6. ISOPERIMETRY AND CONCENTRATION INEQUALITIES

6.1. Azuma's Inequality

The last chapter provided a rate of convergence for the means of subadditive and superadditive Euclidean functionals L^p and L^p_B, but it didn't tell us how these functionals are concentrated around their means in terms of deviation inequalities. This chapter will show that smoothness conditions lead to sharp concentration estimates via isoperimetric methods. We begin by reviewing some of the deviation estimates which are basic to our subject.

When $d = 2$, Steele (1981b) showed by means of the jackknife inequality of Efron and Stein that the variance of $T(U_1, ..., U_n)$ is bounded independently of n. In fact, writing $T(n)$ for $T(U_1, ..., U_n)$ where U_i, $i \geq 1$, are i.i.d. with the uniform distribution on $[0,1]^2$, he proved that

$$(6.1) \qquad \sum_{n=1}^{\infty} P\{|T(n)/n^{1/2} - \alpha(T_B^1, 2)| > \epsilon\} < \infty$$

for all $\epsilon > 0$. We recall that (6.1) expresses the *complete convergence* of $T(n)/n^{1/2}$ and is of course stronger than a.s. convergence. Steele (1981b) was motivated to show (6.1) in order to rigorously justify Karp's (1976,1977) algorithm for the traveling salesman problem under the *independent model of problem generation*, a topic described at the beginning of Chapter 4.

The estimate (6.1) provides a weak deviation estimate for the TSP functional about its mean. In this chapter we show that many Euclidean functionals enjoy similar and even more refined concentration inequalities. It might be anticipated that estimates of the type (6.1) hold for Euclidean functionals which are reasonably well controlled and which do not change too much when the underlying vertex set is perturbed. This is indeed the case and we will see shortly that (6.1) admits a generalization to all *smooth* Euclidean functionals.

Deviation estimates which are more refined than (6.1) may often be established via martingale methods, as first discovered by Rhee and Talagrand (1987). This approach has the advantage of versatility and applies to a wide range of Euclidean functionals.

We first recall the martingale difference sequence representation of an arbitrary random variable $X \in L^1(\Omega, \mathcal{A}, P)$. Given a filtration

$$(\emptyset, \Omega) = \mathcal{A}_0 \subset \mathcal{A}_1 \subset ... \subset \mathcal{A}_n = \mathcal{A}$$

of σ-algebras of $\mathcal{A}$, let $E(X|\mathcal{A}_i)$ denote the conditional expectation of X with respect to $\mathcal{A}_i$. For each $1 \leq i \leq n$ define the martingale difference

$$d_i := E(X|\mathcal{A}_i) - E(X|\mathcal{A}_{i-1})$$

so that $X - EX = \sum_{i=1}^n d_i$. This generic representation of $X - EX$ will allow us to express the high concentration of X around its mean in terms of the size of the differences d_i.

Let $\|d_i\|_\infty$ denote the essential supremum of d_i. The following fundamental inequality, due to Azuma (1967), provides deviation bounds for the above martingale decomposition.

Theorem 6.1. *(Azuma's inequality) For all $t > 0$*

$$P\{|\sum_{i=1}^n d_i| \geq t\} \leq 2exp\left(\frac{-t^2}{2\sum_{i=1}^n \|d_i\|_\infty^2}\right).$$

Proof. We first note that when X is a mean zero random variable such that $|X| \leq 1$ almost surely, then for any real a

$$E\exp(aX) \leq \exp(a^2/2).$$

To see this, note from the convexity of $f(x) = \exp(ax)$ and from $ax = a(1+x)/2 - a(1-x)/2$ that, for any real $|x| \leq 1$ we have

$$\exp(ax) \leq \cosh a + x\sinh a.$$

Taking expectations and using the elementary bound $\cosh x \leq \exp(x^2/2)$ we obtain the claim. It follows for any $i = 1, ..., n$ that

$$E(\exp(ad_i)|\mathcal{A}_{i-1}) \leq \exp(\frac{a^2}{2}\|d_i\|_\infty^2).$$

Iterating this inequality and using the properties of conditional expectation we get

$$\begin{aligned}
E\exp(t\sum_{i=1}^n d_i) &= E\left(E\exp(t\sum_{i=1}^n d_i)|\mathcal{A}_{n-1}\right) \\
&= E\left(\exp(t\sum_{i=1}^{n-1} d_i)\ E(\exp(td_n)|\mathcal{A}_{n-1})\right) \\
&\leq E\exp(t\sum_{i=1}^{n-1} d_i)\ \exp(\frac{t^2}{2}\|d_n\|_\infty^2) \\
&\dots \\
&\leq \exp(\frac{t^2}{2}\sum_{i=1}^n \|d_i\|_\infty^2).
\end{aligned}$$

From Markov's inequality we obtain for all $t > 0$

$$P\{\sum_{i=1}^n d_i > \lambda\} \leq \exp(-\lambda t)\exp(\frac{t^2}{2}\sum_{i=1}^n \|d_i\|_\infty^2).$$

Letting $t = \lambda(\sum_{i=1}^n \|d_i\|_\infty^2)^{-1}$ we obtain

$$P\{\sum_{i=1}^n d_i > \lambda\} \le \exp\left(\frac{-\lambda^2}{2\sum_{i=1}^n \|d_i\|_\infty^2}\right).$$

Applying this inequality to the sum $-\sum_{i=1}^n d_i$ yields Azuma's inequality as desired. □

There are several ways to refine Azuma's inequality and they depend largely on bounds for $E(d_i^2|\mathcal{A}_{i-1})$ and $\max_{i\le n} i\|d_i\|_\infty$. We refer to Ledoux and Talagrand (1991) for a complete treatment. Azuma's inequality and the generic representation $X - EX = \sum_{i=1}^n d_i$ are used in many contexts, especially in studying the deviations of $\|\sum_{i=1}^n X_i\|$ where X_i, $i \ge 1$, are independent Banach space valued random variables; see Ledoux and Talagrand (1991).

Our immediate interest is the application of Azuma's inequality to problems in geometric probability. To illustrate, let U_j, $j \ge 1$, be i.i.d. random variables with the uniform distribution on $[0,1]^d$ and let $\mathcal{A}_i$ denote the σ-field generated by the random variables $U_1, ..., U_i$. Given a Euclidean functional L^p, we abbreviate notation and write $L^p(n)$ for $L^p(U_1, ..., U_n)$. Consider the martingale differences

$$d_i := E\left(L^p(n)|\mathcal{A}_i\right) - E\left(L^p(n)|\mathcal{A}_{i-1}\right)$$

and notice that $L^p(n) - EL^p(n)$ admits the martingale decomposition

$$L^p(n) - EL^p(n) = \sum_{i=1}^n d_i.$$

When L is the shortest tour functional T, then the martingale increments satisfy the bound

$$\|d_i\|_\infty \le C(d)(n-i+1)^{-1/d}, \quad d \ge 2,$$

as shown by Rhee and Talagrand (1987). From this they easily obtain a deviation inequality for $T(n)$:

$$P\{|T(n) - ET(n)| > t\} \le \begin{cases} 2\exp(-ct^2/\log n), & d = 2 \\ 2\exp(-ct^2/n^{(d-2)/d}), & d \ge 3. \end{cases} \tag{6.2}$$

When seeking quick and easy deviation inequalities for Euclidean functionals, Azuma's inequality often suffices once bounds for the martingale differences $(d_i)_{i\ge1}$ are in hand. For example, Talagrand (1991) uses a modification of this approach to find deviation bounds for the *directed* TSP, a topic addressed in Chapter 8.

However, the method of martingale differences, while general and simple, does not always yield optimal tail estimates, especially for problems in geometric probability. We will now use Azuma's inequality to develop improved estimates. These improved estimates, which are concentration inequalities obtained via isoperimetric methods, are crucial to our approach.

6.2. The Rhee and Talagrand Concentration Inequalities

The seminal work of Talagrand (1995, 1996a) develops general isoperimetric inequalities for product measures which can be used to refine Theorem 6.1. Talagrand (1989,1994c) was motivated to develop his isoperimetric and concentration inequalities in order to investigate and settle various open problems related to sums of independent vector valued random variables. His results are a remarkable illustration of the power of abstract concentration of measure ideas; these ideas trace back to the work of V. Milman on the local theory of Banach spaces and in fact originate with Milman's proof of Dvoretzky's theorem on almost spherical sections of convex sets.

Loosely speaking, the concentration of measure phenomenon, which is at the heart of Talagrand's isoperimetric inequalities, says that if (X, d, μ) is a compact metric space, and the Borel set $B \subset X$ has a μ measure of at least one half, then "most" of the points in X are "close" to B. Talagrand's main contribution is to clarify in a mathematical sense the meaning of the words "most" and "close" for some natural families of spaces (X, d, μ). For example, if X is the Euclidean n-sphere S^n, d is the geodesic distance, μ is the normalized rotationally invariant measure, and B_ϵ represents the ϵ fattening of the Borel set B, $\mu(B) \geq 1/2$, then the words "most" and "close" mean that

$$\mu(B_\epsilon) \geq 1 - (\frac{\pi}{8})^{1/2} \exp(-\epsilon^2(n-1)/2). \tag{6.3}$$

This inequality follows from Lévy's isoperimetric inequality; Talagrand's work has developed a new approach to providing isoperimetric inequalities which hold in more general settings. It may be shown that (6.3) implies that any "nice" function f on S^n, i.e, $f \in C(S^n)$, is close to a constant (its median value) everywhere except on a set whose measure is of the order $\exp(-\epsilon^2(n-1)/2)$. In other words, to use Milman's words, *a well-behaved function is "almost" a constant on "almost" all of the space.* This is the powerful concentration of measure phenomenon.

In a series of brilliant papers, Talagrand (1995,1996a,1996b) developed an entirely new set of isoperimetric inequalities which depends heavily upon novel ways to enlarge or fatten a set. Talagrand applies his inequalities to give a large number of applications in geometric probability, probability in Banach spaces, and percolation. We refer to Ledoux (1996) for a thorough and completely accessible treatment of isoperimetry.

Using his isoperimetric approach, Talagrand (1995) shows that when the Euclidean functional L is based on the TSP, MST, or Steiner MST problem, then $L(n) := L(U_1, ..., U_n)$ is concentrated around its mean in a remarkable way and actually exhibits sub-Gaussian behavior:

$$P\{|L(n) - EL(n)| \geq t\} \leq C \exp(-t^2/C). \tag{6.4}$$

Notice that (6.4) goes well beyond the classical approach of Theorem 6.1, which depends upon the martingale difference method.

When L is the minimal matching functional S, then it is unclear whether the sub-Gaussian estimate (6.4) holds. This is due to the apparent lack of good regularity properties in minimal matching. Knowledge of the behavior of S on a set F does not in general tell much about the behavior of S on a modification of F. Indeed, small changes in F could lead to drastic changes in $S(F)$. Another approach, due to Rhee (1994b), gives the best currently available concentration estimates for S. The first of Rhee's results holds for the two dimensional case. As is customary by now, we let $S(n)$ denote $S(U_1, ..., U_n)$.

Theorem 6.2. *(concentration for minimal matching, $d = 2$) There is a constant $0 < C < \infty$ such that for all $t > 0$*

$$P\{|S(n) - ES(n)| \geq t\} \leq C\exp(-t^2/C(\log n)^2). \tag{6.5}$$

A simple consequence of (6.5) is that the variance of $S(n)$ is at most $C(\log n)^2$. An apparently difficult question is whether the variance of $S(n)$ is bounded independently of n.

When the dimension d exceeds 2, Rhee (1993b) obtained sharper deviation estimates for $S(n)$. Her estimate is a consequence of a general deviation inequality for Euclidean functionals. We now discuss this important inequality, which forms the cornerstone of the entire theory.

We have seen that isoperimetric methods essentially show that "well-behaved" functions are close to their medians. In the setting of Euclidean functionals, it is reasonable to expect that if a functional L^p is smooth of order p (Hölder continuous) in the usual sense

$$|L^p(F \cup G) - L^p(F)| \leq (\text{card} G)^{(d-p)/d},$$

then it is well behaved and thus close to its mean. This is indeed the case as shown by the following general result, which applies to Euclidean functionals on sample points which need not be uniform.

Theorem 6.3. *(concentrations for Euclidean functionals, $d \geq 2$) Let X_i, $i \geq 1$, be independent random variables with values in $[0,1]^d$, $d \geq 2$. Let L^p, $0 < p < d$, be a Euclidean functional which is smooth of order p (3.8). Then there is a constant $C := C(L^p, d)$ such that for all $t > 0$ we have*

$$\begin{aligned} &P\{|L^p(X_1, ..., X_n) - EL^p(X_1, ..., X_n)| > t\} \\ &\leq C\exp\left(\frac{-(t/C_3)^{2d/(d-p)}}{Cn}\right). \end{aligned} \tag{6.6}$$

As we will soon see, Rhee (1993b) deduces (6.6) as a consequence of Azuma's inequality and an isoperimetric inequality for the Hamming distance. The upshot of (6.6) is that with high probability the functional $L^p(X_1, ..., X_n)$ and its mean $EL^p(X_1, ..., X_n)$ do not differ by more than $C(n \log n)^{(d-p)/2d}$. One of the most useful consequences of Rhee's concentration estimate (6.6) is that it reduces the problem of showing complete convergence of L^p to one of showing the convergence of the mean of L^p, a fact which we used heavily in the proof of the basic limit theorems of Chapter 4. We will also draw on this fact in the proof of our basic umbrella theorem of Chapter 7.

Corollary 6.4. *(convergence of means implies complete convergence) Let X_i, $i \geq 1$, be i.i.d. random variables with values in $[0,1]^d$ and law μ_X. Let L^p be a smooth Euclidean functional of order p, $0 < p < d$. If the mean of L^p converges in the sense that*

$$\lim_{n\to\infty} EL^p(X_1, ..., X_n)/n^{(d-p)/d} = \alpha(L^p_B, d, \mu_X),$$

then

$$\lim_{n\to\infty} L^p(X_1, ..., X_n)/n^{(d-p)/d} = \alpha(L^p_B, d, \mu_X) \quad c.c. \tag{6.7}$$

Proof. For all $\epsilon > 0$ we have by (6.6)

$$\sum_{n=1}^{\infty} P\left\{\left|\frac{L^p(X_1, ..., X_n) - EL^p(X_1, ..., X_n)}{n^{(d-p)/d}}\right| > \epsilon\right\}$$
$$\leq C \sum_{n=1}^{\infty} \exp\left(-(\frac{\epsilon}{C_3})^{2d/(d-p)} \frac{n}{C}\right).$$

Thus $\left|\frac{L^p(X_1,...,X_n) - EL^p(X_1,...,X_n)}{n^{(d-p)/d}}\right|$ converges completely to zero and the proof is complete. □

Another consequence of Theorem 6.3 is that, together with the rate results of Chapter 5, we can deduce high probability rate results for the functional $L^p(U_1, ..., U_n)$. Combining Theorem 6.3 and (5.3), for example, gives for all $d \geq 3$ and $0 < p < d$ the high probability estimate

$$|L^p(U_1, ..., U_n) - \alpha(L^p_B, d) n^{(d-p)/d}| \leq C \left(n^{(d-p-1)/d} \vee (n \log n)^{(d-p)/2d}\right).$$

6.3. Isoperimetry

In this section we describe the isoperimetric methods which are behind the proof of Theorem 6.3 and which form one of the central themes of this monograph. The approach is guided by Steele (1997).

The path to estimates of the type (6.3) and (6.4) proceeds via isoperimetric inequalities. While there are many such inequalities we will focus on that which involves the Hamming distance H on n-fold product spaces Ω^n.

The Hamming distance H on Ω^n measures the distance between x and y by the number of coordinates in which x and y disagree, that is

$$H(x,y) := \text{card}\{i : x_i \neq y_i\}.$$

To formulate our isoperimetric inequality, we will take $\Omega := [0,1]^d$ and μ a measure on $[0,1]^d$; μ^n denotes the product measure on the product space $([0,1]^d)^n$. Given $A \subset \Omega^n$ and $y \in \Omega^n$, we define the Hamming distance between y and A by

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

We will assume that A satisfies $\mu^n(A) \geq 1/2$.

Changing one of the n coordinates of y produces a change of at most 1 in $H(x,y)$ and therefore the martingale differences d_i, $1 \leq i \leq n$, appearing in the martingale difference representation of $\phi_A(y)$ are bounded by 1. By Azuma's inequality it follows that

$$\mu^n(y : \ |\phi_A(y) - \alpha| \geq t) \leq 2\exp(-t^2/2n), \tag{6.8}$$

where $\alpha := \int \phi_A(y) d\mu^n$. Using $\phi_A(y) = 0$ for $y \in A$, we see that the left side of (6.8) is at least as large as $\mu^n(A)$ when $t = \alpha$. Since $\mu^n(A) \geq 1/2$ it follows that $1/2 \leq 2\exp(-\alpha^2/2n)$ or $\alpha \leq (2n \log 4)^{1/2}$. By (6.8) it follows that

$$\mu^n(\phi_A(y) \geq t + (2n \log 4)^{1/2}) \leq 2\exp(-t^2/2n)$$

and therefore for $t \geq 2(2n \log 4)^{1/2}$ we obtain

$$\begin{aligned} \mu^n(\phi_A(y) \geq t) &= \mu^n\left(\phi_A(y) \geq t - (2n\log 4)^{1/2} + (2n \log 4)^{1/2}\right) \\ &\leq 2\exp(-t^2/8n) \end{aligned} \tag{6.9}$$

since $(t - (2n \log 4)^{1/2})^2 \geq (t/2)^2$ when t is in the range $2(2n\log 4)^{1/2} \leq t < \infty$. In the range $0 \leq t \leq 2(2n \log 4)^{1/2}$ we have $2\exp(-t^2/8n) \geq 1/2$ and so (6.9) holds for all $0 \leq t < \infty$ provided that the coefficient 2 is replaced by 4.

We have thus proved

Proposition 6.5. *(isoperimetry) If $A \subset \Omega^n$ satisfies $\mu^n(A) \geq 1/2$ then*

$$\mu^n(\{y \in \Omega^n : \phi_A(y) \geq t\}) \leq 4\exp(-t^2/8n). \tag{6.10}$$

Proposition 6.5 is an isoperimetric inequality for the Hamming distance H. This inequality goes back to Milman and Schechtman (1986) and its relevance to optimization problems was recognized by Rhee (1993b, Proposition 3). For further refinements of (6.10) with sharper constants we refer to Talagrand (1995). To understand how Proposition 6.5 captures isoperimetry and to see its relation to the classic inequality (6.3), consider the following. As is customary in isoperimetry, we define the t-enlargement of $A \subset \Omega^n$:

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

Proposition 6.5 now says that if $\mu^n(A) \geq 1/2$ then

$$\mu^n(A_t) \geq 1 - 4\exp(-t^2/8n). \tag{6.11}$$

If $t \geq Cn^{1/2}$ then (6.11) implies that the measure of the fattened set A_t is almost 1. We will use both (6.10) and (6.11) in the sequel.

We now use Proposition 6.5 to prove Theorem 6.3.

Proof of Theorem 6.3. We will closely follow Steele's (1997) exposition of Rhee's (1993b) proof. Let $Z := L^p(X_1, ..., X_n)$ and let $M := M(n)$ denote a median of Z. If $A := \{x = (x_1, ..., x_n) \in ([0,1]^d)^n : L^p(x) \leq M\}$ then for each $y = (y_1, ..., y_n) \in ([0,1]^d)^n$ there is an $x \in A$ such that $H(x,y) := \phi_A(y)$. Given y we let $F := F(y,x)$ denote the coordinates of x agreeing with those in y, i.e.

$$F := \{x_i,\ 1 \leq i \leq n :\ x_i = y_i\}.$$

We let $G := G(y,x)$ denote the remaining coordinates, i.e., $G := \{x_i, 1 \leq i \leq n : x_i \neq y_i\}$. Note that $\mathrm{card}G = \phi_A(y)$.

By the assumed smoothness, we have that both $|L^p(y) - L^p(F)|$ and $|L^p(F) - L^p(x)|$ are bounded by $C_3(\mathrm{card}G)^{(d-p)/d}$. Writing

$$L^p(y) \leq |L^p(y) - L^p(F)| + |L^p(F) - L^p(x)| + L^p(x),$$

we obtain

$$L^p(y) \leq 2C_3\phi_A(y)^{(d-p)/d} + M$$

since $L^p(x) \leq M$.

Letting μ be the law of X_1 we obtain by Proposition 6.5

$$\begin{aligned} P\{Z \geq M + t\} &= \mu^n(\{y \in \Omega^n : L^p(y) \geq M + t\}) \\ &\leq \mu^n(\phi_A(y)^{(d-p)/d} \geq \frac{t}{2C_3}) \\ &\leq 4\exp\left(\frac{-t^{2d/d-p}}{8n(2C_3)^{2d/d-p}}\right). \end{aligned}$$

Similar arguments show that $P\{Z \le M - t\}$ satisfies the same bound and therefore

$$P\{|Z - M| \ge t\} \le 8\exp\left(\frac{-t^{2d/d-p}}{8n(2C_3)^{2d/d-p}}\right). \tag{6.12}$$

Integrating this tail bound it follows that $E|Z - M| \le Cn^{(d-p)/2d}$ and thus $|EZ - M| \le Cn^{(d-p)/2d}$. Consequently, by (6.12) we have

$$P\{|Z - EZ| \ge t + Cn^{(d-p)/2d}\} \le 8\exp\left(\frac{-t^{2d/d-p}}{8n(2C_3)^{2d/d-p}}\right). \tag{6.13}$$

We convert (6.13) into a tail bound for $|Z - EZ|$. When $t \ge 2Cn^{(d-p)/2d}$ we write $t = t - Cn^{(d-p)/2d} + Cn^{(d-p)/2d}$ and obtain

$$P\{|Z - EZ| \ge t\} \le 8\exp\left(\frac{-t^{2d/d-p}}{8n(4C_3)^{2d/d-p}}\right), \tag{6.14}$$

where we use $(t - Cn^{(d-p)/2d})^{2d/(d-p)} \ge (t/2)^{2d/d-p}$ for $t \ge 2Cn^{(d-p)/2d}$. In the range $0 \le t \le 2Cn^{(d-p)/2d}$ the right side of (6.14) is bigger than a positive constant and therefore $P\{|Z - EZ| \ge t\}$ is bounded by the right side of (6.14) divided by this constant. This is precisely the desired inequality (6.6). $\square$

6.4. Isoperimetry and the Power-Weighted MST

In this section we illustrate one of the many applications of isoperimetric methods. We show that isoperimetry may be used to strengthen Theorem 4.6 to a complete convergence result. Similar methods apply to Theorem 4.7 as well. Specifically we will prove:

Theorem 6.6. *Let U_i, $i \ge 1$, be i.i.d. with the uniform distribution on $[0,1]^d$. Then*

$$\lim_{n\to\infty} M^d(U_1, ..., U_n) = \alpha(M_B^d, d) \quad c.c. \tag{6.15}$$

Proof. The proof uses a simple variant of Proposition 6.5 to derive a concentration inequality for the power-weighted MST about its median. As before we set $\Omega = [0,1]^d$; we let μ^n denote the uniform measure on Ω^n.

We let $M := M(n)$ denote a median of $M^d(U_1, ..., U_n)$ and we let $A \subset \Omega^n$ consist of those n-tuples $x := \{x_1, ..., x_n\} \in \Omega^n$ for which

$$M^d(x) := M^d(x_1, ..., x_n) \ge M.$$

Recall that for all $t > 0$ the t-enlargement of A is

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

We define two more sets. Let $B \subset \Omega^n$ consist of those points $x := (x_1, ..., x_n)$ such that the edges in the MST graph on $\{x_i\}_{i=1}^n$ have length at most $C(\log n/n)^{1/d}$. Let

$$D := \{x = (x_1, ..., x_n) \in \Omega^n : \max_{j \le n} d(g_j, \{x_i\}_{i=1}^n) \le C(\log n/n)^{1/d}\}$$

where $\{g_j\}_{j=1}^n$ denotes grid points in $[0,1]^d$ and where $d(x, F)$ denotes the distance between the point x and the set F.

Since $\mu^n(A) \ge 1/2$ and since B and D are high probability sets for μ^n, we easily have $\mu^n(A \cap B \cap D) \ge 1/3$. By a simple variant of Proposition 6.5 it follows that if $\mu^n(A \cap B \cap D) \ge 1/3$, then

$$\mu^n(\{y \in \Omega^n : \ \phi_{A \cap B \cap D}(y) \ge t\}) \le 6\exp(-t^2/8n).$$

Therefore the enlarged set $(A \cap B \cap D)_{tn^{1/2}}$ occurs with high probability:

$$\mu^n((A \cap B \cap D)^c_{tn^{1/2}}) \le 6\exp(-t^2/8).$$

Now define $E := B \cap D \cap (A \cap B \cap D)_{tn^{1/2}}$ and note for all $\beta > 1$ that a suitable choice for C yields the high probability bound:

$$\mu^n(E^c) \le n^{-\beta} + 6\exp(-t^2/8).$$

If $x = (x_1, ..., x_n) \in E$ then there is a point $y = y(x) = (y_1, ..., y_n)$ in $A \cap B \cap D$ such that the following conditions hold: $H(x,y) \le tn^{1/2}$, y is close to x in the sense that $\max_{i \le n} d(x_i, \{y_j\}_{j=1}^n) \le C(\log n/n)^{1/d}$ and $\max_{i \le n} d(y_i, \{x_j\}_{j=1}^n) \le C(\log n/n)^{1/d}$, and the edges in the graph of the minimal spanning tree on y have length bounded by $C(\log n/n)^{1/d}$.

We now claim that for this choice of y we have

$$|M^d(x) - M^d(y)| \le Ct \log n/n^{1/2}.$$

Indeed, to see that

$$M^d(x) \le M^d(y) + Ct \log n/n^{1/2},$$

consider the coordinates of x which differ from the coordinates in y. Join these coordinates to points in $\{y_i\}_{i=1}^n$ at a cost of at most $H(x,y) \cdot \log n/n$. This produces a spanning graph G on the union $\{x_i\}_{i=1}^n \cup \{y_i\}_{i=1}^n$. To obtain a spanning graph on just $\{x_i\}_{i=1}^n$ we may modify the edges in G which are linked to those y coordinates which do not appear in x. Using the approach of Lemma 4.8 we see that this may be done at a cost of at most $H(x,y) \cdot \log n/n$, thus showing the desired inequality. The proof of the reverse inequality

$$M^d(y) \le M^d(x) + Ct \log n/n^{1/2}$$

holds for the same reasons and this proves the stated claim.

Therefore if $x \in E$ we have

$$M^d(x) \ge M^d(y) - |M^d(x) - M^d(y)| \ge M - Ct \log n/n^{1/2}.$$

Thus for all $0 \le t \le n^{1/2}/2$ it follows that

$$\begin{aligned}
&P\{M^d(U_1,...,U_n) \le M - Ct \cdot \log n/n^{1/2}\} \\
&\le \mu^n(E^c) \\
&\le n^{-\beta} + 6\exp(-t^2/8).
\end{aligned}$$

Using a similar argument for the reverse inequality

$$P\{M^d(U_1,...,U_n) \ge M + Ct \cdot \log n/n^{1/2}\}$$

we obtain for all $0 \le t \le n^{1/2}/2$

$$P\{|M^d(U_1,...,U_n) - M| \ge Ct \cdot \log n/n^{1/2}\} \le 2n^{-\beta} + 12\ \exp(-t^2/8).$$

Setting $t := \epsilon \cdot n^{1/2}/\log n$, where $\epsilon > 0$ is arbitrary but fixed, yields the concentration inequality

$$P\{|M^d(U_1,...,U_n) - M| \ge \epsilon\} \le 2n^{-\beta} + 12\exp(-\epsilon^2 n/8(\log n)^2). \tag{6.16}$$

The arbitrariness of ϵ and the Borel-Cantelli lemma imply that

$$\lim_{n\to\infty} |M^d(U_1,...,U_n) - M| = 0 \quad c.c.$$

Integrating (6.16) also shows that $\lim_{n\to\infty} E|M^d(U_1,...,U_n) - M| = 0$ and thus

$$\lim_{n\to\infty} |EM^d(U_1,...,U_n) - M| = 0.$$

Since Theorem 4.6 gives

$$\lim_{n\to\infty} |EM^d(U_1,...,U_n) - \alpha(M_B^d, d)| = 0$$

it follows from the triangle inequality that

$$\lim_{n\to\infty} M^d(U_1,...,U_n) = \alpha(M_B^d, d) \quad c.c.$$

as desired. □

6.5. Large Deviations

In this section we use the deviation estimate (6.4) and the superadditivity of the boundary TSP functional T_B to arrive at a large deviation principle for T_B. This principle will show that the sub-Gaussian tail behavior of (6.4) cannot be sharpened and in this way it further supports the conjecture that the TSP functional has an underlying Gaussian (normal) structure. The approach originates in discussions with Amir Dembo and Ofer Zeitouni and it is a pleasure to thank them for their ideas.

For all $1 \le i \le 4$, let U_{ij}, $j \ge 1$, be i.i.d. uniform random variables in the subsquare Q_i and let $N_i(n)$ be an independent Poisson random variable with parameter n. The superadditivity of T_B gives

$$T_B(U_{11}, ..., U_{1N_1(n)}, U_{21}, ..., U_{2N_2(n)}, U_{31}, ..., U_{3N_3(n)}, U_{41}, ..., U_{4N_4(n)})$$
$$\ge \sum_{i=1}^{4} T_B(U_{i1}, ..., U_{iN_i(n)}, Q_i).$$

Write $T_B(k) := T_B(U_1, ..., U_{N(k)})$ where U_i, $i \ge 1$, are i.i.d. uniform random variables with values in $[0,1]^d$ and $N(k)$ is an independent Poisson random variable with parameter k. Then the left side of the above is equal in distribution to $T_B(4n)$ and by scaling the right side is equal in distribution to $\frac{1}{2}\sum_{i=1}^{4} T_B^{(i)}(n)$, where $T_B^{(i)}(n)$ are independent copies of $T_B(n)$. It follows that for all $t > 0$

$$P\{T_B(4n) > t\} \ge P\{\frac{1}{2}\sum_{i=1}^{4} T_B^{(i)}(n) > t\}$$

and thus

$$P\{T_B(4n)/(4n)^{1/2} > t\} \ge P\left\{\frac{\sum_{i=1}^{4} T_B^{(i)}(n)}{4n^{1/2}} > t\right\}$$
$$\ge \left(P\{T_B(n)/n^{1/2} > t\}\right)^4.$$

More generally, for all positive integers m and n we have

$$P\left\{\frac{T_B(4^m n)}{(4^m n)^{1/2}} > t\right\} \ge \left(P\{\frac{T_B(n)}{n^{1/2}} > t\}\right)^{4^m}.$$

If we set for all $n \in \mathbb{N}$

$$\phi(n) := -\log\ P\{\frac{T_B(n)}{n^{1/2}} > t\}$$

then the above relation tells us that $\phi(4^m n) \le 4^m \phi(n)$. Homogenizing, we arrive at

$$\frac{\phi(4^m n)}{4^m n} \le \frac{\phi(n)}{n}$$

and more generally, letting $n = 4^k$ we have

$$\frac{\phi(4^{m+k})}{4^{m+k}} \le \frac{\phi(4^k)}{4^k} \tag{6.17}$$

for all positive integers m and k. If we now set $\alpha(j) := \phi(4^j)/4^j$ then it follows that $\alpha(j)$ is decreasing by (6.17). On the other hand, it follows by the sub-Gaussian estimate (6.4) that if $t > C_2 := \limsup_{j\to\infty} \frac{ET_B(4^j)}{2^j}$ then for j large

$$\begin{aligned}P\{T_B(4^j)/2^j > t\} &\le P\{|T_B(4^j) - ET_B(4^j)| > t \cdot 2^j - ET_B(4^j)\}\\ &\le P\{|T_B(4^j) - ET_B(4^j)| > (t - C_2) \cdot 2^j\}\\ &\le C \cdot \exp(-(t - C_2)^2 \cdot 4^j/C),\end{aligned}$$

and therefore for j large, $\alpha(j)$ is bounded below by a positive constant depending only on t. Since $\alpha(j)$ is decreasing, we get the limit

$$\lim_{j\to\infty} \alpha(j) = C(t),$$

where $C(t)$ depends only on t. We have thus shown the following large deviation principle:

Theorem 6.7. *For $t > C_2 := \limsup_{j\to\infty} \frac{ET_B(4^j)}{2^j}$ there is a positive finite constant $C(t)$ such that*

$$\lim_{j\to\infty} \frac{-\log P\{T_B(4^j)/2^j > t\}}{4^j} = C(t). \tag{6.18}$$

The importance of (6.18) is that it essentially shows that the sub-Gaussian tail behavior (6.4) cannot be improved and that it is of the correct order of magnitude. Theorem 6.7 complements the work of Rhee (1991), who obtains lower bounds for the tails of the TSP. Writing, as usual, $T(n)$ for $T(U_1, ..., U_n)$, Rhee (1991) proves that there is a universal positive constant C such that for all $0 < t \le C^{-1}n^{1/2}$ we have

$$P\{T(n) \le ET(n) - t\} \ge C^{-1}\exp(-t^2 C). \tag{6.19}$$

This shows that the sub-Gaussian estimate (6.4) is "sharp" in the setting of the TSP.

Still, Theorem 6.7 raises some additional questions. For example, can we prove an analogous result for the random variables $T_B(n)$ when n ranges over all integers and not just powers of 4? Is it possible to find the analog of (6.18) for the standard TSP functional T?

Since the sub-Gaussian estimate (6.4) also holds for the MST functional, it is clear that Theorem 6.7 also holds for the boundary MST functional M_B. However, it is not yet clear that it holds for the boundary minimal matching functional S_B.

We anticipate that Theorem 6.7 can be turned into a genuine large deviation principle. Given a Euclidean functional L on $\mathbb{R}^2$ of order 1, let $L(n) := L(U_1, ..., U_n)$ and let $A := A(L) := \sup_n L(n)/n^{1/2}$. We anticipate that the following large deviation principle holds:

Conjecture 6.8. *Let L be a subadditive Euclidean functional on $\mathbb{R}^2$ which is smooth of order 1 and pointwise close to the superadditive boundary functional L_B. There is a rate function $I(x)$ such that for all closed sets $F \subset [0, A]$ we have*

$$\limsup_{n\to\infty} \frac{\log P\{L(n)/n^{1/2} \in F\}}{n} \leq -\inf_{x\in F} I(x), \tag{6.20}$$

and for all open sets $O \subset [0, A]$ we have

$$\liminf_{n\to\infty} \frac{\log P\{L(n)/n^{1/2} \in O\}}{n} \geq -\inf_{x\in O} I(x). \tag{6.21}$$

Remark.

It is *a priori* not clear whether (6.20) and (6.21) give the correct rates of convergence. However, a little reflection shows that the rates are indeed the right ones, at least when L is the TSP functional T. Indeed, by the sub-Gaussian estimate (6.4) we obtain for all $t > \beta := \sup_n ET(n)/n^{1/2}$ that

$$P\{T(4^k)/2^k \geq t\} \leq C\exp\left(-(t-\beta)^2 4^k/C\right).$$

Therefore we have

$$\limsup_{k\to\infty} \frac{\log P\{T(4^k)/2^k \geq t\}}{4^k} \leq -(t-\beta)^2/C < 0.$$

By Rhee's lower bounds (6.19) we similarly obtain for all $t < \beta$ the lower estimate

$$\liminf_{k\to\infty} \frac{\log P\{T(4^k)/2^k \leq t\}}{4^k} \geq -C(\beta - t)^2 > -\infty.$$

Thus the rates given by (6.20) and (6.21) are of the right order.

Notes and References

1. Section 6.3 follows closely the exposition of Steele (1997) and Rhee (1993b), who treat the case $p = 1$. Talagrand (1996c) suggested the use of isoperimetry to obtain the complete convergence of the power-weighted MST. For extensions of Theorem 6.6 to non-uniform random variables see Yukich (1997b).

2. It would be useful to obtain a concentration estimate for subadditive Euclidean functionals L^p when $p \geq d$. This would extend Theorem 6.3.

3. Azuma's inequality is a standard tool and is useful in many contexts. Our treatment is based on Ledoux and Talagrand (1991).