Seedlings in the Theory of Shortest Paths

J. Michael Steele

Abstract

This article explores three developments that arise from the fundamental theorem of Beardwood, Halton, and Hammersley on the asymptotic behavior of the shortest path through n random points. The first development concerns the role of martingales in the theory of shortest paths, especially their role in large deviation inequalities. The second development concerns the use of Lipschitz spacefilling curves to obtain analytical bounds in the theory of the TSP, and it provides some bounds that refine those of Bartholdi and Platzman on the worst case performance of the spacefilling heuristic for the TSP. The final topic addresses the relationship between Karp's partitioning heuristic and the BHH theorem.

1. Introduction

In 1959 Beardwood, Halton, and Hammersley established the following theorem:

If $X_i, 1 \leq i < \infty$ are independent identically distributed random variables with bounded support in $\mathbb{R}^d$, then the length L_n under the usual Euclidean metric of the shortest path through the points $\{X_1, X_2, \ldots, X_n\}$ satisfies

$$n^{-(d-1)/d} L_n \to c_d \int_{\mathbb{R}^d} f(x)^{(d-1)/d}\, dx \quad \textit{almost surely.} \tag{1.1}$$

Here, $f(x)$ is the density of the absolutely continuous part of the distribution of the X_i.

This result has proved fruitful in most of the ways that are open to a mathematical discovery. In particular, it has lead to interesting applications, provoked useful generalizations and inspired new techniques of analysis. The intention of this article is to review and contribute to three developments associated with the Beardwood, Halton, Hammersley theorem.

The first development concerns the extent to which (1.1) can be complemented by large deviation results. This exploration leads us to consider some basic results of large deviations for martingales, particularly Azuma's inequality, for which we give two proofs. While exploring the relationship of the TSP martingale theory, we also examine the demands it places on results like the square function inequality of Burkholder, bounds on Hermite moments, and related ideas. In the course of the review we give new proofs of two inequalities of Rhee and Talagrand, and we examine essentially all of the available information concerning the tail of the probability distribution of L_n.

We next address the use of spacefilling curves in the analytical theory of the TSP. Such techniques are relatively new, but their simplicity and generality suggests that their use will grow. The fact that underlies this development is the existence of measure preserving transformations from $[0,1]$ onto $[0,1]^d$ that are Lipschitz of order $1/d$. A basic objective of Section 4 is to review the background of a problem of Platzman and Bartholdi on the ratio of the length of the tour provided by the spacefilling heuristic and the length of an optimal tour is bounded independently of n.

The third development concerns the role of (1.1) in Karp's polynomial time partitioning algorithm for the TSP. This topic is addressed briefly, but two results are reviewed that will make clear how one can show the effectiveness of Karp's algorithm without resort to the full force of (1.1).

In the concluding section, we discuss some open problems and promising research directions. Finally there are two appendices that stand somewhat apart from our basic themes. The first of these gives S. Lalley's previously unpublished proof of the Beardwood, Halton, Hammersley Theorem in $d = 2$ for random variables with the uniform distribution on $[0,1]^2$. This proof uses minimal machinery and illustrates a technique that is applicable to many related problems. The second appendix develops an inequality for martingales that R.E.A.C. Paley introduced for Walsh functions. Paley's old argument is examined for the suggestions it provides about how one might pursue large deviation inequalities for L_n without paying the price of bounds on L^∞ norms as demanded by Azuma's inequality.

2. Martingale Bounds for the TSP

For $X_i, 1 \le i < \infty$, independent and uniformly distributed in $[0,1]^d$, the length L_n of the shortest path through $\{X_1, X_2, \ldots, X_n\}$ is a random variable that we can show to be tightly concentrated about its mean. In $d = 2$, for example, we know that Var L_n is bounded independently of n. This fact is proved in Steele (1981b) by means of the jackknife inequality of Efron and Stein (1981), but one can provide a proof that offers considerably more potential for further development by following Rhee and Talagrand (1987) and introducing martingale arguments.

If $\mathbf{F}_k$ is the σ-field generated by $\{X_1, X_2, \ldots, X_k\}$ and

$$d_i = E(L_n \mid \mathbf{F}_i) - E(L_n \mid \mathbf{F}_{i-1}), \tag{2.1}$$

then $d_i, 1 \leq i \leq n$, is a sequence of martingale differences that satisfy

$$L_n - EL_n = \sum_{i=1}^{n} d_i. \tag{2.2}$$

This well-known representation is available for any integrable random variable, but there are features that make it particularly effective for L_n. The most central of these is that the d_i can be related to the change that takes place in L_n as one of the X_i is changed. In this respect, the analysis of L_n by means of martingale differences comes to rely on calculations that are quite close to those that made the jackknife inequality effective. By working out the details of the L^p theory associated with the martingale representation, we are led to some of the basic themes of martingale theory.

For each $1 \leq i \leq n$, let $L_n^{(i)}$ denote the length of the shortest path through $S_i = \{X_1, X_2, \ldots, X_{i-1}, \hat{X}_i, X_{i+1}, \ldots, X_n\}$ where the random variables $\{\hat{X}_i : 1 \leq i \leq n\}$ are independent, uniformly distributed and also independent of the variables in the set $S = \{X_i : 1 \leq i \leq n\}$. Since $E(L_n^{(i)} \mid \mathbf{F}_i) = E(L_n \mid \mathbf{F}_{i-1})$, we have the key observation that

$$d_i = E(L_n - L_n^{(i)} \mid \mathbf{F}_i). \tag{2.3}$$

Since one can build a path through S_i by following the minimal path through S and making a detour from X_j to $\hat{X}_i$ and back for some $j \neq i$, we have

$$L_n^{(i)} - L_n \leq 2 \min_{j:j\neq i} |\hat{X}_i - X_j|. \tag{2.4a}$$

By the same reasoning but starting from the optimal tour for S_i we have

$$L_n - L_n^{(i)} \leq 2 \min_{j:j\neq i} |X_i - X_j|, \tag{2.4b}$$

and, moreover, the right hand sides of (2.4a) and (2.4b) both have the same distribution. Next we note that simple geometric considerations as applied in Steele (1981b) give us a bound on the tail of these distributions:

$$P(\min_{j:j\neq i} |\hat{X}_i - X_j| \geq t) \leq Ae^{-Bnt^d}, \quad t > 0, \tag{2.5}$$

where $A = A_d$ and $B = B_d$ are constants that depend on d (but not on n or t). From (2.3), (2.4a,b), (2.5), and Jensen's inequality, we therefore find for any $p \geq 1$ that

$$E|d_i|^p \leq 4^p p d^{-1} A(nB)^{-p/d} \Gamma(p/d). \tag{2.6}$$

Finally, in terms of L^p norms, we find from Stirling's formula that

$$\|d_i\|_p \leq C_1(p/n)^{1/d} \tag{2.7}$$

where C_1 is a constant that depends only on $d \geq 2$.

Inequality (2.7) is a basic one for the theory of the traveling salesman problem. In particular, since the d_i are orthogonal random variables, we find that if we restrict attention to $p = 2$ and $d = 2$, then (2.7) completes the proof of the rather surprising uniform bound Var $L_n \leq 2A_2B_2^{-1}$ mentioned earlier.

For large p inequality (2.7) is not as effective as one would hope since from (2.4a,b) it is already immediate that the norms $\|d_i\|_\infty$ are bounded by $2d^{1/2}$. Still, by applying the argument used in (2.6) to the conditional probabilities (2.3), we can get a sharper bound on the $\|d_i\|_\infty$. In particular, if we relax the bounds (2.4a,b) to

$$|L_n^{(i)} - L_n| \leq 2 \min_{j:j>i} |\hat{X}_i - X_j| + 2 \min_{j:j>i} |X_i - X_j|,$$

then from (2.3) we find

$$|d_i| \leq 2E\{\min_{j:j>i} |\hat{X}_i - X_j|\} + 2E\left\{\min_{j:j>i} |X_i - X_j| \big| X_i\right\}. \tag{2.8}$$

Using $\|d_i\|_\infty \leq 2d^{1/2}$ to deal with $i = n$, we thus can find a constant C_2 that depends only on $d \geq 2$, so for all $1 \leq i \leq n$ we have

$$\|d_i\|_\infty \leq C_2(n - i + 1)^{-1/d}. \tag{2.9}$$

The beauty of (2.9) is that it permits us to use traditional martingale techniques to obtain reasonably sharp large deviation inequalities on $L_n - EL_n$. To develop one such inequality we first note that for any $y \geq 0$,

$$e^{xy} \leq \cosh y + x \sinh y \quad \text{for all} \quad |x| \leq 1, \tag{2.10}$$

because (2.10) trivially holds for $x \in \{-1, 0, 1\}$, e^{xy} is convex, and the right hand side is linear in x. If we now let $x = d_i/\|d_i\|_\infty$ and $y = t\|d_i\|_\infty$, we find for $1 \leq k \leq n$ that

$$\exp\left(t \sum_{i=1}^{k} d_i\right) \leq \prod_{i=1}^{k} \left(\cosh t\|d_i\|_\infty + \frac{d_i(\sinh t\|d_i\|_\infty)}{\|d_i\|_\infty}\right).$$

Taking expectations and using the fact that the d_i are martingale differences gives us

$$\begin{aligned} E\exp\left(t \sum_{i=1}^{k} d_i\right) &\leq \prod_{i=1}^{k} \cosh(t\|d_i\|_\infty) \\ &\leq \exp\left(\frac{t^2}{2} \sum_{i=1}^{k} \|d_i\|_\infty^2\right), \end{aligned} \tag{2.11}$$

where in the last inequality we used the elementary bound $\cosh x \le e^{x^2/2}$. From (2.11) and the fact that the right hand bound is an even function of t, we find for all $t \ge 0$ that

$$P\left(\left|\sum_{i=1}^{k} d_i\right| \ge \lambda\right) \le 2e^{-\lambda t} \exp\left(\frac{t^2}{2}\sum_{i=1}^{k} \|d_i\|_\infty^2\right), \tag{2.12}$$

so letting $t = \lambda\left(\sum_{i=1}^{k} \|d_i\|_\infty^2\right)^{-1}$ we conclude

$$P\left(\left|\sum_{i=1}^{k} d_i\right| \ge \lambda\right) \le 2\exp\left(-\lambda^2 \Big/ \left(2\sum_{i=1}^{k} \|d_i\|_\infty^2\right)\right). \tag{2.13}$$

This inequality is valid for any martingale difference sequence $\{d_i\}$, and it is due to Azuma (1967). When we apply (2.13) to our particular $\{d_i\}$ satisfying (2.1) we find a theorem which was established in the case for $d = 2$ in Rhee and Talagrand (1987).

THEOREM 2.1. *There is a constant C_3 depending only on d such that for all $n \ge 1$ and $\lambda > 0$ we have*

$$P\left(|L_n - EL_n| \ge \lambda\right) \le \begin{cases} 2\exp(-C_3\lambda^2/\log n) & \text{if } d = 2 \\ 2\exp(-C_3\lambda^2 n^{(2-d)/d}) & \text{if } d \ge 3. \end{cases} \tag{2.14}$$

The technique used to obtain Azuma's inequality (2.13) is apparently quite crude, and one might hope to do better in several ways. One natural idea is to try to generalize (2.10) to

$$e^{xy} \le xf(y) + g(y), \quad |x| \le 1,\ y \ge 0, \tag{2.15}$$

for f and g that might be more effective than sinh and cosh. To see why this idea does not succeed, we let $x = \pm 1$ in (2.15) and add the two resulting inequalities. We find that (2.15) forces the bound $\cosh y \le g(y)$, and thus no inequality like (2.15) serves us any better than that used in the argument leading to (2.13).

A second seedling concerning Azuma's technique and the TSP comes from viewing (2.10) as a separation of variables for the bivariate function e^{xy}. A classical approach to such separation might call on the generating function for Hermite polynomials:

$$G(x, y) = e^{2xy - y^2} = \sum_{n=0}^{\infty} \frac{H_n(x) y^n}{n!}. \tag{2.16}$$

This approach has not been developed very far, but it seems rich enough to deserve a brief digression. Because of the basic orthogonality relation

$$\int_{-\infty}^{\infty} H_m(x)H_n(x)e^{-x^2}\,dx = \sqrt{\pi}2^n n!\,\delta_{mn},$$

it is not difficult to give a condition on the L^2 norm of $H_n(L_n - EL_n)$ that implies a large deviation inequality of Gaussian type. In fact it suffices to assume that

$$EH_n^2(Z) \le A^n n!$$

for some constant A.

Before closing this digression on separation of variables in e^{xy}, we should note that (2.16) is closely related to (2.10); in particular from (2.16) we easily find expressions for $\sinh y$ and $\cosh y$ in terms of odd and even Hermite polynomial (see e.g. Section 8.957 of Gradshteyn and Ryzhik 1963). Still, because of special properties of Hermite polynomials such as their recursion relation, one might expect some progress through Lemma 2.1.

Returning to the direct exploration of large deviation inequalities, we should note their easy application to moments. Thus, we multiply (2.13) by $p\lambda^{p-1}$ and integrate over $[0,\infty)$ to find for $p \ge 1$ that

$$E\left|\sum_{i=1}^{k} d_i\right|^p \le 2p\Gamma(p/2)\left\{2\sum_{i=1}^{k}\|d_i\|_\infty^2\right\}^{p/2},$$

or, in terms of norms,

$$\left\|\sum_{i=1}^{k} d_i\right\|_p \le C_4 p^{1/2}\left(\sum_{i=1}^{k}\|d_i\|_\infty^2\right)^{1/2}, \tag{2.17}$$

where C_4 is a universal constant which does not even depend on d. When (2.17) is specialized to $\{d_i\}$ satisfying (2.3), we find from (2.9) that for all $n \ge 1$

$$\|L_n - EL_n\|_p \le \begin{cases} C_5 p^{1/2}(\log n)^{1/2} & \text{if } d = 2, \\ C_5 p^{1/2} n^{(d-2)/(2d)} & \text{if } d \ge 3. \end{cases} \tag{2.18}$$

Inequality (2.8) can be obtained in another way that also provides an interesting proof of Azuma's inequality. The key idea comes from work of Jakubowski and Kwapień (1979), and, in our context, the main point is that if we let $r_k(s)$ be the kth Rademacher function (i.e. $r_k(s) = \operatorname{sign}(\sin 2^k\pi s),\ 0 \le s \le 1$) then

$$f(\omega, s) = \prod_{k=1}^{n}\left(1 + \frac{r_k(s)d_k(\omega)}{\|d_k\|_\infty}\right) \tag{2.19}$$

is a density function with respect to the product measure $ds\,dP$. The identities that make (2.19) effective are

$$\sum_{k=1}^{n} \frac{a_k d_k(\omega)}{\|d_k\|_\infty} = \int_0^1 \sum_{k=1}^{n} a_k r_k(s) f(\omega, s)\, ds \tag{2.20a}$$

and

$$1 = \int f(\omega, s)\, dP. \tag{2.20b}$$

The proof of (2.20a) just requires expanding (2.19) and using the fact that the Rademacher functions $\{r_k(s)\}_{1 \le k \le n}$ have mean zero and variance 1. Similarly, (2.20b) follows from expanding (2.19) and using the martingale property. Since $F(\omega, s) \ge 0$ we also see from (2.20b) that $f(\omega, s)$ must indeed be a density with respect to $ds\,dP$.

To get our second proof of Azuma's inequality we first apply Jensen's inequality in (2.20a), integrate with respect to P, and change order of integration:

$$\begin{aligned} E \exp\Bigl(t \sum_{k=1}^{n} a_k d_k \Big/ \|d_k\|_\infty\Bigr) &\le E\left(\int_0^1 \exp\Bigl(t \sum_{k=1}^{n} a_k r_k(s)\Bigr) f(\omega, s)\, ds\right) \\ &= \int_0^\infty \exp\Bigl(t \sum_{k=1}^{n} a_k r_k(s)\Bigr)\, ds \\ &= \prod_{k=1}^{n} \cosh(t a_k). \end{aligned} \tag{2.21}$$

In the second line of (2.21) we used (2.20b), and in the last we used the fact that the r_k are Bernoulli random variables. If we now let $a_k = \|d_k\|_\infty$ in (2.21), we find that (2.21) reduces to the same bound as (2.11), so one can complete the proof of Azuma's inequality just as before.

The direct application of the Jakubowski-Kwapień representation (2.20) also provides a route to L^p bounds on $\sum_{i=1}^{n} d_i$. Letting $a_k = \|d_k\|_\infty$ in (2.20a) we have

$$\sum_{k=1}^{n} d_k = \int_0^1 \Bigl(\sum_{k=1}^{n} \|d_k\|_\infty r_k(s)\Bigr) f(\omega, s)\, ds, \tag{2.22}$$

so if we raise both sides to the pth power, apply Jensen's inequality on the right and then use (2.20b), we have

$$E\Bigl|\sum_{k=1}^{n} d_k\Bigr|^p \le \int_0^1 \Bigl|\sum \|d_k\|_\infty r_k(s)\Bigr|^p\, ds. \tag{2.23}$$

Since the $\{r_k\}$ are independent Bernoulli random variables, we can apply Khintchine's inequality (Chow and Teicher 1978 or Haagerup 1982) to obtain

$$E\left|\sum_{k=1}^{n} d_k\right|^p \leq \left(\frac{p+1}{2}\right)^{p/2} \left(\sum_{k=1}^{n} \|d_k\|_\infty^2\right)^{p/2}. \tag{2.24}$$

Comparison of (2.24) with (2.17) shows that (2.24) is not an essential improvement. Still, the approach via the representations seem to be a bit better, at least it simplified tracking the constant. An intriguing feature of both approaches is the appearance of the sum of squares of the L_∞ norms. Possibly this quantity is really rooted in the large deviation problem, but more likely, it is a coincidental artifact of the approaches. In the next section we systematically pursue the relationship of moments and large deviations in the context of the TSP. By introducing a few additional martingale tools, we can extract almost all of the information available on the tails of behavior of L_n.

3. Large Deviations and Moment Inequalities

We begin with a lemma that must be classical. It reminds us that the hunt for large deviation inequalities of Gaussian type can be conducted by pursuing appropriate L^p bounds. The interest in this observation comes from the fact that for some variables closely connected with L_n those bounds are easily proved.

LEMMA 3.1. *For any random variable Z, a necessary and sufficient condition that*

$$P(|Z| \geq t) \leq Ae^{-Bt^2}, \quad t \geq 0, \tag{3.1}$$

for some constants $A > 0$ and $B > 0$ is that for all $p \geq 1$

$$\|Z\|_p \leq Cp^{1/2} \tag{3.2}$$

for some constant C.

PROOF: If (3.1) holds, we multiply by pt^{p-1} and integrate as in (2.6) to obtain (3.2). For the converse, we just note by (3.2) and Markov's inequality that

$$P(|Z| \geq t) \leq \frac{1}{t^p} C^p p^{p/2} = e^{p \log C + (1/2)p \log p - p \log t},$$

so, choosing p such that $\log p = 2(\log t - \log C) - 1$, or $p = t^2 C^{-2} e^{-1}$, yields (3.1) with $A = 1$ and $B = (2C^2 e)^{-1}$. ∎

A central theme in the theory of martingales is that for any martingale difference sequence $\{Y_i, 1 \le i \le n\}$ the square function,

$$S_n = \left(\sum_{i=1}^{n} Y_i^2\right)^{1/2}, \tag{3.3}$$

and the maximal function,

$$M_n^* = \sup_{1\le k\le n}\left|\sum_{i=1}^{k} Y_i\right|,$$

share many properties with the underlying martingale

$$M_k = \sum_{i=1}^{k} Y_i, \quad 1 \le k \le n.$$

In particular, the inequalities of Doob and Burkholder tell us, among other things, that if any one of S_n, M_n^*, or M_n is in L^p for some $1 < p < \infty$ then all three are in L^p. The comparability of the moments S_n and M_n is particularly interesting for the theory of the TSP in $\mathbb{R}^d$ because, as we see in the next lemma, the L^p-norm of S_n can be bounded with enough precision to yield powerful large deviation inequalities. In fact, for $d = 2$ the resulting L^p bound is good enough to guarantee a large deviation inequality of Gaussian type.

LEMMA 3.2. *For the TSP martingale summands d_i of (2.1), we have for even integers $p \ge 2$ and any set $S \subset \{1, 2, \ldots, n\}$ that*

$$\left\|\left(\sum_{i\in S} d_i^2\right)^{1/2}\right\|_p \le C_1 p^{1/d}|S|^{1/2} n^{-1/d}, \tag{3.4}$$

where C_1 is the same constant as given in (2.7) and $|S|$ is the cardinality of S.

PROOF: We first expand and apply the generalized Hölder inequality:

$$\begin{aligned} E\left(\sum_{i\in S} d_i^2\right)^p &= \sum_{i_1\in S}\sum_{i_2\in S}\cdots\sum_{i_p\in S} E d_{i_1}^2 d_{i_2}^2 \ldots d_{i_p}^2 \\ &\le \sum_{i_1\in S}\sum_{i_2\in S}\cdots\sum_{i_p\in S} (E d_{i_1}^{2p})^{1/p}(E d_{i_2}^{2p})^{1/p}\ldots(E d_{i_p}^{2p})^{1/p}. \end{aligned}$$

Next, using the bound from (2.7), together with $\|d_i\|_{2p} \le C_1(2p/n)^{1/d}$ or $E d_i^{2p} \le C_1^{2p}(2p/n)^{2p/d}$, we find

$$E\left(\sum_{i\in S} d_i^2\right)^p \le |S|^p C_1^{2p}(2p/n)^{2p/d},$$

and hence for even integers p we conclude

$$\left\| \left(\sum_{i=1}^{n} d_i^2 \right)^{1/2} \right\|_p \leq C_1 p^{1/d} |S|^{1/2} n^{-1/d}. \qquad \blacksquare$$

This bound is of particular interest for $d = 2$ and $S = \{1, 2, \ldots, n\}$, since it is then of the form required in Lemma 3.1, i.e.

$$\left\| \left(\sum_{i=1}^{n} d_i^2 \right)^{1/2} \right\|_p \leq C p^{1/2}. \tag{3.5}$$

Thus for $d = 2$ the square function associated with the TSP martingale differences of (2.1) satisfies a large deviation inequality of Gaussian type (3.1).

One hope raised by (3.4) and (3.5) is that of extracting a Gaussian type large deviation inequality for L_n from that available for the square function $\left(\sum d_i^2\right)^{1/2}$ associated with L_n. To assess this possibility we first recall the square function inequalities of Burkholder (1966, 1973):

For $1 < p < \infty$ and any sequence of martingale differences Y_i with associated square function S_n defined by (3.3), we have

$$(18 p^{1/2} q)^{-1} \|S_n\|_p \leq \left\| \sum_{i=1}^{n} Y_i \right\|_p \leq 18 q^{1/2} p \|S_n\|_p \tag{3.6}$$

where $1/p + 1/q = 1$.

To see how (3.6) relates to the inequalities considered earlier, we note that we always have

$$|S_n|^p \leq \left(\sum_{i=1}^{n} \|d_i\|_\infty^2 \right)^{p/2}, \tag{3.7}$$

so, in particular, the second inequality of (3.3) gives us a bound like (2.16) which expresses the L^p version of Azuma's inequality. In this instance there is a critical difference in that the factor $p^{1/2}$ is inflated to p. Since large deviation results depend on the L^p inequalities for large p, this change in the constant is a major concern.

Still, when $d = 2$ we can use Lemma 3.2 to get good bounds on the tail probabilities of $L_n - EL_n$. We will give two illustrations of this approach. The first consists of showing that the moment generating function of $L_n - EL_n$ can be bounded independently of n.

To begin we note that for $|td_i| < 1$, the Taylor expansion of $\log(1 + td_i)$ gives us

$$\prod_{i=1}^{n} (1 + td_i) = \exp\left(\sum_{k=1}^{\infty} (-1)^{k+1} \beta_k t^k / k \right) \tag{3.8}$$

where $\beta_k = d_1^k + d_2^k + \cdots + d_n^k$. We next note for $k \geq 3$ that

$$|\beta_k| \leq \sum_{j=1}^{n} \|d_j\|_\infty^k \leq \sum_{j=1}^{n} C_1^k (n-j+1)^{k/2}$$
$$\leq C_1^k \sum_{j=1}^{\infty} j^{-3/2} = C_1^k \zeta(3/2) \tag{3.9}$$

where $\zeta(s) = \sum_{j=1}^{\infty} j^{-s}$, so from (3.8) and (3.9) we find

$$\exp\bigl(\beta_1 t - \tfrac{1}{2}\beta_2 t^2\bigr) \leq \prod_{i=1}^{n}(1+td_i)\exp\biggl(\zeta(3/2)\sum_{k=3}^{\infty} C_1^k t^k/k\biggr).$$

After taking expectations, we see

$$E\exp\bigl(\beta_1 t - \tfrac{1}{2}\beta_2 t^2\bigr) \leq \exp\biggl(\zeta(3/2)\sum_{k=3}^{\infty} C_1^k t^k/k\biggr) \equiv \phi(t), \tag{3.10}$$

so writing $\exp(\beta_1 t) = \exp(\beta_1 t - \beta_2 t^2)\exp(\beta_2 t^2)$ and applying Schwarz's inequality gives

$$E\exp(\beta_1 t) \leq \phi(2t)^{1/2}\bigl(E\exp(2\beta_2 t^2)\bigr)^{1/2}. \tag{3.11}$$

By (3.5) and Lemma 3.1 we know there is a constant $A > 0$ not depending on n such that $P(\beta_2^{1/2} > t) \leq Ae^{-t^2/A}$; hence, we have for $|t| < A^{-1}$ that

$$E\exp(t\beta_2) \leq \frac{1}{1-At} \tag{3.12}$$

and the bound (3.11) does not depend upon n.

The uniform bound on the moment generating function given by (3.11) and (3.12) naturally give a large deviation bound. For reference purposes we record the following consequence of (3.11) and (3.12) that was first obtained in Rhee and Talagrand (1988a) by different means.

Proposition 3.1. *For $d = 2$, there is a constant C such that for all $n \geq 2$ and $t > 0$*

$$P(|L_n - EL_n| \geq t) \leq Ce^{-Ct}. \tag{3.13}$$

A stronger result than (3.13) can be obtained by the use of Burkholders inequality. In fact, the following theorem seems to be about as much as one can obtain without going beyond the information on the TSP that is incorporated in (2.7) and (2.9).

THEOREM 3.1. *For $d = 2$, there is a constant C such that for all $n \geq 1$ and $p \geq 1$,*

$$\|L_n - EL_n\|_p \leq Cp^{1/2}(\log p)^{1/2}. \tag{3.14}$$

PROOF: We rely on the martingale representation (2.3) and split the representing sum into two terms,

$$\|L_n - EL_n\|_p = \left\|\sum_{i=1}^{n} d_i\right\|_p \leq \left\|\sum_{i \leq \alpha n} d_i\right\|_p + \left\|\sum_{i > \alpha n} d_i\right\|_p \tag{3.15}$$

for any $0 < \alpha < 1$. To the first summand we apply (2.17), the L^p version of Azuma's inequality, and to the second we apply Burkholder's second inequality to find

$$\|L_n - EL_n\|_p \leq C_4 p^{1/2}\Bigl(\sum_{i \leq \alpha n} \|d_i\|_\infty^2\Bigr)^{1/2} + 18pq^{1/2}\left\|\Bigl(\sum_{i > \alpha n} d_i^2\Bigr)^{1/2}\right\|_p. \tag{3.16}$$

Now we apply (2.9) to the first sum and (3.4) to the second,

$$\begin{aligned}\|L_n - &EL_n\|_p \\ &\leq C_1C_4p^{1/2}\Bigl(\sum_{i \leq \alpha n}(n-i+1)^{-1}\Bigr)^{1/2} + 18pq^{1/2}C_1p^{1/2}(1-\alpha)^{1/2} \\ &\leq C_6p^{1/2}\bigl(\log 1/(1-\alpha)\bigr)^{1/2} + C_6p^{3/2}q^{1/2}(1-\alpha)^{1/2}. \end{aligned} \tag{3.17}$$

When we let $(1-\alpha)^{1/2} = p^{-1}$, we find (3.14). ∎

COROLLARY. *There is a constant B such that for $d = 2$ we have*

$$P(|L_n - EL_n| \geq t) \leq 2e^{-Bt^2/\log(1+t)} \tag{3.18}$$

for all $t \geq 0$.

The proof of (3.18) from (3.14) follows just as in Lemma 3.1. This time the proper choice of p is $t^2/(C\log t)$ where C is the constant of (3.8).

Inequality (3.18) was also first established in Rhee and Talagrand (1988a). Their proof grew out of the idea of interpolating between the $d = 2$ case of (2.14) where the tails have quadratic exponential behavior that depends on n, and on (3.13), where the bound is independent of n but is linear exponential. Rhee and Talagrand (1988a) bring these two bounds together to prove (3.18) by use of interpolation results from Bergh and Lofstrom (1976). The present proof via (3.8) is simpler than that of

Rhee and Talagrand, at least so far it relies on methods that are familiar to probabilists. Still, even now, the Burkholder inequalities might not be regarded as completely commonplace tools, and the proof of (3.18) is not yet elementary.

The quest that has been traced here, the derivation of a Gaussian type large deviation bound for $L_n - EL_n$, has very recently come to fruition through Rhee and Talagrand (1988b). By combining their basic martingale approach with a bare-handed investigation of the geometry of an n-sample from $[0,1]^2$, they show that one can indeed remove the logarithmic factor from (3.18). The resulting inequality for the TSP in $d = 2$ stands as both the natural end to a line of investigation and as a hard challenge. What can one say for $d \geq 2$? What other functionals permit a comparable analysis?

4. Analytical Bounds from Spacefilling Curves

For many problems concerning combinatorial optimization in $\mathbb{R}^d$ one can obtain useful bounds by appealing to the existence of a map ϕ from $[0,1]$ onto $[0,1]^d$ that is Lip α with $\alpha = 1/d$, i.e. $|\phi(s) - \phi(t)| \leq c|s-t|^{1/d}$ for a constant c and all $0 \leq s \leq t \leq 1$. Moreover, Milne (1980) established that one can further require ϕ to be measure preserving, and from our perspective, the benefit of that fact is that it lets us use spacefilling curve techniques to get probabilistic inequalities, at least in the case of uniformly distributed random variables.

For our first example we again consider the traveling salesman problem in $\mathbb{R}^d$, but this time we take the cost of travel from x to y to be $|x-y|^p$, the pth power of the Euclidean distance. If $S = \{x_1, x_2, \ldots, x_n\}$ is a set of n points in $[0,1]^d$, how can we bound $\tilde{L}(S)$, the length of the shortest tour through the points of S under this metric, i.e. how can we bound

$$\tilde{L}(S) = \min_{\sigma} \sum_{i=1}^{n-1} |x_{\sigma(i)} - x_{\sigma(i+1)}|^p \tag{4.1}$$

where the minimum is over all cyclic permutations?

Since ϕ is a surjection, each $x_i \in S \subset [0,1]^d$ has a pre-image $y_i \in [0,1]$. If we choose a cyclic permutation σ so that $y_{\sigma(1)} \leq y_{\sigma(2)} \leq \cdots \leq y_{\sigma(n)}$, then a heuristic tour of the $\{x_i\}$ can be formed by visiting them in the order of the $\{y_i\}$. For this heuristic we find

$$\begin{aligned}\tilde{L}(S) &\leq \sum_{i=1}^{n-1} \big|\phi(y_{\sigma(i)}) - \phi(y_{\sigma(i+1)})\big|^p \\ &\leq c^p \sum_{i=1}^{n-1} \big|y_{\sigma(i)} - y_{\sigma(i+1)}\big|^{p/d} \\ &\leq c^p n^{(d-p)/d} \end{aligned} \tag{4.2}$$

where we applied Hölder's inequality and the fact that $\sum |y_{\sigma(i)} - y_{\sigma(i+1)}|$ is bounded by 1. The key idea of (4.2), i.e. building a path through $\{x_n, 1 \leq i \leq n\}$ by visiting the points in the linear ordering of the $\{y_i, 1 \leq i \leq n\}$, is called the *spacefilling heuristic*. For application to the TSP, this idea was first proposed by Bartholdi and Platzman (1982) and independently by D.H. Fremlin (see e.g. Fremlin 1982). Both for heuristic algorithms and analytic bounds, the idea of using a spacefilling map to exploit the linear ordering, or simple geometry, of $[0,1]$ has many natural applications, and the breadth of these variations can be seen by consulting the survey by Bartholdi and Platzman (1988), the papers by Glass (1985) and Imai (1986), or the recent thesis by Bertsimas (1988).

For $p = 1$, inequality (4.2) recaptures the familiar $O(n^{(d-1)/d})$ bound, but for $p = d$ it provides new information by providing a $O(1)$ bound. In contrast, one only obtains the weaker inequality

$$\tilde{L}(S) \leq c \log n. \tag{4.3}$$

by classical arguments that rest on the fact that any set of n points in $[0,1]^d$ contains a pair within $cn^{-1/d}$ of each other.

The argument used for sharper bound (4.2) was also applied in Steele (1988) to show that the sum of the dth powers of the lengths of the edges of a minimal spanning tree of n points in $[0,1]^d$ can be bounded independent of n. For $d = 2$ the uniform boundedness of the sum of squares of the edge lengths had been established earlier by Gilbert and Pollak (1968), but their delicate geometric argument has no natural analogue for $d > 2$. In contrast, the bound provided by the spacefilling heuristic works pleasantly in all $d \geq 2$. For the spacefilling heuristic applied to the TSP the most interesting problems concern the ratio of the length of the tour produced by the spacefilling curve to the length of the optimal tour. In $\mathbb{R}^2$ Platzman and Bartholdi (1988) provided a bound of order $O(\log n)$, and they conjectured that there is a uniform bound on the ratio. Bertsimas and Grigni (1989) settled the conjecture by giving an example that shows the ratio can be as bad as $c \log n$. The following special case of work in Steele (1989) complements the results of Platzman and Bartholdi (1988) in a way that may be useful in algorithmic applications. The proof does not require any detailed properties of the spacefilling curve in order to provide ratio bounds, except that the curve is measure preserving and is as smooth as feasible.

THEOREM 4.1. *Let ϕ be a measure preserving transformation of $[0,1]$ onto $[0,1]^2$ that is Lipschitz of order $\alpha = 1/2$, i.e.*

$$|\phi(x) - \phi(y)| \leq c|x - y|^{1/2} \tag{4.4}$$

for some c and all $x, y \in [0,1]$. If H_n is the length of the path through the points $\{x_1, x_2, \ldots, x_n\} \subset [0,1]^2$ that is constructed using the spacefilling *heuristic based on ϕ, then for $n \geq 2$*

$$H_n \leq L_n\big\{1 + 2c^2 \log(m/\bar{e})\big\} + \pi c^2 m, \tag{4.5a}$$

where L_n is the length of the optimal path through $\{x_1, x_2, \ldots, x_n\}$, m is the length of the longest edge in the heuristic path, and $\bar{e}$ is the average length of the edges in the optimal path.

Corollary.

$$H_n \leq \big(1 + \pi c^2 + 2c^2 \log n\big) L_n. \tag{4.5b}$$

Proof: We suppose the heuristic tour visits the points in the order $x_1, x_2, \ldots, x_n$, i.e. we suppose there are $t_i \in [0,1]$ such that $t_1 \leq t_2 \leq \cdots \leq t_n$ with $x_i = \phi(t_i)$. For $\lambda > 0$ we introduce two basic subsets of $\{1, 2, \ldots, n-1\}$ by

$$U(\lambda) = \big\{i : |t_{i+1} - t_i| > \lambda, 1 \leq i < n\big\}$$

and

$$V(\lambda) = \big\{i : |\phi(t_{i+1}) - \phi(t_i)| > \lambda, 1 \leq i < n\big\}.$$

For $i \in V(\lambda)$ inequality (4.4) implies

$$c|t_i - t_{i+1}|^{1/2} \geq |\phi(t_{i+1}) - \phi(t_i)| \geq \lambda$$

so $i \in V(\lambda)$ implies $i \in U(\lambda^2/c^2)$, i.e.

$$V(\lambda) \subset U(c^{-2}\lambda^2). \tag{4.6}$$

If $g(\lambda)$ is the cardinality of $V(\lambda)$, we also have

$$H_n = \int_0^m g(\lambda)\, d\lambda, \tag{4.7}$$

where $m = \max_{1 \leq i < n} |\phi(t_{i+1}) - \phi(t_i)|$, so our goal is now to use (4.6) to bound $g(\lambda)$.

For $i \in U(\lambda)$ the intervals $[t_i, t_i + \lambda]$ are non-intersecting, so if we set

$$A_i = \phi([t_i, t_i + \lambda])$$

then since ϕ preserves measure, each A_i has Lebesgue measure $\lambda = \mu(A_i)$ and $A_i \cap A_j$ has measure zero for any pair $i \neq j$, $i, j \in U(\lambda)$.

We let $D(x, C) \subset [0,1]^2$ denote the set of all points within distance x of the curve C, and let T_n be an optimal tour of $\{x_1, x_2, \ldots, x_n\}$ with length L_n. By (4.4) and the fact that each x_i is somewhere on the path, we have for each $i \in S$ that

$$A_i \subset D(c\lambda^{1/2}, T_n). \tag{4.8}$$

Naiman's inequality on the volume of tubes (e.g. Naiman 1986, or the easier version of the basic result given in Johnstone and Siegmund 1989) tells us that for any rectifiable curve C of length L one has

$$\mu\big(D(x,C)\big) \le 2xL + \pi x^2, \tag{4.9}$$

for all $x \ge 0$. If $f(\lambda)$ denotes the cardinality $U(\lambda)$ we then have by (4.8) and (4.9) that

$$\lambda f(\lambda) = \mu\Big(\bigcup_{i\in U(\lambda)} A_i\Big) \le \mu\big(D(c\lambda^{1/2}, T_n)\big) \le 2c\lambda^{1/2}L_n + \pi c^2\lambda,$$

so

$$f(\lambda) \le 2c\lambda^{-1/2}L_n + \pi c^2. \tag{4.10}$$

By (4.6) and (4.10) we find our basic bound

$$g(\lambda) \le 2c^2 L_n \lambda^{-1} + \pi c^2. \tag{4.11}$$

For any $0 < \alpha < m$, we can apply the trivial bound $g(\lambda) \le n-1$ for $\lambda \in [0,\alpha]$ and apply (4.11) for $\lambda \in [\alpha, m]$; so, when we integrate in (4.7), we find

$$H_n \le \alpha(n-1) + 2c^2 L_n \log(m/\alpha) + \pi c^2(m-\alpha). \tag{4.12}$$

Finally, since $L_n \le H_n \le (n-1)m$ we have for $\alpha = L_n/(n-1) = \bar{e}$ that $\alpha \in [0,m]$, so we can let $\alpha = \bar{e}$ in (4.12) to find (4.5a). To see that (4.5b) follows from (4.5a) we just invoke the very crude bound $m \le L_n$ and $\bar{e} = L_n/(n-1)$. ∎

The argument used in the proof of Theorem 4.1 uses several ideas from Bartholdi and Platzman (1988), and it makes progress mainly by being systematic in the exploitation of the bound (4.9).

The next section deals more directly with the geometry and topology of spacefilling curves.

5. Schoenberg's Map and Smoother Maps

Section 4 made use of smooth spacefilling curves, but it did not provide concrete examples. This section engages the problem of constructing spacefilling curves, especially curves that are as smooth as possible and that preserve Lebesgue measure. It also points out a topological barrier to the sharpening of Theorem 4.1.

We begin by considering a method of Schoenberg (1938) that gives perhaps the shortest classical example of a continuous map from $[0,1]$ onto

$[0,1]^2$. Schoenberg's map is not as smooth as we need, but it points the way to a map that is both simpler and smoother. We first define a real valued even function f of period 2 by taking $f(t) = 0$ in $(0, 1/3)$, $f(t) = 1$ in $(1/3, 1)$, and making $f(t)$ linear in $(1/3, 2/3)$. We then define Schoenberg's spacefilling curve by the explicit formulas

$$x(t) = \frac{1}{2}f(t) + \frac{1}{2^2}f(3^2t) + \frac{1}{2^3}f(3^4t) + \cdots \tag{5.1a}$$

and

$$y(t) = \frac{1}{2}f(3t) + \frac{1}{2^2}f(3^3t) + \frac{1}{2^3}f(3^5t) + \cdots \ . \tag{5.1b}$$

To prove the map $t \to (x(t), y(t))$ is surjective, we first note that if $\{a_k\}$ is any infinite sequence of 0's and 1's, then a typical point in the Cantor set $\mathbf{C} \subset [0,1]$ can be written uniquely as

$$t_0 = \frac{2a_0}{3} + \frac{2a_1}{3^2} + \frac{2a_2}{3^3} + \dots \ . \tag{5.2}$$

By straightforward, but tedious, bounds one can also show that f can be used to extract the kth term in the ternary expansion of t_0, specifically

$$f(3^k t_0) = a_k. \tag{5.3}$$

Now, given any $(x_0, y_0) \in [0,1]^2$, we can use the binary expansion of x_0 and y_0 together with the explicit formulas (5.1) and (5.2) to write down a point in $\mathbf{C}$ that ϕ maps to (x_0, y_0), so ϕ is a surjection of $[0,1]$ onto $[0,1]^2$.

One important aspect of the explicit formulas (5.1) and (5.2) is their computational feasibility. Not only do we know that for every point (x_0, y_0) of $[0,1]^2$ that there *exists* a point of $[0,1]$ that maps onto (x_0, y_0), but we can also quickly compute a point $t \in \mathbf{C}$ such that $\phi(t) = (x_0, y_0)$.

Now we need to assess the smoothness of Schoenberg's map $\phi(t) = (x(t), y(t))$. By uniform convergence, we see ϕ is continuous on $[0,1]$. In fact, it is easy to show there is an α so that ϕ is in Lip α, and we can even determine the best value of α. First, just consider $x(t)$ and note that f satisfies the two naive bounds $|f(s) - f(t)| \leq 3|s - t|$ and $|f(t)| \leq 1$. Thus we have for any $n \geq 1$ that

$$|x(s) - x(t)| \leq 3\sum_{k=1}^{n} 2^{-k}|3^{2k-2}s - 3^{2k-2}t| + 2\sum_{k=n+1}^{\infty} 2^{-k}, \tag{5.4}$$

so for all integers n we have

$$|x(s) - x(t)| = O\big(|s - t|(9/2)^n + 2^{-n}\big). \tag{5.5}$$

Finally, by choosing n to be the integer nearest $-(\log_2 |s-t|)/\log_2 9$, we find $|x(s)-x(t)| = O(|s-t|^\alpha)$ where $\alpha = (2\log_2 3)^{-1}$.

Having achieved an α for which $f \in \text{Lip } \alpha$, we will show that $f \notin \text{Lip } \alpha'$ for any $\alpha' > \alpha$ by using some elementary facts about Hausdorff dimension. In fact, we use the result of Hausdorff (1919) that the dimension of the Cantor ternary set equals $\log 2/\log 3$. If we let $N(\epsilon)$ be the least number of intervals $\{I_i\}$ of length 2ϵ, $0 < \epsilon < 1$, that cover the Cantor set $\mathbf{C}$, then in terms of $N(\epsilon)$, the fact that $\mathbf{C}$ has Hausdorff dimension $1/\log_2 3$ tells us that for any $\delta > 0$ there are constants A and B such that $A\epsilon^{-\beta-\delta} > N(\epsilon) > B\epsilon^{-\beta+\delta}$, where $\beta = 1/\log_2 3$.

Now suppose ψ is any map of $\mathbf{C}$ onto $[0,1]^2$, and suppose that ψ is also Lip α'. If λ denotes Lebesgue measure in $\mathbb{R}^2$, then since the compact set $\psi(\mathbf{C})$ covers $[0,1]^2$, and since we have a collection of $N(\epsilon)$ intervals $\{I_i\}$ of length 2ϵ that cover $\mathbf{C}$, we have

$$1 \le \lambda(\psi(\mathbf{C})) \le \sum_{i=1}^{N(\epsilon)} \lambda(\psi(I_i)) \le N(\epsilon)\pi(c\epsilon^{\alpha'})^2 = O(N(\epsilon)\epsilon^{2\alpha'}). \tag{5.6}$$

From (5.6) and the arbitrariness of $\delta > 0$, we conclude that $\beta \ge 2\alpha'$, i.e. $\alpha \le 1/(2\log_2 3)$ for any Lip α map of the Cantor set onto $[0,1]^2$. We have thus established that Schoenberg's spacefilling curve is precisely of smoothness type Lip α with $\alpha = 1/(2\log_2 3)$.

Although Schoenberg's mapping is a rich source of insight, one has to put in considerable modification in order to attain the maximal level of smoothness that one can have. Still the Lip 1/2 measure preserving property is shared by several of the classical spacefilling curves, particularly those due to Hilbert and Lebesgue. For a proof of these features of the classical curves as well as some remarkable analytical applications of spacefilling curves, one can consult Milne (1980). Also, to show one cannot find a map smoother than Lip 1/2 from $[0,1]$ onto $[0,1]^2$, we just use the fact that the Hausdorff dimension of $[0,1]$ is 1 and repeat the argument given for the lower bound of smoothness for Schoenberg's map.

There is nothing more we need to say about the construction of smooth spacefilling curves, but there are some final issues concerning the spacefilling heuristic and the topology of $[0,1]^2$. The bound on the ratio H_n/L_n that was given in Section 4 really relied on bounding the ratio H_n^*/L_n where

$$H_n^* = \sum_{i=1}^{n-1} |t_{i+1} - t_i|^{1/2} \tag{5.7}$$

and $\phi(t_i) = x_i, 1 \le i \le n$. To see a subtlety in this process, we first recall that the dimension theorem of general topology tells us that there is no continuous bijection between $[0,1]$ and $[0,1]^2$ (see e.g. Dugundji 1970, p.

359). Thus, every continuous surjection must have a double point. The investigation of multiple points was pursued further by Pólya (1913) who gave a spacefilling curve with multiple points with multiplicity bounded by three. This explicit line of investigation was completed by Hurewicz (1933) who showed that any surjection of $[0,1]$ onto $[0,1]^2$ must have a triple point. These facts can be used to show that bounding of H_n^* can be slippery.

For example, suppose (x, y) is the triple point guaranteed by Hurewicz and therefore suppose we have $t_1 < t_2 < t_3$ with $\phi(t_i) = (x, y)$. Now, if $s_{1,j} < s_{2,j} < s_{3,j}$ and $s_{i,j} \to t_i$ as $j \to \infty$ for each $i \in \{1, 2, 3\}$ we have that $L_3 = L\big(\phi(s_{1,j}), \phi(s_{2,j}), \phi(s_{3,j})\big) \to 0$ as $j \to \infty$. On the other hand,

$$H_3^* = \sum_{1 \le i \le 3} |s_{i,j} - s_{i+1,j}|^{1/2} \ge \tfrac{1}{2}|t_3 - t_1| \tag{5.8}$$

for all sufficiently large j. We thus have that H_3^*/L_3 can be made arbitrarily large, and, at first blush, this fact might seem to cast doubt on (4.5a) or (4.5b). There is no contradiction between (5.8) and the earlier bounds, but (5.8) nicely shows that one cannot rely too heavily on H_n^* for a detailed understanding of H_n.

6. Karp's Partitioning Algorithm

The Euclidean traveling salesman problem is the task of *computing* the shortest path through a set of points in $\mathbb{R}^d$. As a computational challenge, the TSP has become an essential test problem for combinatorial optimization, and, as one can see by considering the range of techniques in *The Traveling Salesman Problem: A Guided Tour of Combinatorial Optimization* (Lawler, et al. 1985), the TSP has provided the inspiration for some of the most fundamental developments in the field.

One such development took place when Karp (1976, 1977) used the Beardwood, Halton, Hammersley theorem to show how a simple partitioning algorithm yields a solution to the TSP that is (1) computable in polynomial time and (2) asymptotically optimal in an appropriate probabilistic sense. In this section, we will review Karp's basic idea and make a point that deserves to be more widely known. The asymptotic optimality of Karp's algorithm can be obtained independently of the Beardwood, Halton, Hammersley theorem. In fact, we will see that one can justify Karp's algorithm with results that are considerably less refined than the BHH theorem.

The simplest version of Karp's algorithm addresses the case of the uniform distribution, and, for ease of exposition, we will keep to that case. Let $X_i, 1 \le i < \infty$, be independent random variables with the uniform distribution in $[0,1]^d$, and suppose k_n is a sequence of integers that grows more slowly than $n^{1/d}$. Karp's method for obtaining a path through the points of $S = \{X_1, X_2, \ldots, X_n\}$ is as follows:

(1) Partition $[0,1]^d$ into k_n^d congruent subcubes $\{Q_j\}$.

(2) For each $j, 1 \leq j \leq k_n^d$, calculate an optimal path P_j through the points $S \cap Q_j$.

(3) Join the endpoints of the P_j to form a heuristic path H through all the points of S.

This description is incomplete until we specify k_n, provide a method for finding the paths P_j, and spell out how the P_j are joined to form H. As it happens, virtually any reasonable choices will suffice. For example, we can calculate the P_j by complete enumeration of the possible orders of visiting the points of $Q_j \cap S$, or we can use dynamic programming. Either of these methods will be fast enough to yield a polynomial time algorithm if k_n is chosen appropriately. Thus, for the moment, our concern is just with the effectiveness of the algorithm.

Still, we need to pick a specific rule concerning the connection of the partial paths. Thus, for each $1 \leq i \leq k_n^d$, we label the two end points of the partial path of $S \cap Q_i$ by a_i and b_i, and we connect b_i to a_{i+1} where the Q_i have been ordered lexicographically according to the vertex within each square that is lexicographically minimal. With these procedures assumed, one can show the following:

THEOREM 6.1. *If L_n^K denotes the length of the path produced by Karp's method and if k_n is any unbounded increasing sequence such that $n/k_n^d \to \infty$, then for any $\epsilon > 0$ we have*

$$\sum_{n=1}^{\infty} P\left\{\frac{L_n^K}{L_n} \geq 1 + \epsilon\right\} < \infty. \tag{6.1}$$

The proof of (6.1) follows from the next two lemmas. The first guarantees that under the hypotheses on $\{X_i\}$ that L_n cannot be too short.

LEMMA 6.1. *There exist constants $A > 0$ and $0 < \rho < 1$ such that for all $n \geq 1$, we have*

$$P\left\{L_n < An^{(d-1)/d}\right\} \leq \rho^n. \tag{6.2}$$

The proof of (6.2) is easily achieved by dividing $[0,1]^d$ into n subcubes of volume $1/n$, applying standard occupancy results, and a little geometry. The second lemma is more challenging.

LEMMA 6.2. *There is an r_n depending only on n and k_n such that for all n,*

$$L_n \leq L_n^K \leq L_n + r_n \tag{6.3}$$

and

$$r_n \leq c \left\{ n^{(d-2)/(d-1)} k_n^{1/(d-1)} + k_n^{d-1} \right\} \tag{6.4}$$

where the constant c depends on d and the sequence $\{k_n\}$.

Remark. When $k_n = o(n^{1/d})$ then (6.4) says that $r_n = o(n^{(d-1)/d})$. Since this result holds everywhere, (6.4) and (6.2) yield (6.1).

It is relatively easy to sketch a proof of (6.4). Let $\{F_i\}$ be the set of faces of the k_n^d subcubes $\{Q_j\}$. We will use the optimal path P and some additional edges in order to bound L_n^K. If e is an edge of the optimal path, P, we associate with e a set of points that we will call *pierce points.* If e is interior to some Q_j, then the set of pierce points created by e is just the null set. On the other hand, if $e = (a, b)$ where $a \in Q$ and $b \in Q'$ and Q and Q' are distinct subcubes, then e will create a set of two pierce points. In particular, if F and F' are the faces of Q and Q' that intersect the line from a to b, then $p = e \cap F$ and $p' = e \cap F'$ are called the pierce points associated with a and b, respectively.

We will now build a set of points that may have rather large cardinality, but that can be proved to lie on a relatively short path. First, note the set of pierce points has cardinality bounded by n since each X_i is associated with at most one pierce point. Next, to each face F of each cube Q, we associate a set S_F consisting of (1) its 2^{d-1} subfaces of dimension zero (i.e. its vertices) and (2) its set of pierce points. For each F the set S_F is contained in a $d-1$ dimensional cube of edge length k_n^{-1}, so, by classical bounds (e.g. Few 1955), there is a tour through the points of S_F of length bounded by $ck_n^{-1}|S_F|^{(d-2)/(d-1)}$, where $|S_F|$ denotes the cardinality of S_F. Now consider the union of all of the tours through S_F for all F together with the optimal path P. This set of edges has the property that for each j it contains a path that is contained in Q_j and goes through all the points of S that are contained in Q_j.

We finally see that L_n^K can be bounded by three terms: (1) the length of the edges in the optimal tour, (2) the sum of the edges needed to tour S_F for all F, and (3) the cost of the edges required in Step 3 of Karp's heuristic. We thus see that

$$L_n^K \leq L_n + c \sum_F k_n^{-1} |S_F|^{(d-2)/(d-1)} + ck_n^{d-1}. \tag{6.5}$$

The bound on r_n given in (6.4) now follows from (6.5) by Hölder's inequality, the fact that the sum of the $|S_F|$ is $O(n)$, and the fact that there are $O(k_n^d)$ faces of the cubes $\{Q_i\}$.

To some extent, the preceding sketch follows the lines of Halton and Terada (1982) which one can consult for additional details. From the present perspective, the main point of interest is that one requires so little probability theory. All one needs is the elementary occupancy theory in Lemma 6.1.

7. Concluding Remarks

In 1959 the work of Beardwood, Halton and Hammersley was a singular event in the sense that prior to that date and for many years subsequent one finds no comparable work relating probability theory and combinatorial optimization. The power and beauty of the Beardwood, Halton, Hammersley theorem were immediately present, but considerable time needed to elapse before wide appreciation was possible. The key step in the process toward that appreciation is the work of Karp (1976). By connecting the asymptotic result of Beardwood, Halton, and Hammersley with the possibility of effective algorithms, Karp created an eager audience for both the original work and for results that complement it. In Karp and Steele (1985) and the recent thesis of Bertsimas (1988), one can find a review of that development. This article also provides a review, but here the focus is narrowed to the developing roles of martingale theory and of spacefilling curves.

The field of martingale inequalities is so rich that the applications in Sections 2 and 3 only offer a hint of future possibilities. Connections between martingale theory and problems like the TSP can be counted on to develop vigorously in the next few years.

Among the concrete problems that may, or may not, be attacked via martingales, the one that stands out most concerns the completion of our understanding of the tails of $L_n - EL_n$ in $d = 2$. More broadly we would like to understand the ways in which $L_n - EL_n$ behaves like a Gaussian random variable. In particular, we would like to know if $L_n - EL_n$ converges in distribution to a Gaussian limit.

The force behind applications of the spacefilling heuristic is not as great as that behind martingale theory, but one can still expect vigorous activity. The strong interest in the geometry of fractals provides one motivation, but the fact that the heuristic is easily coded also helps. Even though the conjecture of Bartholdi and Platzman is formally settled by the example of Bertsimas and Grigni (1989), many questions remain. As suggested in Section 5, one can expect some more negative results. Nevertheless, one may be able to provide further positive results like Theorem 4.1 that are of use in practical problems.

John Hammersley coined the inviting phrase 'seedlings of research', and throughout his work one finds a generous willingness to reveal interesting ideas that still have room to grow. The intention of this article has been to try to live up to that tradition while engaging the shortest path through many points.

Acknowledgement

I wish to thank S. Lalley for permitting me to include his proof of the critical case of the Beardwood, Halton, Hammersley theorem. I also thank I. Karatzas and L. Platzman for their comments on parts of this article, and I thank D. Bertsimas, D.H. Fremlin, I. Johnstone, W.-S. Rhee, and D. Siegmund for making available their unpublished work. This research was supported in part by National Science Foundation Grant # DMS-8812868.

References

Azuma, K. (1967). Weighted sums of certain dependent random variables. *Tôhoku Mathematical Journal* 19, 359–367.

Bartholdi, J.J. and Platzman, L.K. (1982). An $O(n \log n)$ planar travelling salesman heuristic based on spacefilling curves. *Operations Research Letters* I (4), 121–125.

——— (1988). Heuristics based on spacefilling curves for combinatorial problems in Euclidean space. *Management Science* 34, 291–305.

Beardwood, J., Halton, J.H., and Hammersley, J. (1959). The shortest path through many points. *Proceedings of the Cambridge Philosophical Society* 55, 299–327.

Bergh, J. and Lofstrom, J. (1976). *Interpolation Spaces: An Introduction.* Springer-Verlag, New York.

Bertsimas, D. (1988). *Probabilistic Combinatorial Optimization Problems.* Ph. D. Thesis, Operations Research Center, Massachusetts Institute of Technology, Cambridge, Massachusetts.

Bertsimas, D. and Grigni, M. (1989). On the spacefilling curve heuristic for the Euclidean traveling salesman problem. Technical Report, Operations Research Center, Massachusetts Institute of Technology.

Bially, T. (1969). Spacefilling curves: their generation and their application to bandwidth reduction. *IEEE Transactions on Information Theory* II-15, 658–664.

Burkholder, D.L. (1966). Martingale transforms. *Annals of Mathematical Statistics* 37, 1494–1504.

——— (1973). Distribution function inequalities for martingales: The 1971 Wald Memorial Lectures. *Annals of Probability* 1, 19–42.

Chow, Y.S. and Teicher, H. (1978). *Probability Theory: Independence, Interchangeability, Martingales.* Springer-Verlag, New York.

Dugundji, J. (1970). *Topology.* Allyn and Bacon, Boston.

Efron, B. and Stein, C. (1981). The Jackknife estimate of variance. *Annals of Statistics* 9, 586–596.

Few, L. (1955). The shortest path and the shortest road through n points in a region. *Mathematika* 2, 141–144.

Fremlin, D.H. (1982). On the traveling salesman problem. University of Essex, Note of 6 May 1982, unpublished.

Gilbert, E.N. and Pollak, H.O. (1968). Random minimal trees. *SIAM Journal of Applied Mathematics* 16, 376–387.

Glass, M. (1985). The application of spacefilling curves to some VLSI design problems. Unpublished manuscript, Fermi National Accelerator Laboratory, Batavia, Illinois.

Gradshteyn, I.S. and Ryzhik, I.M. (1963). *Tables of Integrals, Sums, Series and Products*. 4th edn., Fiznatgizdat, Moscow.

Haagerup, U. (1982). The best constants in the Khintchine inequality. *Studia Mathematica* 52, 231–283.

Halton, J.H. and Terada, R. (1982). A fast algorithm for the Euclidean traveling salesman problem, optimal with probability one. *SIAM Journal of Computing* 11, 28–46.

Hausdorff, F. (1919). Dimension und äußeres Maß. *Mathematische Annalen* 79, 157–179.

Hurewicz, W. (1933). Über dimensionser höhende steitige Abbildungen. *Journal für Reine und Angewandte Mathematik* 169, 71–78.

Imai, H. (1986). Worst-case analysis for planar matching and tour heuristics with bucketing techniques and spacefilling curves. *Journal of the Operations Research Society of Japan* 29, 43–67.

Jakubowski, J. and Kwapień, S. (1979). On multiplicative systems of functions. *Bulletin de L'Académie Polonaise des Sciences* 27, 689–694.

Johnstone, I. and Siegmund, D. (1989). On Hotelling's formula for the volume of tubes and Naiman's inequality. *Annals of Statistics* 16, to appear.

Kakutani, S. (1986). *Collected Works*. Ed. R.R. Kallman, Birkhäuser, Boston.

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

——— (1977). Probabilistic analysis of partitioning algorithms for the traveling salesman problem in the plane. *Mathematics of Operations Research* 2, 209–224.

Karp, R.M. and Steele, J.M. (1985). Probabilistic analysis of heuristics. In *The Traveling Salesman Problem: A Guided Tour of Combinatorial Optimization*, eds. E.L. Lawler et al., John Wiley and Sons, New York, 181–206.

Lalley, S. (1984). Personal communication.

Lawler, E.L., et al. (1985). *The Traveling Salesman Problem: A Guided Tour of Combinatorial Optimization*. Editors, John Wiley and Sons, New York.

Milne, S.C. (1980). Peano curves and smoothness of functions. *Advances in Mathematics* 35, 129–157.

Naiman, D.Q. (1986). Conservative confidence bounds in curvilinear regression. *Annals of Statistics* 14, 896–906.

Paley, R.E.A.C. (1932). A remarkable series of orthogonal functions (I). *Proceedings of the London Mathematical Society* 34, 241–264.

Platzman, L.K. and Bartholdi, J.J. (1988). Spacefilling curves and the planar traveling salesman problem. *Journal of the Association of Computing Machinery*, to appear.

Pólya, G. (1913). Über eine Peanosche Kurve. *Bulletin de l'Académie de Sciences, Cracovie A* 305–313 (also *George Pólya's Collected Works* 3, 1–8,

eds. J. Hersch and G.-C. Rota, MIT Press, Cambridge, Massachusetts).
Rhee, W.T. and Talagrand, M. (1987). Martingale inequalities and NP-complete problems. *Mathematics of Operations Research* 12, 177–181.
——— (1988a). Martingale inequalities, interpolation, and NP-complete problems. *Mathematics of Operations Research*, to appear.
——— (1988b). A sharp deviation for the stochastic traveling salesman problem. *Annals of Probability*, to appear.
Schoenberg, I.J. (1938). On the Peano curve of Lebesgue. *Bulletin of the American Mathematical Society* 44, 519.
Steele, J.M. (1981a). Subadditive Euclidean functionals and non-linear growth in geometric probability. *Annals of Probability* 9, 365–376.
——— (1981b). Complete convergence of short paths and Karp's algorithm for the TSP. *Mathematics of Operations Research* 6, 374–378.
——— (1988). Probabilistic and worst case analyses of some classical problems of combinatorial optimization in Euclidean space. Technical Report, Program in Statistics and Operations Research, Princeton University.
——— (1989). Efficacy of spacefilling heuristics in Euclidean combinatorial optimization. Technical Report, Program in Statistics and Operations Research, Princeton University.

Appendix I. Lalley's Proof of BHH

S. Lalley (1984) provided a remarkable proof of the most interesting case of the Beardwood, Halton, Hammersley theorem. Lalley's previously unpublished proof wins the prize for using minimal machinery. Moreover, his proof serves as a model of the power of similarity arguments and provides a line of attack that is applicable to many other functionals.

Let $U_1, U_2, \ldots$ be independent random variables each having the uniform distribution on $[0,1]^2$, and let L_n be the length of the shortest path through $U_1, U_2, \ldots, U_n$. Observe that L_n is nondecreasing in n. We are to prove that $n^{-1/2} L_n \to C$ a.s. for a constant $C \in (0, \infty)$. For this it suffices to prove that if $N(t),\ t \geq 0$, is a Poisson process with rate 1, then as $t \to \infty$ we have

$$t^{-1/2} L_{N(t)} \to C \quad \text{almost surely.} \tag{AI.1}$$

Partition the square $[0,1]^2$ into squares $Q_1, Q_2, \ldots, Q_{m^2}$ of side m^{-1}, and define $\lambda_t^m(Q_i)$ to be the length of the shortest path through $\{U_1, U_2, \ldots, U_{N(t)}\} \cap Q_i$. It is easy to see that for each $t > 0$ and each $m = 1, 2, \ldots$, the random variables $\lambda_t^m(Q_1), \lambda_t^m(Q_2), \ldots, \lambda_t^m(Q_{m^2})$ are independent and identically distributed. Moreover, $m\lambda_{m^2 t}^m(Q_i)$ has the same distribution as $L_{N(t)}$. Finally, we note $\mathrm{Var}(L_{N(t)}) < \infty$ for each $t \geq 0$ as one can see from the trivial bound $L_{N(t)} \leq 2^{1/2} N(t)$.

Lemma 1. *For each $t > 0$ and each $m = 1, 2, \ldots,$ we have*

$$-6m + \sum_{i=1}^{m^2} \lambda_t^m(Q_i) \le L_{N(t)} \le m\sqrt{5} + \sum_{i=1}^{m^2} \lambda_t^m(Q_i). \tag{AI.2}$$

Proof: To prove the right inequality we only need to obtain a path through $U_1, U_2, \ldots, U_{N(t)}$. For $1 \le i \le m^2$ we first find the shortest path through $\{U_1, U_2, \ldots, U_{N(t)}\} \cap Q_i$ then knit these m^2 paths together by joining endpoints in adjacent squares ordered in snake raster order. Since points in adjacent squares are not separated by a distance greater than $\sqrt{5}\, m^{-1}$, the resulting path has length no greater than $m\sqrt{5} + \sum_{i=1}^{m^2} \lambda_t^m(Q_i)$, establishing the right hand inequality. One should note that it does not hurt this bound if some of the sets $\{U_1, U_2, \ldots, U_n\} \cap Q_i$ are empty.

To prove the left hand inequality consider the shortest path γ through $U_1, U_2, \ldots, U_{N(t)}$. If the two endpoints of γ do not lie in $\bigcup_i \partial Q_i$, extend the path γ so that the endpoints of the extended path $\bar{\gamma}$ lie in $\bigcup_i \partial Q_i$; this can be done in such a way that the length of $\bar{\gamma}$ is bounded by $|\gamma| + 2/m$, where $|\gamma|$ denotes the length of γ. Fix a square Q_i. The intersection $Q_i \cap \bar{\gamma}$ consists of a finite number of paths $\gamma_1, \gamma_2, \ldots, \gamma_k$ in Q_i each having its endpoints on ∂Q_i. Clearly, each point in $\{U_1, U_2, \ldots, U_{N(t)}\} \cap Q_i$ lies in $\bigcup_j \gamma_j$. The paths $\gamma_1, \gamma_2, \ldots, \gamma_k$ may be joined together by cutting and pasting and adding arcs $\beta_1, \ldots, \beta_{k-1}$ on ∂Q_i in such a way that no point in ∂Q_i lies on more than one of $\beta_1, \ldots, \beta_{k-1}$. Consequently, $\lambda_t^m(Q_i) \le |\gamma_j| + 4/m$. Summing over $i = 1, \ldots, m^2$ yields the left hand inequality. ■

Lemma 2. *For each $t > 0$,*

$$\lim_{m\to\infty} m^{-1} \sum_{i=1}^{m^2} \lambda_{m^2 t}^m(Q_i) = EL_{N(t)} \quad \textit{almost surely.}$$

Proof: This does not quite follow from the strong law of large numbers. But, since for each m the random variables $m\lambda_{m^2t}^m(Q_1), \ldots, m\lambda_{m^2t}^m(Q_{m^2})$ are i.i.d. with the same distribution as $L_{N(t)}$, we have by Chebyshev's inequality that

$$P\left\{\left|\sum_{i=1}^{m^2} \frac{\lambda_{m^2t}^m(Q_i)}{m} - EL_{N(t)}\right| > \epsilon\right\} \le \frac{\mathrm{Var}(L_{N(t)})}{m^2\epsilon^2}.$$

The assertion therefore follows from the Borel-Cantelli lemma. ■

LEMMA 3. *There exists $C \in (0, \infty)$ such that*

$$\lim_{t \to \infty} \frac{EL_{N(t)}}{t^{1/2}} = C.$$

PROOF: Take expectations in (AI.2) and use the fact that $E(m\lambda^m_{m^2t}(Q_i)) = EL_{N(t)}$ to obtain

$$-6 + EL_{N(t)} \leq \frac{EL_{N(m^2t)}}{m} \leq \sqrt{5} + EL_{N(t)}.$$

It follows that for any $\epsilon > 0$ there exists t sufficiently large that

$$\left| \frac{EL_{N(m^2t)}}{mt^{1/2}} - \frac{EL_{N(t)}}{t^{1/2}} \right| < \epsilon$$

for all $m = 1, 2, \ldots$. Since L_n is nondecreasing in n, this implies that

$$\begin{aligned} E\left(\frac{L_{N(t)}}{t^{1/2}}\right) - \epsilon &\leq \liminf_{s \to \infty} \frac{EL_{N(s)}}{s^{1/2}} \\ &\leq \limsup_{s \to \infty} \frac{EL_{N(s)}}{s^{1/2}} \\ &\leq \frac{EL_{N(t)}}{t^{1/2}} + \epsilon. \end{aligned}$$

Since $\epsilon > 0$ is arbitrary, it follows that $EL_{N(s)}/s^{1/2} \to C$ as $s \to \infty$ for some $0 \leq C < \infty$. To prove that $C > 0$, note that $EL_{N(t)} \to \infty$ as $t \to \infty$, by an elementary argument. Choose t sufficiently large that $EL_{N(t)} > 4$; then (AI.4) implies that

$$\liminf_{m \to \infty} \frac{EL_{N(m^2t)}}{m} > 0. \qquad \blacksquare$$

The proof of (AI.1) may now be completed. By (AI.2), for each $t > 0$ and $m = 1, 2, \ldots,$

$$\begin{aligned} -6 + m^{-1} \sum_{i=1}^{m^2} \lambda^m_{m^2t}(Q_i) &\leq \frac{L_{N(m^2t)}}{m} \\ &\leq \sqrt{5} + m^{-1} \sum_{i=1}^{m^2} \lambda^m_{m^2t}(Q_i), \end{aligned}$$

so Lemma 2 implies that almost surely

$$\begin{aligned} -6t^{-1/2} + \frac{EL_{N(t)}}{t^{1/2}} &\le \liminf_{m\to\infty} \frac{L_{N(m^2t)}}{mt^{1/2}} \\ &\le \limsup_{m\to\infty} \frac{L_{N(m^2t)}}{mt^{1/2}} \\ &\le \sqrt{5}\,t^{-1/2} + \frac{EL_{N(t)}}{t^{1/2}}. \end{aligned}$$

Now Lemma 3 implies that if t is sufficiently large then almost surely

$$\begin{aligned} C - \epsilon &\le \liminf_{m\to\infty} \frac{L_{N(m^2t)}}{mt^{1/2}} \\ &\le \limsup_{m\to\infty} \frac{L_{N(m^2t)}}{mt^{1/2}} \\ &\le C + \epsilon. \end{aligned}$$

Since L_n is nondecreasing in n it follows that

$$\begin{aligned} C - \epsilon &\le \liminf_{s\to\infty} \frac{L_N(s)}{s^{1/2}} \\ &\le \limsup_{s\to\infty} \frac{L_{N(s)}}{s^{1/2}} \\ &\le C + \epsilon \end{aligned}$$

almost surely. Now (AI.1) follows by letting $\epsilon \to 0$. ∎

Appendix II. Paley's Square Function Argument

This Appendix develops an argument for martingales that was introduced in Paley (1932) for Walsh functions. The only real changes made here to Paley's method are those required to provide explicit bounds on the basic constant. As one should expect, the constant is not as sharp as that given in Burkholder (1973), but the reason for reviewing Paley's argument is rather to show how the maximal function can be used to bound L_p norms of martingales. Other features of the proof are discussed at the end of the appendix.

Consider a martingale difference sequence $\{y_i : 0 \le i \le n\}$ with $y_0 \equiv 0$ and its associated martingale $M_k = y_1 + y_2 + \cdots + y_k,\ 1 \le k \le n$. To keep to the essentials, we will stick to the case of even integers p. We first compute the difference sequence of pth moments:

$$\begin{aligned} E\{M_{k+1}^p - M_k^p\} &= E\{(M_k + y_{k+1})^p - M_k^p\} \\ &= E\left\{ p y_{k+1} M_k^{p-1} + \binom{p}{2} y_{k+1}^2 M_k^{p-2} + \cdots + y_{k+1}^p \right\} \\ &= E\left\{ \binom{p}{2} y_{k+1}^2 M_k^{p-2} + \binom{p}{3} y_{k+1}^3 M_k^{p-3} + \cdots + y_{k+1}^p \right\} \end{aligned} \tag{AII.1}$$

where only in the last inequality is the martingale property invoked. We then use Hölder's inequality on the right hand side to bring the powers of M_k up to the same level. Specifically, for $3 \leq j \leq p-1$ we use

$$E y_{k+1}^j M_k^{p-j} \leq (E y_{k+1}^2 M_k^{p-2})^{\theta} (E y_{k+1}^p)^{1-\theta} \tag{AII.2}$$

where $\theta = (p-j)/(p-2)$. Since $0 < \theta \leq 1$, inequality (AII.2) can be relaxed to

$$E y_{k+1}^j M_k^{p-j} \leq E y_{k+1}^2 M_k^{p-2} + E y_{k+1}^p, \tag{AII.3}$$

so we can crudely bound the sum of the binomial coefficients to find

$$\left| E\{M_{k+1}^p - M_k^p\} \right| \leq 2^p \left\{ E y_{k+1}^2 M_k^{p-2} + E y_{k+1}^p \right\}. \tag{AII.4}$$

Finally, we sum over $0 \leq k < n$ to find

$$\begin{aligned} E M_n^p &\leq 2^p E \left\{ \left(\sum_{k=1}^n y_k^2 \right) \max_{1 \leq k \leq n} M_k^{p-2} \right\} + 2^p E \sum_{k=0}^{n-1} y_{k+1}^p \\ &\leq 2^p \left\{ E \left(\sum_{k=1}^n y_k^2 \right)^{p/2} \right\}^{2/p} \left(E \max_{1 \leq k \leq n} M_k^p \right)^{(p-2)/p} + 2^p E \left(\sum_{k=1}^n y_k^2 \right)^{p/2} \end{aligned} \tag{AII.5}$$

where in the first summand we used Hölder's inequality, and in the second summand we used the elementary real variable inequality for $p \geq 2$

$$a_1^p + a_2^p + \cdots + a_n^p \leq (a_1^2 + a_2^2 + \cdots + a_n^2)^{p/2}.$$

Our motivation for moving from (AII.4) to (AII.5) is to use Doob's maximal inequality, or rather its consequence for $1 < p < \infty$ that

$$\left\{ E \left(\max_{1 \leq k \leq n} |M_k|^p \right) \right\}^{1/p} \leq q (E M_n^p)^{1/p}, \tag{AII.6}$$

where q is the conjugate index to p (i.e. $q = p/(p-1)$). From (AII.6) we thus find

$$\|M_n\|_p^p \le q2^p \left\| \left(\sum_{k=1}^n y_k^2 \right)^{1/2} \right\|_p^2 \|M_n\|_p^{p-2} + 2^p \left\| \left(\sum_{k=1}^n y_k^2 \right)^{1/2} \right\|_p^p. \qquad \text{(AII.7)}$$

Finally, we note inequality (AII.7) is of the form $x^p \le ay^2x^{p-2} + by^p$ which implies $x \le \{(2a)^{1/2} + (2b)^{1/p}\}y$, so we find our modest version of Burkholder's inequality for even integers p:

$$\|M_n\|_p \le \alpha_p \left\| \left(\sum_{k=1}^n y_k^2 \right)^{1/2} \right\|_p \qquad \text{(AII.8)}$$

where $\alpha_p \le q^{1/2}2^{(p+1)/2} + 2^{(p+1)/2} \le q^{1/2}2^{(p+3)/2}$. ■

The constant α_p is larger than the $18q^{1/2}p$ we know to be sufficient, so some comment seems needed to justify our enthusiasm for this more-than-fifty year old argument. First, it uses very little about martingales; e.g. in (AII.1) we use a weak consequence of the definition, and the only other fact we need is a maximal inequality of Doob's type as given in (AII.6). Second, the differencing applied to pth powers in (AII.1) can be applied to other functions f of M_k, provided that $f(M_k + y_k) - f(M_k)$ can be bounded by a useful expression. Finally, since the argument is free of stopping times, its parts are amenable to more individual attention. In particular, the use of bounds on $\|y_i\|_p$, $1 \le p \le \infty$, can be tried out in AII.5, AII.6, or AII.7.

Added in Proof: The idea of using a spacefilling curve to sequence visits to points in the square is evidently much older than recent references seem to indicate. From the comments of R. Adler in the *Collected Works of S. Kakutani* (Kakutani 1986, V.II, p. 445), Kakutani had presented the idea as early as the spring of 1966.

Program in Statistics and Operations Research
School of Engineering and Applied Science
Princeton University
Princeton
New Jersey 08544.