$

Proof. In the proof Lemma 2.5 is used to get a concentration inequality for the MDVRP about its median. Note that in the used notation $y = (y_1, \ldots, y_n, y_{n+1}, \ldots, y_{n+k}) \in \Omega^{n+k}$ the points are given by the first n coordinates and the depots by the last k coordinates. Let M denote a median of $L(y)$, $y \in \Omega^{n+k}$, and let A be the set of $y \in \Omega^{n+k}$ for which

$$L(y) \geq M.$$

By the definition of the median we have $\mu^{n+k}(A) \geq \frac{1}{2}$. As before, for all $t > 0$ the t-enlargement of A is

$$A_t := \{x \in \Omega^{n+k} : \exists_{y \in A} \text{ s.t. } H(x,y) \leq t\}.$$

Let G be a regular grid with k gridpoints in $[0,1]^2$. Let $B \subseteq \Omega^{n+k}$ defined by

$$B := \left\{ y \in \Omega^{n+k} : \max_{1 \leq i \leq k} d(g_i, D) \leq C \left(\frac{\log k}{k} \right)^{1/2} \right\},$$

where $d(g_i, D)$ denotes the Euclidean distance between a gridpoint g_i and the set of depots D. So for $y \in B$ all points may be connected to a depot with two edges of length at most $C \left(\frac{\log k}{k}\right)^{1/2}$. B is a high probability set: we are given a set D of k depots in $[0,1]^2$ by iid random variables with uniform distribution. Let g be a gridpoint in $[0,1]^2$, we bound the probability that all depots of D are at a distance of at least $C \left(\frac{\log k}{k}\right)^{1/2}$ from g:

$$\mathbb{P}\left[d(g,D) \geq \left(C \frac{\log k}{k} \right)^{1/2} \right] \leq \left(1 - \frac{C\pi \log k}{4k} \right)^k \leq \exp\left(-\frac{k\pi \log k^C}{4k} \right) \leq \frac{1}{k^C}.$$

Thus, we get $\mu^{n+k}(B^c) \leq 1/k^{C'}$, where $C' > 1$ with an appropriate choice of C in the definition of B.

We have $\mu^{n+k}(A \cap B) \geq \frac{1}{3}$, because $\mu^{n+k}(A) \geq \frac{1}{2}$ and B is a high probability set. Applying Lemma 2.5, we have

$$\mu^{n+k}\left(\{ y \in \Omega^{n+k} : \Phi_{A \cap B}(y) \geq t \} \right) \leq 6 \exp\left(-t^2/8(n+k) \right).$$

For the complement of the $t\sqrt{n+k}$-enlargement $(A \cap B)_{t\sqrt{n+k}}$ we have

$$\mu^{n+k}((A \cap B)^c_{t\sqrt{n+k}}) \leq 6 \exp\left(-t^2/8 \right).$$

Consider $E := B \cap (A \cap B)_{t\sqrt{n+k}}$, we obtain

$$\begin{aligned} \mu^{n+k}(E^c) &= \mu^{n+k}\left(B^c \cup (A \cap B)^c_{t\sqrt{n+k}} \right) \\ &\leq \mu^{n+k}(B^c) + \mu^{n+k}\left((A \cap B)^c_{t\sqrt{n+k}} \right) \\ &\leq \frac{1}{k^{C'}} + 6 \exp\left(-t^2/8 \right). \end{aligned}$$

If $x \in E$ then there is a $y \in A \cap B$ such that:

- $H(x,y) \leq t\sqrt{n+k}$.

- all points may be connected to a depot with two edges of length at most $C\left(\frac{\log k}{k}\right)^{1/2}$.

In the following we show that for this choice of x and y

$$|L(x) - L(y)| \leq Ct\sqrt{n+k}\left(\frac{\log k}{k}\right)^{1/2}.$$

Let G be a graph associated to $L(y)$. In order to modify G into a feasible MDVRP graph for x, we remove in each cycle of G the points of $y \setminus x$ by shortcuts. We replace each depot of $y \setminus x$ in a cycle of G with the nearest depot x: For each depot of $y \setminus x$ in a cycle of G there is a depot of x at distance at most $C\left(\frac{\log k}{k}\right)^{1/2}$, since x is an element of B. Thus, we can connect in each cycle with a depot in $y \setminus x$ the depot to the nearest depot of x with two edges of length at most $C\left(\frac{\log k}{k}\right)^{1/2}$ and remove the depot in $y \setminus x$ by shortcuts. Furthermore, we connect each point in $x \setminus y$ to its nearest depot by two edges of length at most $C\left(\frac{\log k}{k}\right)^{1/2}$. If a depot is now part of more than two cycles, all cycles are merged into a single cycle by shortcuts. Altogether we add at most $2H(x,y)$ edges of length at most $C\left(\frac{\log k}{k}\right)^{1/2}$ during the modification. Thus,

$$L(x) \leq L(y) + t\sqrt{n+k}C\left(\frac{\log k}{k}\right)^{1/2}.$$

Swapping the roles of x and y, it is easy to see that

$$L(y) \leq L(x) + t\sqrt{n+k}C\left(\frac{\log k}{k}\right)^{1/2},$$

and the above assumption is proven. So for an $x \in E$,

$$L(x) \geq L(y) - |L(y) - L(x)| \geq M - Ct\sqrt{n+k}\left(\frac{\log k}{k}\right)^{1/2}.$$

Thus it follows for all $z \in \Omega^{n+k}$ that

$$\begin{aligned}\mathbb{P}\left[L(z) \leq M - Ct\sqrt{n+k}\left(\frac{\log k}{k}\right)^{1/2}\right] &\leq \mu^{n+k}(E^c) \\ &\leq \frac{1}{k^{C'}} + 6\exp\left(-t^2/8\right).\end{aligned}$$

Let A' be the set of $y \in \Omega^{n+k}$ for which

$$L(y) \leq M.$$

We can show the reverse inequality with the same reasoning

$$\mathbb{P}\left[L(z) \geq M + Ct\sqrt{n+k}\left(\frac{\log k}{k}\right)^{1/2}\right] \leq \frac{1}{k^{C'}} + 6\exp\left(-t^2/8\right),$$

and we get

$$\mathbb{P}\left[|L(z) - M| \geq Ct\sqrt{n+k}\left(\frac{\log k}{k}\right)^{1/2}\right] \leq \frac{1}{k^{C'}} + 12\exp\left(-t^2/8\right).$$

For $t := \gamma\frac{n}{C((n+k)\log k)^{1/2}}$, where $\gamma > 0$ is fixed, we obtain the concentration inequality

$$\mathbb{P}\left[\left|\frac{L(z) - M}{nk^{-1/2}}\right| \geq \gamma\right] \leq \frac{1}{k^{C'}} + 12\exp\left(-\gamma^2\frac{n^2}{C(n+k)(\log k)}\right). \tag{5.1}$$

We have

$$\begin{aligned}
&\sum_{n=1}^{\infty}\mathbb{P}\left[\left|\frac{L(z) - M}{n(n^{1+\varepsilon})^{-1/2}}\right| \geq \gamma\right] \\
&\leq \sum_{n=1}^{\infty}\frac{1}{(n^{1+\varepsilon})^{C'}} + \sum_{n=1}^{\infty}12\exp\left(-\gamma^2\frac{n^2}{C(n+n^{1+\varepsilon})(\log(n^{1+\varepsilon}))}\right) \\
&< \infty,
\end{aligned}$$

since the first sum is a converging series($C' > 1$), and the second sum is also converging: note that for all $\beta > 0$ and large n we have $n^\beta > \log n$. There is a $\delta > 0$ such that $\sum_{n=1}^{\infty} 12\exp\left(-\gamma^2\frac{n^2}{C(n+(n^{1+\varepsilon}))(\log(n^{1+\varepsilon}))}\right) \leq \sum_{n=1}^{\infty} 12\exp\left(-\gamma^2 n^\delta/C\right)$. Since $\sum_{n=1}^{\infty} 12\exp\left(-\gamma^2 n^\delta/C\right) < \infty$ for all $\gamma > 0$,

$$\lim_{n\to\infty}\left|\frac{L(z) - M}{nk^{-1/2}}\right| = 0 \text{ c.c.}$$

Since $\left|\frac{L(z)-M}{nk^{-1/2}}\right| \leq \frac{Cn}{nk^{-1/2}} = Cn^{(1+\varepsilon)/2}$, we have with (5.1)

$$\mathbb{E}\left[\left|\frac{L(z) - M}{nk^{-1/2}}\right|\right] \leq \int_0^{Cn^{(1+\varepsilon)/2}} \frac{1}{k^{C'}} + 12\exp\left(-\gamma^2\frac{n^2}{C(n+k)(\log k)}\right)d\gamma.$$

Note $C' > 1$, so integration of the inequality above yields

$$\lim_{n\to\infty}\mathbb{E}\left[\left|\frac{L(z) - M}{nk^{-1/2}}\right|\right] = 0$$

and hence

$$\lim_{n\to\infty}\left|\frac{\mathbb{E}[L(z)] - M}{nk^{-1/2}}\right| = 0.$$

According to Lemma 5.5,

$$\lim_{n\to\infty} \frac{\mathbb{E}[L(z)]}{nk^{-1/2}} = \frac{2\Gamma(\frac{1}{d})\Gamma(\frac{d}{2}+1)^{1/d}}{d\sqrt{\pi}} = 1,$$

so the triangle inequality yields

$$\begin{aligned}
&\lim_{n\to\infty} \frac{|L(z) - nk^{-1/2}|}{nk^{-1/2}} \\
&= \lim_{n\to\infty} \left| \frac{L(z) - M + \mathbb{E}[L(z)] - nk^{-1/2} + M - \mathbb{E}[L(z)]}{nk^{-1/2}} \right| \\
&\leq \lim_{n\to\infty} \left[\left| \frac{L(z) - M}{nk^{-1/2}} \right| + \left| \frac{\mathbb{E}[L(z)] - nk^{-1/2}}{nk^{-1/d}} \right| + \left| \frac{M - \mathbb{E}[L(z)]}{nk^{-1/2}} \right| \right] \\
&= 0 \text{ c.c.}
\end{aligned}$$

Therefore,

$$\lim_{n\to\infty} \frac{L(z)}{nk^{-1/d}} = 1 \text{ c.c.}$$

□

Chapter 6

The MDVRP with General Distributions

We consider in this chapter not only uniformly distributed random variables, but all distributions with an absolutely continuous part of the density. Consider a partition of $[0,1]^d$ into congruent subcubes. The idea is to approximate the general distribution on $[0,1]^d$ by a linear combination of uniform distributions on the subcubes, recall that Theorem 4.1 gives the asymptotic behavior of the functional on the subcubes. We have for the MDVRP the following result:

Theorem 6.1. *Let $D = \{D_1, \ldots, D_k\}$ and $P = \{P_1, \ldots, P_n\}$ be sets of depots and points in $[0,1]^d$, $d \geq 2$, given by iid random variables. The optimal length $L(D,P)$ of an MDVRP tour through D and P satisfies*

(i) $\lim_{n\to\infty} \frac{L(D,P)}{n^{(d-1)/d}} = \alpha(L_{TSP}, d) \int_{[0,1]^d} f(x)^{(d-1)/d} dx$ *c.c., if* $k = o(n)$,

(ii) $\lim_{n\to\infty} \frac{L(D,P)}{n^{(d-1)/d}} = \alpha(L_{MDVRP}, \lambda, d) \int_{[0,1]^d} f(x)^{(d-1)/d} dx$ *c.c., if* $k = \lambda n + o(n)$ *for a constant* $\lambda > 0$,

where f is the density of the absolutely continuous part of the law of D_1 and where $\alpha(L_{TSP}, d)$ is the constant for the TSP and $\alpha(L_{MDVRP}, \lambda, d) > 0$ is a positive constant.

Note that the constant $\alpha(L_{MDVRP}, \lambda, d)$ does not depend on the $o(n)$ term, it depends only on λ and the dimension d.

The proof of the first part of the theorem is almost verbatim the proof of Theorem 4.1 (i), we only have to replace $\alpha(L_{TSP}, d)$ by $\alpha(L_{TSP}, d) \int_{[0,1]^d} f(x)^{(d-1)/d} dx$.

For the second part of the theorem, we need the following two lemmas. The first one was proven by Strassen [Str65] in 1965.

Lemma 6.2 ([Str65]). *Suppose P and Q are probability measures on a bounded subset of $\mathbb{R}^d$ and suppose also that there is an $\varepsilon > 0$ such that P and Q satisfy $P(A) \leq Q(A)+\varepsilon$ for all closed A. There is a probability measure $\hat{\mu}$ on the product space $\mathbb{R}^d \times \mathbb{R}^d$ such that*

$$\hat{\mu}(\cdot, \mathbb{R}^d) = P(\cdot),\ \hat{\mu}(\mathbb{R}^d, \cdot) = Q(\cdot) \text{ and } \hat{\mu}\left(\{(x,y) : x \neq y\}\right) \leq \varepsilon.$$

The second lemma is an extension of a lemma that was proven by Steele in 1988 [Ste88]. It shows that it suffices to prove the limit result for the expectation for a special class of distributions, the so-called blocked distributions, in order to show the limit result for the expectation when the density of the distribution is a mixture of absolutely continuous and singular laws. We extend the lemma in order to handle functionals that are defined on two point sets.

Lemma 6.3. *Let F be a smooth subadditive Euclidean functional and suppose that for all sequences of iid random variables $(X_i)_{i\geq 1}$ and $(Y_i)_{i\geq 1}$ distributed with a blocked distribution $\mu(x) := \phi(x)dx + \mu_s$ with $\phi(x) := \sum_{i=1}^{m^d} \gamma_i \mathbf{1}_{Q_i}(x)$, where γ_i are constants and Q_i for $1 \leq i \leq m^d$ form a partition of $[0,1]^d$ into m^d congruent subcubes, we have that*

$$\lim_{n\to\infty} \frac{\mathbb{E}[F(X_1,\ldots,X_{\lambda n}, Y_1, \ldots, Y_n)]}{n^{(d-1)/d}} = \alpha(F,\lambda,d) \int_{[0,1]^d} \phi(x)^{(d-1)/d} dx$$

for a positive constant λ. Whenever $(A_i)_{i\geq 1}$ and $(B_i)_{i\geq 1}$ are independent and identically distributed with respect to any probability measure on $[0,1]^d$ with an absolutely continuous part given by $f(x)dx$, we then have that

$$\lim_{n\to\infty} \frac{\mathbb{E}[F(A_1,\ldots,A_{\lambda n}, B_1, \ldots, B_n)]}{n^{(d-1)/d}} = \alpha(F,\lambda,d) \int_{[0,1]^d} f(x)^{(d-1)/d} dx.$$

Proof. Let $(Q_i)_{i=1}^{m^d}$ be partition of $[0,1]^d$ into m^d congruent subcubes. If the $(A_i)_{i\geq 1}$ and $(B_i)_{i\geq 1}$ are distributed according to $f(x)dx+\mu_s$, where μ_s is singular, we take an approximation $g_m(x)dx + \mu_s$, where $g_m(x) = \sum_{i=1}^{m^d} \gamma_i \mathbf{1}_{Q_i}$ and $\gamma_i = \int_{Q_i} f(x)dx$. We have $\lim_{m\to\infty} \int_{A\subseteq[0,1]^d} |g_m(x) - f(x)|dx = 0$.

Thus, for measures M and M' defined by

$$M(A) = \int_A f(x)dx + \mu_s(A) \text{ and } M'(A) = \int_A g_m(x)dx + \mu_s(A),$$

we have $|M(A)-M'(A)| \leq \int_A |g_m(x)-f(x)|dx \leq \varepsilon$ for all $m \geq m(\varepsilon)$ and $\varepsilon > 0$. By Lemma 6.2 there is a probability measure $\hat{\mu}$ on $\mathbb{R}^d \times \mathbb{R}^d$ such that

$$\hat{\mu}(\cdot, \mathbb{R}^d) = M'(\cdot),\ \hat{\mu}(\mathbb{R}^d, \cdot) = M(\cdot) \text{ and } \hat{\mu}\left(\{(x,y) : x \neq y\}\right) \leq \varepsilon.$$

We define sequences of random variables $(X_i, A_i)_{i\geq 1}$ and $(Y_i, B_i)_{i\geq 1}$, where (X_i, A_i) respectively (Y_i, B_i) are the i-th vector of an independent sequence of random vectors with distribution given by the measure $\hat{\mu}$. So we have $\mathbb{E}[|\{i \in \{1,\ldots,\lambda n\} : X_i \neq A_i\}|] \leq \lambda n \varepsilon$, respectively $\mathbb{E}[|\{i \in \{1,\ldots,n\} : Y_i \neq B_i\}|] \leq n\varepsilon$. So by smoothness and Jensen's inequality it follows that

$$\begin{aligned}
&|\mathbb{E}\left[F(X_1,\ldots,X_{\lambda n},Y_1,\ldots,Y_n)\right] - \mathbb{E}\left[F(A_1,\ldots,A_{\lambda n},B_1,\ldots,B_n)\right]| \\
&\quad\leq\ C\,\mathbb{E}\left[\left(|\{i \in \{1,\ldots,\lambda n\} : X_i \neq A_i\}| + |\{i \in \{1,\ldots,n\} : Y_i \neq B_i\}|\right)^{(d-1)/d}\right] \\
&\quad\leq\ C(n\varepsilon)^{(d-1)/d},
\end{aligned}$$

where C depends also on λ now. By the triangle inequality we get

$$\begin{aligned}
&\limsup_{n\to\infty}\left|\frac{\mathbb{E}[F(A_1,\ldots,A_{\lambda n},B_1,\ldots,B_n)]}{n^{(d-1)/d}} - \alpha(F,\lambda,d)\int_{[0,1]^d}\phi(x)^{(d-1)/d}dx\right| \\
&\quad=\ \limsup_{n\to\infty}\left|\frac{\mathbb{E}[F(A_1,\ldots,A_{\lambda n},B_1,\ldots,B_n)] - \mathbb{E}[F(X_1,\ldots,X_{\lambda n},Y_1,\ldots,Y_n)]}{n^{(d-1)/d}}\right. \\
&\qquad\left.+\frac{\mathbb{E}[F(X_1,\ldots,X_{\lambda n},Y_1,\ldots,Y_n)]}{n^{(d-1)/d}} - \alpha(F,\lambda,d)\int_{[0,1]^d}\phi(x)^{(d-1)/d}dx\right| \\
&\quad\leq\ C\varepsilon^{(d-1)/d}.
\end{aligned}$$

Since for all $a,b \geq 0$ we have

$$|a^{(d-1)/d} - b^{(d-1)/d}| \leq |a-b|^{(d-1)/d},$$

using Hölder's inequality it follows that

$$\begin{aligned}
&\left|\int_{[0,1]^d} f(x)^{(d-1)/d}dx - \int_{[0,1]^d}\phi(x)^{(d-1)/d}dx\right| \\
&\quad\leq\ \int_{[0,1]^d}|f(x)-\phi(x)|^{(d-1)/d}dx \\
&\quad\leq\ \varepsilon^{(d-1)/d}.
\end{aligned}$$

Combining the two inequalities and letting ε tend to zero completes the proof. □

The proof of Theorem 6.1 (ii) is similar to proof of Theorem 7.1 in [Yuk98]. The idea of the proof is to generalize the proof of Theorem 4.1. The general random variables are approximated by linear combinations of uniform random variables and the result is shown for these blocked distributions. By Lemma 6.3 we get the result for the general distributions.

Proof of Theorem 6.1 (ii). We prove the theorem for the boundary functional for blocked distributions and apply Lemma 6.3. Let Q_i for $1 \le i \le m^d$ form a partition of $[0,1]^d$ into m^d congruent subcubes and $\mu(x) := \phi(x)dx + \mu_s$ with $\phi(x) := \sum_{i=1}^{m^d} \gamma_i \mathbf{1}_{Q_i}(x)$, where γ_i are constants. Fix $\varepsilon > 0$ with $\varepsilon > m^{-1}$ w.l.o.g. Let E denote the singular support of μ and let λ denote the Lebesgue measure on the cube. We may assume that m is chosen so that

- $E \subseteq A \cup B$, where A and B are disjoint, $\lambda(A) = 0$ and $\mu(A) \le \varepsilon$.
- $B := \bigcup_{i \in J} Q_i$ for some $J \subset I := \{1, \ldots, m^d\}$ and $\lambda(B) \le \varepsilon$.

Since the boundary functional is smooth and subadditive in the sense of Lemma 3.3, we have

$$\begin{aligned}
\mathbb{E}[L^B(D,P)] &\le \mathbb{E}[L^B(D \setminus A, P \setminus A)] + C((1+\lambda)n\varepsilon)^{(d-1)/d} \\
&\le \sum_{i \in I \setminus J} \mathbb{E}[L^B((D \setminus A) \cap Q_i, (P \setminus A) \cap Q_i)] \\
&\quad + \sum_{i \in J} \mathbb{E}[L^B((D \setminus A) \cap Q_i, (P \setminus A) \cap Q_i)] \\
&\quad + C m^{d-1} \left(\frac{n}{m^d}\right)^{(d-2)/(d-1)} + C((1+\lambda)n\varepsilon)^{(d-1)/d}.
\end{aligned}$$

Let $(U_i)_{i \ge 1}$ and $(W_i)_{i \ge 1}$ be iid random variables with uniform distribution on $[0,1]^d$. The number of depots respectively points that fall into a subcube Q_i, $i \in I \setminus J$, of volume m^{-d} are given by binomial random variables X respectively Y with distribution $B(k, \gamma_i m^{-d})$ and $B(n, \gamma_i m^{-d})$. As in the proof of Lemma 4.4 we have by smoothness, homogeneity and Jensen's inequality for the the first sum that

$$\begin{aligned}
&\sum_{i \in I \setminus J} \mathbb{E}[L^B((D \setminus A) \cap Q_i, (P \setminus A) \cap Q_i)] \\
&\le m^{-1} \sum_{i \in I \setminus J} \left(\mathbb{E}[L^B((U_i)_{i=1}^{\gamma_i m^{-d} k}, (W_i)_{i=1}^{\gamma_i m^{-d} n})] \right) \\
&\quad + C\,\mathbb{E}\left[\left(|X - \gamma_i m^{-d} k| + |Y - \gamma_i m^{-d} n| \right)^{(d-1)/d} \right] \\
&\le m^{-1} \sum_{i \in I \setminus J} \left(\mathbb{E}[L^B((U_i)_{i=1}^{\gamma_i m^{-d} k}, (W_i)_{i=1}^{\gamma_i m^{-d} n})] \right) \\
&\quad + C(km^{-d})^{(d-1)/2d} + C(nm^{-d})^{(d-1)/2d}.
\end{aligned}$$

Now, we consider the second sum. The expected number of depots and points in $Q_i \setminus A$ is at most $k\mu(Q_i)$ respectively $n\mu(Q_i)$. By Jensen's inequality, smoothness and Hölder's inequality we bound the second sum:

$$\begin{aligned}
&\sum_{i\in J}\mathbb{E}[L^B((D\setminus A)\cap Q_i,(P\setminus A)\cap Q_i)]\\
&\quad\le\ m^{-1}C\sum_{i\in J}(k\mu(Q_i)+n\mu(Q_i))^{(d-1)/d}\\
&\quad\le\ C\left((1+\lambda)n\right)^{(d-1)/d}\sum_{i\in J}(m^{-d})^{1/d}\mu(Q_i)^{(d-1)/d}\\
&\quad\le\ C\left((1+\lambda)n\right)^{(d-1)/d}\left(\sum_{i\in J}m^{-d}\right)^{1/d}\\
&\quad\le\ C\left((1+\lambda)n\right)^{(d-1)/d}\left(\lambda(B)\right)^{1/d}\\
&\quad\le\ C\left((1+\lambda)n\right)^{(d-1)/d}\varepsilon^{1/d},
\end{aligned}$$

since $\lambda(B)\le\varepsilon$.We obtain by these two estimates

$$\begin{aligned}
&\mathbb{E}[L^B(D,P)]\\
&\quad\le\ m^{-1}\sum_{i\in I\setminus J}\left(\mathbb{E}[L^B((U_i)_{i=1}^{\gamma_i m^{-d}k},(W_i)_{i=1}^{\gamma_i m^{-d}n})]\right)\\
&\qquad+C(km^{-d})^{(d-1)/2d}+C(nm^{-d})^{(d-1)/2d}+C\left((1+\lambda)n\right)^{(d-1)/d}\varepsilon^{1/d}\\
&\qquad+Cm^{d-1}\left(\frac{n}{m^d}\right)^{(d-2)/(d-1)}+C((1+\lambda)n\varepsilon)^{(d-1)/d}.
\end{aligned}$$

Dividing by $n^{(d-1)/d}$, replacing k by $(1+\lambda)n$ and putting λ into the constant C we obtain

$$\begin{aligned}
&\mathbb{E}[L^B(D,P)]/n^{(d-1)/d}\\
&\quad\le m^{-1}\sum_{i\in I\setminus J}((\gamma_i m^{-d}n)/n)^{(d-1)/d}\left(\mathbb{E}[L^B((U_i)_{i=1}^{\gamma_i m^{-d}k},(W_i)_{i=1}^{\gamma_i m^{-d}n})]/(\gamma_i m^{-d}n)^{(d-1)/d}\right)\\
&\qquad+Cm^{(1-d)/2}n^{(1-d)/2d}+Cm^{1/(d-1)}n^{-1/d(d-1)}+C\varepsilon^{(d-1)/d}+C\varepsilon^{1/d}.
\end{aligned}$$

The right hand of the inequality contains the functional L^B over two sequences of iid random variables with uniform distribution. Letting n tend to infinity and using Theorem 4.1 to evaluate the right side we get

$$\begin{aligned}
&\limsup_{n\to\infty}\mathbb{E}[L^B(D,P)]/n^{(d-1)/d}\\
&\quad\le\ \sum_{i\in I\setminus J}\gamma_i^{(d-1)/d}m^{-d}\alpha(L_{MDVRP},\lambda,d)+C\varepsilon^{(d-1)/d}+C\varepsilon^{1/d}\\
&\quad=\ \alpha(L_{MDVRP},\lambda,d)\int_{\bigcup_{i\in I\setminus J}Q_i}\phi(x)^{(d-1)/d}dx+C\varepsilon^{(d-1)/d}+C\varepsilon^{1/d}.
\end{aligned}$$

For $\varepsilon \to 0$, m tends to infinity and $\bigcup_{i\in I\setminus J} Q_i \uparrow [0,1]^d$. Thus,

$$\begin{aligned}
&\limsup_{n\to\infty} \mathbb{E}[L^B(D,P)]/n^{(d-1)/d} \\
&\quad\leq \alpha(L_{MDVRP}, \lambda, d) \int \phi(x)^{(d-1)/d} dx.
\end{aligned}$$

The lower bound is established by the same ideas, but we use the superadditivity of the boundary functional instead of the subadditivity.

By the smoothness and the superadditivity of the boundary functional we get

$$\begin{aligned}
\mathbb{E}[L^B(D,P)] &\geq \mathbb{E}[L^B(D\setminus A, P\setminus A)] - C((1+\lambda)n\varepsilon)^{(d-1)/d} \\
&\geq \sum_{i\in I\setminus J} \mathbb{E}[L^B((D\setminus A)\cap Q_i, (P\setminus A)\cap Q_i)] - C((1+\lambda)n\varepsilon)^{(d-1)/d}.
\end{aligned}$$

The following is analogous to the upper bound above. Let $(U_i)_{i\geq 1}$ and $(W_i)_{i\geq 1}$ be iid random variables with uniform distribution on $[0,1]^d$. The number of depots respectively points that fall into a subcube Q_i, $i \in I\setminus J$, of volume m^{-d} are given by binomial random variables X respectively Y with distribution $B(k, \gamma_i m^{-d})$ and $B(n, \gamma_i m^{-d})$. We have by smoothness, homogeneity and Jensen's inequality for the the first sum that

$$\begin{aligned}
&\sum_{i\in I\setminus J} \mathbb{E}[L^B((D\setminus A)\cap Q_i, (P\setminus A)\cap Q_i)] \\
&\quad\geq m^{-1} \sum_{i\in I\setminus J} \left(\mathbb{E}[L^B((U_i)_{i=1}^{\gamma_i m^{-d}k}, (W_i)_{i=1}^{\gamma_i m^{-d}n})]\right) \\
&\qquad -C\,\mathbb{E}\left[\left(|X-\gamma_i m^{-d}k| + |Y-\gamma_i m^{-d}n|\right)^{(d-1)/d}\right] \\
&\quad\geq m^{-1} \sum_{i\in I\setminus J} \left(\mathbb{E}[L^B((U_i)_{i=1}^{\gamma_i m^{-d}k}, (W_i)_{i=1}^{\gamma_i m^{-d}n})]\right) \\
&\qquad -C(km^{-d})^{(d-1)/2d} - C(nm^{-d})^{(d-1)/2d}.
\end{aligned}$$

Using this estimate and dividing by $n^{(d-1)/d}$ we get

$$\begin{aligned}
&\mathbb{E}[L^B(D,P)]/n^{(d-1)/d} \\
&\quad\geq m^{-1} \sum_{i\in I\setminus J} \mathbb{E}\left[L^B((U_i)_{i=1}^{\gamma_i m^{-d}k}, (W_i)_{i=1}^{\gamma_i m^{-d}n})\right]/n^{(d-1)/d} - C((1+\lambda)n\varepsilon/n)^{(d-1)/d} \\
&\qquad -C(km^{-d})^{(d-1)/2d}/n^{(d-1)/d} - C(nm^{-d})^{(d-1)/2d}/n^{(d-1)/d} \\
&\quad\geq \sum_{i\in I\setminus J} m^{-1}(\gamma_i m^{-d} n/n)^{(d-1)/d}\, \mathbb{E}\left[L^B((U_i)_{i=1}^{\gamma_i m^{-d}k}, (W_i)_{i=1}^{\gamma_i m^{-d}n})\right]/(\gamma_i m^{-d} n)^{(d-1)/d} \\
&\qquad -C\varepsilon^{(d-1)/d} - Cm^{(1-d)/2} n^{(1-d)/2d}.
\end{aligned}$$

As n tends to infinity, we obtain by Theorem 4.1

$$\begin{aligned}&\liminf_{n\to\infty} \mathbb{E}[L^B(D,P)]/n^{(d-1)/d} \\ &\geq \sum_{i\in I\setminus J} m^{-1}(\gamma_i m^{-d})^{(d-1)/d}\alpha(L_{MDVRP},\lambda,d) - C\varepsilon^{(d-1)/d}.\end{aligned}$$

As before, for $\varepsilon \to 0$, m tends to infinity and $\bigcup_{i\in I\setminus J} Q_i \uparrow [0,1]^d$. So we obtain

$$\liminf_{n\to\infty} \mathbb{E}[L^B(D,P)]/n^{(d-1)/d} \geq \alpha(L_{MDVRP},\lambda,d)\int \phi(x)^{(d-1)/d}dx.$$

Combining the limes superior and the limes inferior we get

$$\lim_{n\to\infty} \mathbb{E}[L^B(D,P)]/n^{(d-1)/d} = \alpha(L_{MDVRP},\lambda,d)\int \phi(x)^{(d-1)/d}dx.$$

Applying Lemma 6.3, we have Theorem 6.1 for the boundary MDVRP functional. By Lemma 3.4 the MDVRP functional has the same asymptotics as its boundary functional. The asymptotic behavior does not change in case $k = \lambda n + o(n)$, this can be shown as in the proof of Theorem 4.1 (i). $\square$

Chapter 7

Probabilistic Analysis of MDVRP Heuristics

The proof of Theorem 4.1 shows that partitioning the cube into subcubes leads to useful analytical relations. The same insight is used in many algorithms that apply to Euclidean optimization problems: the cube is partitioned into small subcubes and the problem is solved in the smaller subcubes. The solutions in the small cubes are put together by linking edges to construct a solution in the whole cube.

Karp [K76] was the first to observe that under different natural probabilistic models one can use this dissection to construct fast algorithms that yield almost optimal solutions with high probability. It was the first time that someone showed the existence of a polynomial time algorithm for the stochastic version of an NP-complete problem that provides a solution within a factor of $1 + \varepsilon$ times the value of the optimal solution with high probability. In [K77], Karp analyzes the partitioning heuristics for the traveling salesman problem in two dimensions. Using this seminal work as a guide, Karp and Steele [KS85] showed for the d-dimensional case of the traveling salesman problem that the ratio of the lengths of the heuristic tour and the optimal tour converges completely to one.

The result of Karp and Steele is an example of a general phenomenon: For a Euclidean functional L, there is a heuristic with length function L_H so that the ratio of the lengths converge completely to one. Using the general framework of subadditive Euclidean functionals and their associated superadditive boundary functionals, Yukich [Yuk98] extended the work of Karp and Steele to subadditive Euclidean functionals with superadditive boundary functionals over general sequences of random variables.

Theorem 7.1. *[Yuk98] Let L and L^B be subadditive and superadditive Euclidean functionals respectively. Assume that they are close. Then for all $\varepsilon > 0$ and all iid sequences X_i, $i \geq 1$, of random variables with a continuous part of the density function, there exists a heuristic with length function L_H such that*

$$\sum_{i=1}^{\infty} \mathbb{P}\left[\frac{L_H(X_1,\ldots,X_n)}{L(X_1,\ldots,X_n)} \geq 1+\varepsilon\right] < \infty.$$

Although the MDVRP functional is not subadditive, it is possible to construct a partitioning heuristic for the MDVRP. Again, we can exploit the versatile properties of the boundary modification. In Chapter 7.2 we show for the multi depot vehicle routing problem the same result as in Theorem 7.1.

In Chapter 7.1 we consider a heuristic that does not partition the unit cube into equal regions, but that subdivides the unit cube in a different way: the points are clustered by assigning each point to its nearest depot. Thus, the MDVRP reduces to the TSP in each cluster and we can rely on TSP heuristic, which have been the topic of elaborate studies.

7.1 Analysis of Nearest Neighbor Heuristics

We analyze a two phase scheme for the multi depot routing problem in two dimensions that combines a depot clustering heuristic with approximation algorithms for the TSP problem:

Two phase scheme:

1. *Cluster* the points around the depots.
2. *Connect* the points in each cluster and the corresponding depot with a TSP tour.

This is the most common approach to solve the problem practically [FJ81, GTV02], since well-studied TSP heuristic can be applied. We give an analysis for the case that the clustering step is implemented by the *nearest neighbor* heuristic: assign each point to its nearest depot. To our best knowledge this is the first theoretical analysis of this approach that considers an increasing number of depots.

The proof uses the following asymptotic result for Voronoi diagrams by McGivney and Yukich [MY99a] for independently distributed points in $[0,1]^2$. The result extends work of Miles [Mil70] and Avram and Bertsimas [AB93]. Recall the definition of the planar Voronoi diagram: we are given $x_1, \ldots, x_n \in [0,1]^2$, consider the locus of points closer to x_i, $1 \leq i \leq n$, than to any other point. This set of points is a cell and denoted by $\mathcal{C}_i$. The cells $\mathcal{C}_i$, $1 \leq i \leq n$, partition the square into a convex net which is called the Voronoi diagram of $[0,1]^2$.

Theorem 7.2. *[MY99a] Let X_i, $i \geq 1$, be iid random variables on $[0,1]^2$ with a continuous density f_Y which is bounded away from 0 and ∞. Let $V(X_1, \ldots, X_n)$ denote the total edge length of the Voronoi diagram, then*

$$\lim_{n\to\infty} \frac{V(X_1, \ldots, X_n)}{\sqrt{n}} = 2 \int_{[0,1]^2} (f_Y(x))^{1/2} dx \ c.c.$$

Let $\mathcal{C}$ denote a clustering, i.e., an assignment of points to depots. Let $\mathcal{C}^N$ denote the clustering produced by applying the nearest neighbor rule and let $\mathcal{C}^*$ denote the clustering in an optimal tour. For a clustering $\mathcal{C}$, let $\mathcal{T}^*(\mathcal{C})$ denote the total length of an optimal tour for the clustering $\mathcal{C}$.

Lemma 7.3. *Let the point set P and depot set D be given by be iid random variables on $[0,1]^2$ with a continuous density f which is bounded away from 0 and ∞. The nearest neighbor clustering rule satisfies*

$$\limsup_{k\to\infty} \frac{\mathcal{T}(\mathcal{C}^N) - \mathcal{T}(\mathcal{C}^*)}{\sqrt{k}} \leq 6 \int_{[0,1]^2} (f(x))^{1/2} dx \ c.c.$$

Proof. Consider the Voronoi partition V corresponding to the depots in the unit square. Take an optimal tour $\mathcal{T}$ corresponding to the optimal clustering $\mathcal{C}^*$ and modify it to respect the partition V. This is accomplished by "stitching in" a tour along the sides of a Voronoi cell and connecting it to an inner tour respectively depot, see Figure 7.1 and 7.2. This means to cut off the optimal tours keeping only the portions within the Voronoi cell and connecting the portions by edges along the Voroni diagram. The total length of the edges used for the "stichting" is at most $2V(D)$. The total length of the edges needed to connect the tour to the depot is also at most $V(D)$, since the depot of a cell can be connected to its boundary with two edges of total length at most half the perimeter of the cell. □

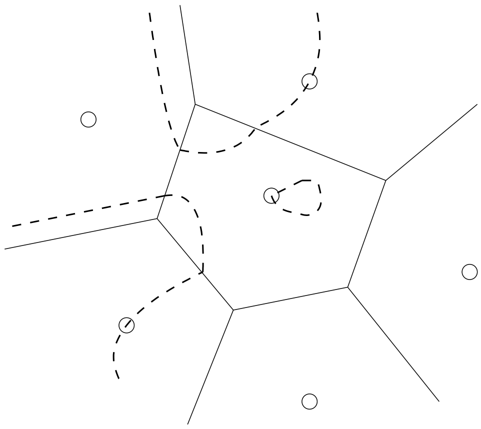

Figure 7.1: Six depots with Voronoi diagram, the optimal tour shown as dashed line

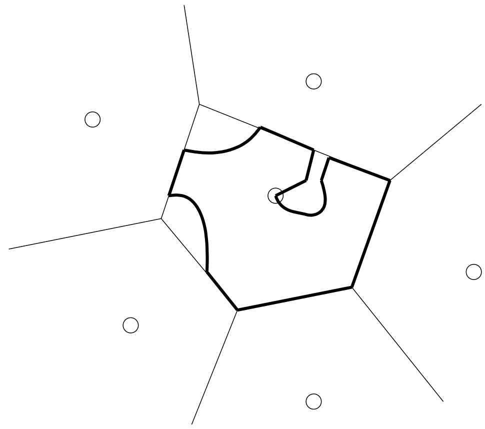

Figure 7.2: The stitched tour of the central depot as a bold line

7.2 Analysis of a Karp-type Heuristic

We apply the well-known ideas of Karp's heuristic for the TSP to the MDVRP. In this way we construct an asymptotical optimal MDVRP heuristic that gives an optimal depot tour almost surely in the cases $k = o(n)$ and $k = \lambda n$. We concentrate on these cases, because for $k \geq n^{1+\varepsilon}$ we have with high probability an all nearest neighbor problem. Since the MDVRP functional does not have the needed properties to apply the ideas of Karp, we construct a heuristic using the boundary functional.

The algorithm consists of dividing the d-dimensional unit cube $[0,1]^d$ into m^d congruent subcubes $Q_1, \ldots, Q_{m^d}$, and determining optimal boundary MDVRP tours in each non-empty subcube Q_i by brute force. A boundary MDVRP in $[0,1]^d$ is constructed by deleting the faces of the subcubes one by one and connecting resulting "loose ends" of the paths. The boundary tour is turned into an MDVRP tour as in Lemma 3.4.

The number of subcubes m^d depends on n, this approach is similar to Karp's[K76, K77]. Let $\sigma(n)$ be an unbounded increasing function and $1 < \frac{n}{\sigma(n)} = 2^{dj(n)}$ for some non-decreasing sequence of integers $j(n)$, $n \geq 1$. We note there is a non-decreasing sequence $j(n)$, $n \geq 1$, such that $C \log n \leq \frac{n}{2^{dj(n)}} \leq \log n$ for some constant $C < 1$. Thus, for this choice we have $\sigma(n) \leq \log n$. Furthermore, we set $m^d := \frac{n}{\sigma(n)}$.

Dissection Algorithm:

Input: Two sets D and P consisting of k respectively n points in $[0,1]^d$.
Output: An MDVRP tour.

1. Partition $[0,1]^d$ into m^d congruent subcubes Q_i.
2. Construct optimal boundary MDVRP tours in each Q_i.
3. Merge the tours in the Q_i into a boundary MDVRP tour in $[0,1]^d$.
4. Alter the boundary MDVRP tour into an MDVRP tour.

Let $L^\sigma(D,P)$ denote the tour length of the MDVRP tour produced by the heuristic with point set P and depot set D.

Lemma 7.4. *The tour length produced by the heuristic satisfies*

$$|L^\sigma(D,P) - L(D,P)| = o(n^{(d-1)/d}).$$

Proof. Let $|P| = n$ and $|D| = k$. First, we compare the total length of optimal boundary tours in the subcubes with the length of the boundary tour constructed in the first three steps, denoted by $L^{B\sigma}(D,P)$. We show

$$L^{B\sigma}(D,P) \leq \sum_{i=1}^{m^d} L^B(D \cap Q_i, P \cap Q_i) + Cn^{(d-1)/d}\sigma(n)^{-1/(d(d-1))}.$$

The first part of the proof is analogous to the proof of Lemma 3.3: Optimal tours in the subcubes Q_i are merged into a boundary tour in $[0,1]^d$ in the following way. Consider two neighboring subcubes Q_i and Q_j. There are two types of paths connected to the separating face between Q_i and Q_j: paths that have their start and endpoint on the face and paths that are only connected with one point of the face. We collect the points on the separating face where paths of the first type meet the boundary in a set B_1 and the points where paths of the second type meet the boundary in a set B_2. If the cardinality of B_2 is odd, a point of the boundary of the separating face is added to B_2. Now, a minimal matching of B_2 is added to the graph. The total length of the added edges is $\frac{C}{m}|B_2|^{(d-2)/(d-1)}$ by Lemma 2.1, since the face is a $d-1$ dimensional unit cube stretched by a factor of $\frac{1}{m}$. Furthermore, we add a minimal perfect matching of B_1 and a TSP tour through B_1 to the graph. In the connected component containing B_1 all vertices have even degree. So there is a Eulerian tour through the component. We turn the Eulerian tour into a TSP tour by shortcuts, delete an edge and connect both endpoints to the boundary of $Q_i \cup Q_j$. The total length of all edges used to connect the first type paths to the boundary of $Q_i \cup Q_j$ is at most $\frac{C}{m}|B_1|^{(d-2)/(d-1)}$, Lemma 2.1. The number of faces that have to removed by this procedure is bounded by dm^d. In every face i let B_1^i and B_2^i denote the number of points on the face where paths of the first respectively second type meet. The total length of all added edges is

$$\sum_{i=1}^{dm^d} C\left(|B_1^i|^{(d-2)/(d-1)} + |B_2^i|^{(d-2)/(d-1)}\right)\frac{1}{m}. \tag{7.1}$$

Each point may be connected to the boundary of its subcube by two edges, so $\sum_{i=1}^{dm^d} |B_1^i| + |B_2^i| \leq 2n$. The sum (7.1) is maximized for $B_j^i = \frac{2n}{2dm^d}$ for $i = 1, \ldots, dm^d$ and $j = 1, 2$. We have

$$L^{B\sigma}(D, P) \leq \sum_{i=1}^{m^d} L^B(D \cap Q_i, P \cap Q_i) + m^{d-1}C\left(\frac{n}{m^d}\right)^{(d-2)/(d-1)}.$$

The substitution $m^d = \frac{n}{\sigma(n)}$ gives

$$L^{B\sigma}(D, P) \leq \sum_{i=1}^{\frac{n}{\sigma(n)}} L^B(D \cap Q_i, P \cap Q_i) + Cn^{(d-1)/d}\sigma(n)^{-1/(d(d-1))}.$$

The boundary tour associated to $L^{B\sigma}(D, P)$ is turned into a multi depot tour as in Lemma 3.4: let B denote the set of points where the graph meets the boundary of $[0, 1]^d$. Since B is set of endpoints of paths, $|B|$ is even. We add to the graph a perfect minimal matching of B and a traveling salesman tour through B. In the new connected component all vertices have even degree, so there is a Eulerian tour. This tour is turned into a TSP tour by shortcuts and we connect it to a depot. As in Lemma 3.4, the total length of all added edges is $O(n^{(d-2)/(d-1)})$.

All in all, we have

$$\begin{aligned}
L(D, P) &\leq L^\sigma(D, P) \\
&\leq L^{B\sigma}(D, P) + Cn^{(d-2)/(d-1)} \\
&\leq \sum_{i=1}^{m^d} L^B(D \cap Q_i, P \cap Q_i) + Cn^{(d-1)/d}\sigma(n)^{-1/(d(d-1))} + Cn^{(d-2)/(d-1)} \\
&\leq L^B(D, P) + Cn^{(d-1)/d}\sigma(n)^{-1/(d(d-1))} + Cn^{(d-2)/(d-1)} \\
&\leq L(D, P) + Cn^{(d-1)/d}\sigma(n)^{-1/(d(d-1))} + Cn^{(d-2)/(d-1)}.
\end{aligned}$$

The last two steps are valid, because the boundary functional is superadditive and $L^B(D, P) \leq L(D, P)$. □

Theorem 7.5. *For independently uniformly distributed depots and points, we have for all $\varepsilon > 0$:*

$$\sum_{n=1}^{\infty} \mathbb{P}\left[\frac{L^{\sigma}(D,P)}{L(D,P)} \geq 1+\varepsilon\right] < \infty.$$

Proof. $L(D,P)/n^{(d-1)/d}$ converges completely to a constant $\alpha(L_{MDVRP}, \lambda, d)$, Theorem 6.1, and from Lemma 7.4 it follows that $L^{\sigma}(D,P)/n^{(d-1)/d}$ converges completely to $\alpha(L_{MDVRP}, \lambda, d)$, too. Thus, the quotient $L^{\sigma}(D,P)/L(D,P)$ converges completely to 1. □

In the remainder of this chapter we show that an n-point and k-depot boundary MDVRP can be solved by dynamic programming in $f(n) = O(3^n)$ time in the cases $k = \lambda n$ and $k = o(n)$. We show that for $\sigma(n) \leq \log n$ the expected execution time of the *Dissection Algorithm* is polynomial if points and depots are given by iid random variables with uniform distribution. The expected execution time of the *Dissection Algorithm* is of the same order as the time to solve the boundary MDVRP in the subcubes. The algorithm is presented on the next page.

The boundary MDVRP is solved in several steps. First, we identify for each subset S of P and each depot the shortest TSP tour on S and the depot, then we determine the shortest boundary TSP tour on each subset S of P. The shortest tour on each S and a depot and the shortest boundary TSP tour on S and the the lengths of these tours will be stored in the variables *depot_tsp*$[S]$ and *boundary_tsp*$[S]$, respectively *cost_dtsp*$[S]$ and *cost_btsp*$[S]$. This is done in subroutines called *Depot_TSP* and *Boundary_TSP*, they will be explained in detail later. For each $S \subseteq P$ the shorter of the two tours is stored in a variable *bestcycle*$[S]$, and the length of the tour in *cost_bc*$[S]$.

Let $\mathcal{C}$ be the weighted set containing the shortest cycles on all subsets of P, i.e. *bestcycle*$[S]$. Note that there is an optimal boundary MDVRP tour on P and D which is a set of disjunct subsets of $\mathcal{C}$ with minimal weight that covers all points of P. Such a minimal set cover of P that is an optimal boundary MDVRP tour can be determined by dynamic programming. For each $S \subseteq P$ with increasing cardinality an optimal cover is identified and the cover and its weight are stored in *cover*$[S]$ respectively *cost_c*$[S]$.

The running time of the algorithm is the sum of the running times of the 5 steps. The running time of the third step is at most $\sum_{j=1}^{n}\binom{n}{j}Cj \leq Cn2^n$, since the innermost loop takes Cj steps for a constant $C > 0$. The fourth step is negligible and the fifth step takes at most $\sum_{j=2}^{n}\binom{n}{j}C2^j \leq C3^n$ steps, since in the innermost loop we determine the minimum of 2^j values.

Boundary MDVRP:

Input: Two sets D and P consisting of k respectively n points in $[0,1]^d$.
Output: An optimal MDVRP tour.

1. $Depot\_TSP$ determines

   $depot\_tsp[S]$ and $cost\_dtsp[S]$ for all $S \subseteq P$

2. $Boundary\_TSP$ determines

   $boundary\_tsp[S]$ and $cost\_btsp[S]$ for all $S \subseteq P$

3. for $j = 1$ to n

   for each subset S of P with $|S| = j$ do

   $cost\_bc[S]$:= min $\{cost\_dtsp[S], cost\_btsp[S]\}$
   store shorter tour in $bestcycle[S]$

4. for each i in P do

   $cover[i]$:=$bestcycle[i]$

   $cost\_c[i]$:=$cost\_bc[i]$

5. for $j = 2$ to n do

   for each subset S of P with $|S| = j$ do

   $cost\_c[S] = \min\{\min_{J \subsetneq S}\{cost\_c[S \setminus J] + cost\_c[J]\}, cost\_bc[S]\}$
   store shortest cover in $cover[S]$

The first two steps are modifications of an algorithm by Held and Karp [HK62], they showed that an optimal solution of the traveling salesman problem on n points may be determined in $O(n^2 2^n)$ steps. In $Depot\_TSP$ we determine for each depot j and each subset S of P the shortest cycle on $S \cup \{j\}$ using the algorithm by Held and Karp. The shortest cycle over all considered depots is stored in *depot_tsp*$[S]$, and its length in *cost_dtsp*$[S]$. In the following we introduce the other variables used in *Depot_TSP*. The variable *cost*$[i, j]$ denotes the distance between two points i, j in $[0, 1]^d$. The variable *bestpath*$[S, i]$ denotes the shortest path through S starting at the considered depot j and ending in i, *cost_bp*$[S, i]$ stores its length. *cost_bc*$[S]$ contains the cost of the shortest cycle through S and the current depot.

Depot_TSP:

Input: Two sets D and P consisting of k respectively n points in $[0, 1]^d$.
Output: For each $S \subseteq P$ the shortest TSP tour on S containing a depot and its length.

1. for all $S \subseteq P$ set *cost_dtsp*$[S] = \infty$.

2. for $j = 1$ to k do
   {

   for $i = 1$ to n do

      cost_bp$[i, i]$:=*cost*$[j, i]$

      bestpath$[i, i]$:=(j,i)

   for $m = 2$ to n do

      for each subset S of P with $|S| = m$ do

         for each $i \in S$ do

            cost_bp$[S, i] := min_{l \in S \setminus \{i\}} \{cost\_bp[S \setminus \{i\}, l] + cost[l, i]\}$

            store shortest path in *bestpath*$[S, i]$

         cost_bc$[S] := \min_{l \in S} \{cost\_bp[S, l] + cost[l, j]\}$

         cost_dtsp$[S] := \min\{cost\_bc[S], cost\_dtsp[S]\}$

         store shorter cycle in *depot_tsp*$[S]$

   }

The running time of $Depot\_TSP$ is $O(\sum_{j=1}^{k} \sum_{l=1}^{n} \binom{n}{l} l^2) = O(kn^2 2^n)$.

In the routine $Boundary\_TSP$ we first determine for each point in P the nearest point on the boundary, the set of boundary points is denoted by B. For each two element set of boundary points $\{a, b\}$ and all $S \subseteq P$ we determine the shortest path through S with endpoints a and b, note that $a = b$ is an allowed choice. We need the additional variable $min\_cost(a, S, b)$, which stores the length of the shortest path through S with endpoints a and b.

Boundary_TSP:

Input: Two sets D and P consisting of k respectively n points in $[0, 1]^d$.
Output: For each $S \subseteq P$ the shortest path through S starting and ending at the boundary and its length.

1. determine set of boundary points B
2. for each $S \subseteq P$ set $cost\_btsp[S] = \infty$.
3. for each $E = \{a, b\} \subset B$ with $|E| = 2$ do
   {

   for $i = 1$ to n do

   $cost\_bp[i, i]$:=$cost[a, i]$

   $bestpath[i, i]$:=(a,i)

   for $t = 2$ to n do

   for each subset S of P with $|S| = t$ do

   for each $i \in S$ do

   $cost\_bp[S, i] := min_{l \in S \setminus \{i\}}\{cost\_bp[S \setminus \{i\}, l] + cost[l, i]\}$

   store shortest path in $bestpath[S, i]$

   $min\_cost(a, S, b) := \min_{j \in S}\{bestpath[S, j] + cost[j, b]\}$

   $cost\_btsp[S] := \min\{min\_cost(a, S, b), cost\_btsp[S]\}$

   store shorter path in $boundary\_tsp[S]$

   }

The running time of $Boundary\_TSP$ is $O((\binom{n}{2} + n)\sum_{l=2}^{n}\binom{n}{l}l^2) = O(n^4 2^n)$.

So all in all the algorithm $Boundary\_MDVRP$ has running time of $O(3^n + kn^2 2^n + n^4 2^n) = O(3^n)$ for n points and k depots.

The order of the running time of the $DissectionAlgorithm$ is bounded by the order of the running time of *Boundary_MDVRP*. The running time of *Boundary_MDVRP* depends on the number of points and depots that are in each cube. Let n_i be the number of points in cube i. Since points and depots are given by iid random variables with uniform distribution, the random variable n_i has binomial distribution, so

$$\mathbb{P}[n_i = l] = \sum_{l=0}^{n} \binom{n}{l} \left(\frac{1}{m^d}\right)^l \left(1 - \frac{1}{m^d}\right)^{n-l}.$$

We only consider the number of points, since we are only considering the cases $k = o(n)$ and $k = \lambda n$. Step 5 of *Boundary_MDVRP* has running time $f(n) = C3^n$. So the order of the running time of the *DissectionAlgorithm* is bounded by the order of

$$R = \sum_{i=1}^{m^d} f(n_i).$$

To determine the mean of R, we have to determine the mean of $f(n_i)$ using $m^d = n/\sigma(n)$:

$$\begin{aligned}
\mathbb{E}\left[f(n_i)\right] &= \sum_{l=0}^{n} C3^l \binom{n}{l} \left(\frac{1}{m^d}\right)^l \left(1 - \frac{1}{m^d}\right)^{n-l} \\
&\leq C\sum_{l=0}^{n} 3^l \binom{n}{l} \left(\frac{\sigma(n)}{n}\right)^l \\
&= C\sum_{l=0}^{n} \frac{3^l \sigma(n)^l}{l!} \frac{n!}{(n-l)!n^l} \\
&\leq C\sum_{l=0}^{n} \frac{3^l \sigma(n)^l}{l!} \leq C \exp\left(3\sigma(n)\right).
\end{aligned}$$

We have for the mean $\mathbb{E}\left[f(n_i)\right] \leq C\exp(3\sigma(n))$. By linearity of expectation the mean of the execution time of the *Dissection Algorithm* is $O(\exp(3\sigma(n))n/\sigma(n))$.

Since we have $\sigma(n) \leq \log n$, the expected execution time of the *Dissection Algorithm* is $O(n^4/\log n)$.

Chapter 8

The b-Degree Minimal Spanning Tree

Yukich [Yuk98] conjectures that the asymptotic behavior of the degree constrained minimal spanning tree also fits into the theory of boundary functionals. We settle this conjecture in the affirmative.

In the b-degree constrained minimal spanning tree problem (bMST) we are given a set P containing n points in $[0,1]^d$, and a degree bound $b \geq 2$. The aim is to find a spanning tree in which the degree of each vertex is at most b of minimum weight, where the weight of an edge is given by its Euclidean length. The total edge length is denoted by $L_{bMST}(P)$.

This is a generalization of the path version of the TSP. Furthermore it is the most basic problem of a family of well-studied problems about finding degree constrained structures. A fine survey is given by Raghavachari in [Rag96]. Concerning complexity, the case $b = 2$ is equivalent to the path version of the traveling salesman problem and hence NP-hard. For $b = 3$ Papadimitriou and Vazirani [PV84] showed that the problem remains NP-hard in the Euclidean plane. They conjecture that the problem is NP-hard for $b = 4$, but this question is still open. For $b = 5$ the problem is polynomially solvable, since there is always a minimal spanning tree with degree at most 5 [MS92] in the Euclidean plane.

Considering approximation algorithms, Arora and Chang [AC04] developed a quasi-polynomial time approximation scheme for the problem using the famous techniques for the TSP [Aro98]. The best polynomial approximation algorithms are given by Chan [Cha03], they can be implemented in linear time if the MST (without degree restriction) is given. He gave a 1.40 approximation for $b = 3$ and a 1.14 approximation for $b = 4$ in $\mathbb{R}^2$, and he gave a 1.63 approximation for $b = 3$ in $\mathbb{R}^d$.

In this chapter we show that the asymptotic behavior of the b-degree constrained MST can be analyzed with the help of its boundary modification and get the following result:

Theorem 8.1. *Let $P = \{P_1, \ldots, P_n\}$ be a set of points in $[0,1]^d$ given by iid random variables with an absolutely continuous part given by f. The optimal length $L_{bMST}(P)$ of a b-degree constrained minimal spanning tree on P satisfies*

$$\lim_{n\to\infty} \frac{L_{bMST}(P)}{n^{(d-1)/d}} = \alpha(L_{bMST}, d) \int_{[0,1]^d} f(x)^{(d-1)/d} dx \ \ c.c.,$$

where $\alpha(L_{bMST}, d)$ is a positive constant. In the case $b = 2$, the b-degree constrained MST has the same behavior like the TSP, we have $\alpha(L_{bMST}, d) = \alpha(L_{TSP}, d)$.

First of all, we consider the special case $b = 2$:

Lemma 8.2. *For $b = 2$, the b-degree constrained MST has the same behavior like the TSP, we have*

$$\lim_{n\to\infty} \frac{L_{2MST}(P)}{n^{(d-1)/d}} = \alpha(L_{TSP}, d) \int_{[0,1]^d} f(x)^{(d-1)/d} dx \ \ c.c.,$$

where $\alpha(L_{TSP}, d)$ is the constant of the TSP.

Proof. Let P be a finite point set in $[0,1]^d$. Deleting an edge in an optimal TSP tour generates a 2MST, so $L_{2MST}(P) \leq L_{TSP}(P)$. Connecting the endpoints of an optimal 2MST by an edge of length at most $\sqrt{d}$ produces a feasible TSP tour. Thus,

$$|L_{TSP}(P) - L_{2MST}(P)| \leq \sqrt{d}.$$

The complete convergence can be shown in the same way as in the proof of Theorem 4.1 (i). □

In the remainder of this chapter we consider $b \geq 3$. For the proof of the main part of Theorem 8.1 we verify in Chapter 8.1 that the b-degree constrained minimal spanning tree functional is a subadditive and smooth Euclidean functional. Then we define in Chapter 8.2 its boundary modification and prove that the functional and the boundary functional are pointwise close. After that we show that the boundary functional is a superadditive and smooth Euclidean functional. So we can directly apply the following theorem by Redmond and Yukich [RY94] to the b-degree constrained MST, formulated by Yukich in [Yuk98].

Theorem 8.3. *[RY94] Let F and F_B be smooth subadditive and superadditive Euclidean functionals, respectively. Let $(X_i)_{i\geq 1}$ be iid random variables with values in $[0,1]^d$, $d \geq 2$. Assume*

$$|\,\mathbb{E}[F(X_1,\ldots,X_n) - \mathbb{E}[F_B(X_1,\ldots,X_n)]| = o(n^{(d-1)/d}).$$

Then

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

where f is the density of the absolutely continuous part of the law of X_1.

8.1 Properties of the bMST Functional

In this section we analyze the properties of the b-degree constrained MST, particularly with regard to the conditions in Theorem 2.8, where the functional L_{bMST} will taking the role of F.

Lemma 8.4. *L_{bMST} is a subadditive and smooth Euclidean functional.*

Proof. It is obvious that L_{bMST} has the translation invariance, homogeneity and normalization properties, and it is also easy to see that the functional is subadditive: consider a finite set P, a d-dimensional rectangle R with diameter $\text{diam}(R)$, a partition of R into two rectangles $R = R_1 \cup R_2$ and let $bMST_1$ and $bMST_2$ be optimal b-degree constrained minimal spanning trees in R_1 respectively R_2. Each tree contains two leaves, vertices with degree 1, and the trees are merged into a single tree by connecting two leaves. The length of the used edge is at most $\text{diam}(R)$. So we have

$$L_{bMST}(P\cap R) \leq L_{bMST}(P\cap R_1) + L_{bMST}(P\cap R_2) + \text{diam}(R).$$

In the second part of the proof we show that the functional is smooth:

$$|L_{bMST}(P_1\cup P_2) - L_{bMST}(P_1)| \leq C|P_2|^{(d-1)/d}.$$

We begin with a bMST on P_1 and add a bMST on P_2. Each of the graphs contains at least two leaves, we connect the graphs by a single edge between two leaves of length at most $\sqrt{d}$. The resulting graph is a feasible b-degree constrained MST on $P_1 \cup P_2$. By Lemma 2.1 the total edge length of the added bMST on P_2 is at most $C|P_2|^{(d-1)/d}$, since the bMST is a subadditive Euclidean functional. Thus,

$$L_{bMST}(P_1\cup P_2) \leq L_{bMST}(P_1) + C|P_2|^{(d-1)/d}.$$

Now we start with a bMST on $P_1 \cup P_2$ and construct a bMST on P_1. All points of P_2 and edges incident with these points are deleted. The deletion generates at most $b|P_2|$ connected components, and each component is a tree. We choose a leaf of each tree and add a TSP tour through these leaves to the graph. An edge of the TSP tour has to be deleted to construct a feasible bMST on $P_1 \cup P_2$. The total length of the added TSP tour is bounded by $C|P_2|^{(d-1)/d}$ by Lemma 2.1. We have

$$L_{bMST}(P_1) \leq L_{bMST}(P_1 \cup P_2) + C|P_2|^{(d-1)/d},$$

hence, the functional is smooth. □

8.2 Properties of the Boundary bMST Functional

As in the boundary modification of the MDVRP functional considered before, edges along the boundary have length zero. So in a boundary bMST graph we have either bMSTs that are all connected to the boundary or a single bMST without a connection to the boundary, see Figure 8.1 and 8.2. Here is the formal definition of the boundary functional of the b-degree constrained MST: For all rectangles $R \subset \mathbb{R}^d$, finite point sets $P \subset R$ and points a on the boundary of R let $L'_{bMST}(P, a)$ denote the length of the minimal b-degree constrained spanning tree on $P \cup \{a\}$. The *boundary bMST functional* L^B_{bMST} is defined by

$$L^B_{bMST}(P) := \min\left\{ L_{bMST}(P),\ \inf\left\{ \sum_i L'_{bMST}(P_i, a_i) \right\} \right\},$$

where the infimum ranges over all sequences $(a_i)_{i\geq 1}$ of points on the boundary of R and all partitions $(P_i)_{i\geq 1}$ of P. We show that the boundary bMST functional is a good approximation of the bMST functional:

Lemma 8.5. *The b-degree constrained MST functional and its boundary functional are pointwise close:*

$$|L_{bMST}(P) - L^B_{bMST}(P)| \leq C|P|^{(d-2)/(d-1)}.$$

Proof. Since $L^B_{bMST}(P) \leq L_{bMST}(P)$, we only have to show that

$$L_{bMST}(P) \leq L^B_{bMST}(P) + C|P|^{(d-2)/(d-1)}.$$

We start with a graph associated to $L^B_{bMST}(P)$ and modify it into a feasible b-degree constrained MST by adding edges of total length at most $C|P|^{\frac{d-2}{d-1}}$: let B

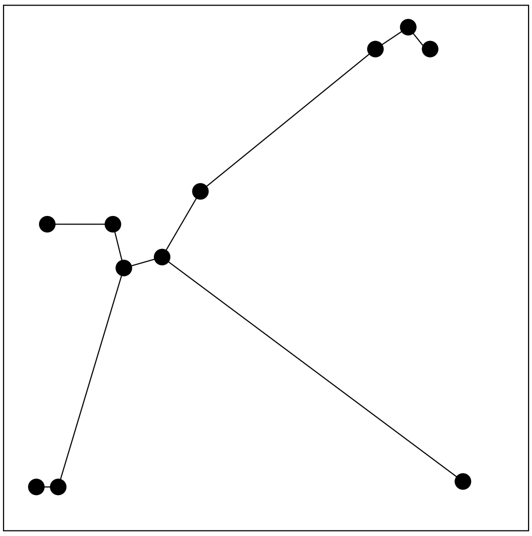

Figure 8.1: A 3-degree constrained MST.

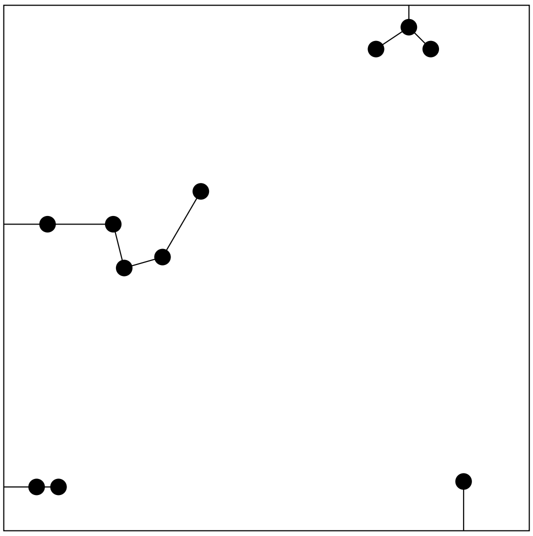

Figure 8.2: A boundary 3-degree constrained MST.

denote the set of points where the graph meets the boundary of $[0,1]^d$. Note that the vertices in B have degree 1. We add to the graph a TSP tour through B with edges lying on the boundary of $[0,1]^d$ and delete an arbitrary edge in order to construct a b-degree constrained MST(Note $b \geq 3$). Since the boundary of $[0,1]^d$ has dimension $d-1$ and the TSP functional is a subadditive Euclidean functional, the total length of the added MST is at most $C|B|^{(d-2)/(d-1)}$, see Lemma 2.1. Due to the fact that $|B| \leq |P|$, we have

$$L_{bMST}(P) \leq L^B_{bMST}(P) + C|P|^{(d-2)/(d-1)}$$

and the claim follows. □

The next lemma shows that the boundary bMST functional has the needed properties for Theorem 2.8. The boundary functional will be used as F^B in the theorem.

Lemma 8.6. *The boundary functional L^B_{bMST} of the b-degree constrained MST is a superadditive and smooth Euclidean functional.*

Proof. It is easy to verify that L^B_{bMST} has the translation invariance, homogeneity and normalization properties. Furthermore the functional is superadditive: consider a finite set P, a d-dimensional rectangle R with a partition into two rectangles $R = R_1 \cup R_2$ and let $bMST^B$ be an optimal boundary b-degree constrained minimal spanning tree in R. The restrictions of $bMST^B$ to R_1 and R_2 define boundary b-degree constrained minimal spanning trees in R_1 respectively

R_2, in case that the restrictions contain paths that start and end at the boundary one has to remove an arbitrary edge in the path. The restrictions are at least as large as $L^B_{bMST}(P \cap R_1)$ respectively $L^B_{bMST}(P \cap R_2)$. Thus,

$$L^B_{bMST}(P \cap R) \geq L^B_{bMST}(P \cap R_1) + L^B_{bMST}(P \cap R_2).$$

It remains to show that the functional is smooth:

$$|L^B_{bMST}(P_1 \cup P_2) - L^B_{bMST}(P_1)| \leq C|P_2|^{(d-1)/d}.$$

We start with a graph associated to $L^B_{bMST}(P_1 \cup P_2)$ and delete all points of P_2 and all edges incident with these points. The resulting graph consists of at most $b|P_2|$ connected components that are not connected to the boundary. These components are trees, so each of them contains vertices with degree 1. Choose a vertex with degree 1 in every component and add a TSP tour through these vertices(note that we are considering $b \geq 3$). Then we delete an arbitrary edge in the tour and choose a vertex with degree 1 in the component and connect it to the boundary in order to construct a feasible boundary bMST on P_1. The total length of all added edges is at most $C|P_2|^{(d-1)/d}$, since the TSP functional is a subadditive Euclidean functional, see Lemma 2.1. Thus,

$$L^B_{bMST}(P_1) \leq L^B_{bMST}(P_1 \cup P_2) + C|P_2|^{(d-1)/d}.$$

To show $L^B_{bMST}(P_1 \cup P_2) \leq L^B_{bMST}(P_1) + C|P_2|^{(d-1)/d}$, we begin with a graph associated to $L^B_{bMST}(P_1)$ and add a bMST on P_2 to the graph. A leaf of the bMST on P_2 and a leaf of the boundary bMST on P_1 are connected by an edge of length at most $\sqrt{d}$ in order to construct a feasible boundary bMST on $P_1 \cup P_2$. Since the bMST functional is a subadditive Euclidean functional, we have by Lemma 2.1 that $L_{bMST}(P_2) \leq C|P_2|^{(d-1)/d}$. Thus,

$$L^B_{bMST}(P_1 \cup P_2) \leq L^B_{bMST}(P_1) + C|P_2|^{(d-1)/d},$$

and all in all the assumption follows:

$$|L^B_{bMST}(P_1 \cup P_2) - L^B_{bMST}(P_1)| \leq C|P_2|^{(d-1)/d}.$$

□

8.3 Concluding Consequences

There are three immediate consequences of the properties of the bMST functional and its boundary version. First of all, Theorem 8.1. Secondly, we are able to apply a lemma by Redmond and Yukich to give a rate convergence of the mean of the bMST functional. At last, by Theorem 2.4 we have a concentration inequality for the functional.

Proof of Theorem 8.1. Putting the Lemmata 8.4, 8.5 and 8.6 together, we have shown that the bMST functional is a smooth and subadditive Euclidean functional which is close to its smooth and superadditive Euclidean boundary functional. We can directly apply Theorem 2.7 and get the assumption of Theorem 8.1. □

In the following we consider points that are given by iid random variables with uniform distribution. Remond and Yukich [RY94] have shown that boundary functionals are an ideal tool to provide rates of convergence of Euclidean functionals. The subadditive structure of a functional is not enough to prove rates of convergence, one gets only one-sided estimates. With the help of the boundary functional, the functional can be made superadditive and one can extract rates of convergence. The idea of modifying functionals to get a superadditive structure was known before the work of Redmond and Yukich, see e.g. Hammersley [Ham74], but they provide a general and simple approach. The formulation of the following lemma is from McGivney and Yukich [MY99b]:

Lemma 8.7 ([MY99b])**.** *Let $U_1, \dots, U_n$ be iid 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*

$$|\,\mathbb{E}\,L[(U_1, \dots, U_n)] - \mathbb{E}[L^B(U_1, \dots, U_n)]| \leq \beta(n),$$

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

$$|\,\mathbb{E}\,L[(U_1, \dots, U_n)] - \alpha(L, d)n^{(d-1)/d)}| \leq \max\{\beta(n), Cn^{(d-1)/2d}\}.$$

Corollary 8.8. *Let $P = \{P_1, \dots, P_n\}$ be a set of points in $[0,1]^d$ given by independent uniformly distributed random variables. The mean of the bMST functional satisfies*

$$|\,\mathbb{E}[L_{bMST}(P_1, \dots, P_n)] - \alpha(L_{bMST}, d)n^{(d-1)/d)}| \leq C|P|^{(d-2)/(d-1)}.$$

Proof. By Lemma 8.5 we have

$$|L_{bMST}(P) - L^B_{bMST}(P)| \leq C|P|^{(d-2)/(d-1)},$$

this clearly implies with the Jensen inequality

$$|\,\mathbb{E}[L_{bMST}(P)] - \mathbb{E}[L^B_{bMST}(P)]| \leq C|P|^{(d-2)/(d-1)}.$$

So we immediately obtain the rate result

$$|\,\mathbb{E}[L_{bMST}(P_1, \dots, P_n)] - \alpha(L_{bMST}, d)n^{(d-1)/d)}| \leq C|P|^{(d-2)/(d-1)}.$$

□

This results gives gives no indication how the functional is concentrated around its mean. But the functional satisfies the conditions of Theorem 2.4.

Corollary 8.9. *Let $P = \{P_1, \dots, P_n\}$ be a set of points in $[0,1]^d$ given by independent uniformly distributed random variables. There are constants C and C' such that for all $t > 0$:*

$$\mathbb{P}\left[\left|L_{bMST}(P_1, \dots, P_n) - \mathbb{E}\left[L_{bMST}(P_1, \dots, P_n)\right]\right| > t\right] \leq Ce^{-\frac{(t/C')^{2d/(d-1)}}{Cn}}.$$

Bibliography

[AHR$^+$92] A. Aggarwal, H. Edelsbrunner, P. Raghavan and P. Tiwari. Optimal time bounds for some proximity problems in the plane. *Information Processing Letters*, 42, 55–60, 1992.

[AY01] A. Aggarwal and P. S. Yu. Outlier Detection for High Dimensional Data. *SIGMOND*, 2001.

[Aro98] S. Arora. Polynomial time Approximation Schemes for Euclidean TSP and other Geometric problems. *J. ACM*, 45(5), 753–782, 1998.

[AC04] S. Arora and K. Chang. Approximation schemes for degree-restricted MST and red-blue separation problems. *Algorithmica*, 40(3),189–210, 2004.

[AB93] F. Avram and D. Bertsimas. On Central Limit Theorems in Geometrical Probability. *The Annals of Applied Probability*, 3(4), 1033–1046, 1993.

[BDS$^+$05] A. Baltz, D. Dubhashi, A. Srivastav, L. Tansini, and S. Werth. Probabilistic analysis for a multiple depot vehicle routing problem. In R. Ramanujam, Sandeep Sen(Eds.), *FSTTCS 2005: Foundations of Software Technology and Theoretical Computer Science*, volume 3821 of *Lecture Notes in Computer Science*, 360–371, 2005.

[BDS$^+$06] A. Baltz, D. Dubhashi, A. Srivastav, L. Tansini, and S. Werth. Probabilistic analysis of a Multidepot Vehickle Routing Problem. *Random Structures and Algorithms*, to appear.

[BHH59] J. Beardwood, J.H. Halton, and J. M. Hammersley. The shortest path through many points. In *Proc. Cambridge Philos. Soc.*, 55, 299–327, 1959.

[Bel62] R. E. Bellman. Dynamic programming treatment of the travelling salesman problem. *J. Assoc. Comput. Mach.*, 9,61–63, 1962.

[BR90] D. Bertsimas and G. van Ryzin. An asymptotic determination of the minimum spanning tree and minimum matching constants in geometrical probability. *Oper. Research Lett.*, 9,223–231, 1990.

[BDO06] A. Bompadre, M. Dror and J.B. Orlin. Probabilistic analysis of unit demand vehicle rounting problems. *Journal of Applied Probability*, to appear.

[Cha03] T. M. Chan. Euclidean Bounded-Degree Spanning Tree Ratios. In *Proc. 19th ACM Sympos. Computatinal Geometry*, 11–19, 2003.

[CGW93] I. M. Chao, B. L. Golden and E. Wasil. A new heuristic for the multi-depot vehicle routing problem that improves upon best-known solutions. *American Journal of Mathematical and Management Sciences*, 13, 371–406, 1993.

[Chr76] N. Christofides. Worst-case analysis of a new heuristic for the traveling salesman problem. Technical Report CS-93-13, Carnegie Mellon University, 1976.

[CDT+05] F. Cicalese, P. Damaschke, L. Tansini and S. Werth. Overlaps Help: Improved Bounds for Group Testing with Interval Queries. In L. Wang (Editor), *Computing and Combinatorics 2005* , volume 3595 of *Lecture Notes in Computer Science*, 935–945, 2005.

[CDT+06] F. Cicalese, P. Damaschke, L. Tansini and S. Werth. Overlaps Help: Improved Bounds for Group Testing with Interval Queries. *Discrete Applied Mathematics* , to appear.

[CGL97] J. F. Cordeau, M. Gendreau, and G. Laporte. A Tabu Search Heuristic for Periodic and Multi-Depot Vehicle Routing Problems. *Networks*, 30, 105–119, 1997.

[Cro58] G. A. Croes. A method for solving traveling salesman problems. *Oper. Res.*, 6,791–812, 1958.

[DR59] G. B. Dantzig and J. H. Ramser. The truck dispatching problem. *Management Science*, 6,80, 1959.

[DHW06] B. Doerr, N. Hebbinghaus, and S. Werth. Improved Bounds and Schemes for the Declustering Problem. *Theoretical Computer Science*, to appear.

[DHW04] B. Doerr, N. Hebbinghaus, and S. Werth. Discrepancy and Declustering. In J. Fiala, V. Koubek, J. Kratochvil (Eds.), *Mathematical Foundations of Computer Science 2004*, volume 3153 of *Lecture Notes in Computer Science*, 760–771, 2004.

[FJ81] M. L. Fisher and R. Jaikumar. A Generalized Assignment Heuristic for Vehicle Routing. *Networks,* 11,109-124, 1981.

[Flo56] M. M. Flood. The traveling-salesman problem. *Oper. Res.*, 4,61–75, 1956.

[GM74] B. E. Gillet and L. R. Miller. A heuristic algorithm for the vehicle dispatch problem. *Oper. Res.* 22, 340–349, 1974.

[GTV02] D. Giosa, L. Tansini and O. Viera. New Assignment Algorithms for the Multi-Depot Vehicle Routing Problem. *Journal of the Operational Research Society* 53(9), 977–984, 2002.

[Ham74] J. M. Hammersley. Postulates for subadditive processes. *Ann. Prob.*, 2, 652–680, 1974.

[HK62] M. Held and R. M. Karp. A dynamic programming approach to sequencing problems. *SIAM J. Appl. Math.*, 10,196–210, 1962.

[JMF99] A. Jain, M. Murty and P. Flynn. Data Clustering: A Review. *ACM Computing Surveys*, 31(3), 264–323, 1999.

[JM02] D. S. Johnson and L. A. McGeoch. Experimental Analysis of Heuristics for the STSP. *The traveling salesman problem and its variations*, G. Gutin and A. P. Punnen (eds), Kluwer Academic Publisher, 369–443, 2002

[K76] R. Karp. The probabilistic analysis of some combinatorial search algorithms. J.F. Traub (ed). *Algorithms and Complexity: New Directions and Recent Results*, Academic Press, New York, 1–19, 1976.

[K77] R. Karp. Probabilistic Analysis of Partitioning Algorithms for the Travelling Salesman Problem in the Plane. *Math. of Operations Research*, 2, 209–224, 1977.

[KS85] R. Karp and J. M. Steele. Probabilistic Analysis of Heuristics. In Lenstra et al. (eds) *The Travelling Salesman Problem*, John Wiley 1985.

[Led96] M. Ledoux. Isoperimetry and Gaussian Analysis. Ecole d'été de Probabilités de Saint-Flour(1994), *Lecture Notes in Mathematics*, Springer-Verlag, 1996.

[LK73] S. Lin and B. W. Kernighan. An effective heuristic algorithm for the traveling-salesman problem. *Operations Res.*, 21,498–516, 1973.

[M97] K. McGivney. Ph.D. Thesis, Department of Mathematics, Lehigh University, Bethelem, Pa, 1997.

[MY99a] K. McGivney and J. E. Yukich. Asymptotics for Voronoi tessellations on random samples. *Stochastic Processes and their Applications*, 83,2, 273–288, 1999.

[MY99b] K. McGivney and J. E. Yukich. Asymptotics for geometric location problems over random samples. *Adv. Appl. Prob.*, 31, 632–642, 1999.

[Mil70] R. Miles. On the homogeneous planar Poisson point process, *Math. Biosci.*, 6, 85–127, 1970.

[MS86] V. Milman and G. Schechtman. Asymptotic Theory of Finite Dimensional Normed Spaces. *Lecture Notes in Mathematics*, 1200, Springer Verlag, 1986.

[MS92] C. Monma and S. Suri. Transitions in geometric minimum spanning trees. *Discrete Comput. Geom.*, 8(3),265–293, 1992.

[NO97] K. Nakano and S. Olariu. An Optimal Algorithm for the Angle-Restricted All Nearest Neighbor Problem on the reconfigurable mesh, with Applications. *IEEE TPDS*, 8(9), 983–990, 1997.

[Pap78] C. H. Papadimitriou. The probabilisitic analysis of matching heuristics. *Proc. of the 15th Allerton Conf. on Communication, Control, and Computing*, 368–378, 1978.

[PV84] C. H. Papadimitriou and U. V. Vazirani. On two geometric problems related to the travelling salesman problem. *J. Algorithms*, 5(2),231–246, 1984.

[PY01] M. Penrose and J. Yukich. Central Limit Theorems for some Graphs in Computational Geometry. *Annals of Applied Prob.*, 11,4, 1005–1041, 2001.

[PS85] F. P. Preparata and and M. I. Shamos. *Computational Geometry*, Springer Verlag, 1985.

[Rag96] B. Raghavachari. Algorithms for Finding Low Degree Structures. , In *Approximation algorithms*, Dorit Hochbaum (ed.), PWS Publishers Inc., 266–295, 1996.

[RY94] C. Redmond and J. E. Yukich. Limit theorems and rates of convergence for Euclidean functionals. *Annals of Applied Prob.*, 4(4),1057–1073, 1994.

[RLB96] J. Renaud, G. Laporte and F. F. Boctor. A tabu search heuristic for the multi-depot vehicle routing problem. *Computers and Operations Research*, 23, 229–235, 1996.

[Rhe92] W. T. Rhee. On the traveling salesperson problem in many dimensions. *Random Structures and Algorithms*, 3,227–233, 1992.

[Rhe93] W. T. Rhee. A matching problem and subadditive Euclidean functionals. *Ann. Appl. Probab.*, 3(3),794–801, 1993.

[S00] M. Smid. Closest-Point Problems in Computational Geometry. *Handbook on Computational Geometry*, J.-R. Sack and J. Urrutia (eds), Amsterdam, Netherlands: North-Holland, 2000.

[Ste81] J. M. Steele. Subadditive Euclidean functionals and non-linear growth in geometric probability. *Ann. Prob.*, 9,365–376, 1981.

[Ste88] J. M. Steele. Growth rates of Euclidean minimal spanning trees with power weighted edges. *Ann. Prob.*, 16,1767–1787, 1988.

[Ste97] J. M. Steele. *Probability theory and combinatorial optimization*, volume 69 of *CBMS-NSF Regional Conference Series in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1997.

[Str65] V. Strassen. The existence of probability measures with given marginals. *Ann. Math. Statist.*, 36, 423–439,1965.

[Tal94] M. Talagrand. Matching theorems and empirical discrepancy computations using majorizing measures. *Journal of the American Mathematical Society*, 7, 455–537, 1994.

[Tal95] M. Talagrand. Concentration of measure and isoperimetric inequalities in product spaces. *Publ. Math. Institut des Hautes Études Scientifiques*, 81, 73–205, 1995.

[TV02] P. Toth and D. Vigo (eds.). *The Vehicle Routing Problem.*, SIAM Monographs on Disrecte Mathematics and Applications, SIAM, 2002.

[Wei78] B. Weide. *Statistical Methods in Algorithm design and Analysis*. Ph.D. Thesis, computer Science Department, Carnegie Mellon University, 1978.

[Yuk98] J. E. Yukich. *Probability theory of classical Euclidean optimization problems*, volume 1675 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 1998.