BEARDWOOD-HALTON-HAMMERSLEY THEOREM FOR STATIONARY ERGODIC SEQUENCES: CONSTRUCTION OF A COUNTEREXAMPLE

ALESSANDRO ARLOTTO AND J. MICHAEL STEELE

Abstract. We construct a stationary ergodic process $\{X_1, X_2, X_3 \ldots\}$ such that each X_t, $1 \leq t < \infty$, has the uniform distribution on the unit square and the length L_n of the shortest path through $\{X_1, X_2, \ldots, X_n\}$ is *not* asymptotic to a constant times the square root of n. In other words, we show that the Beardwood, Halton and Hammersley theorem does not extend from the case of independent uniformly distributed random variables to the case of stationary ergodic sequences with the uniform stationary distribution.

Mathematics Subject Classification (2000): Primary 60D05, 90B15; Secondary 60F15, 60G10, 60G55, 90C27.

Key Words: traveling salesman problem, Beardwood-Halton-Hammersley theorem, subadditive Euclidean functional, stationary ergodic processes, equidistribution, construction of stationary processes.

1. Introduction

Given a set $\mathcal{S}_n = \{x_1, x_2 \ldots, x_n\}$ of n points in the unit square $[0,1]^2$, we let $L(x_1, x_2, \ldots, x_n)$ denote the length of the shortest path through the points in $\mathcal{S}_n$; that is, we have

$$L(x_1, x_2, \ldots, x_n) = \min_{\sigma} \sum_{t=1}^{n-1} |x_{\sigma(t)} - x_{\sigma(t+1)}|, \tag{1}$$

where the minimum is over all permutations $\sigma : [1:n] \to [1:n]$ and where $|x-y|$ denotes the usual Euclidian distance between elements x and y of $[0,1]^2$. The classical theorem of Beardwood, Halton and Hammersley (1959) tells us that if $X_1, X_2, \ldots$ is a sequence of independent random variables with the uniform distribution on $[0,1]^2$, then there is a constant $\beta > 0$ such that

$$\lim_{n\to\infty} \frac{L(X_1, X_2, \ldots, X_n)}{\sqrt{n}} = \beta \quad \text{with probability one.} \tag{2}$$

The Beardwood, Halton and Hammersley (BHH) constant β has been studied extensively, and, although its exact value is still unknown, sophisticated numerical computations (Applegate et al., 2006) suggest that $\beta \approx 0.714\ldots$. The best

Date: June 30, 2013.

A. Arlotto: Fuqua School, Duke University, 100 Fuqua Drive, Durham, NC, 27708. Email address: `alessandro.arlotto@duke.edu`.

J. M. Steele: Wharton School, University of Pennsylvania, Department of Statistics, Huntsman Hall 447, Philadelphia, PA, 19104. Email address: `steele@wharton.upenn.edu`.

available analytical bounds are much rougher; we only know with certainty that $0.62499 \le \beta \le 0.91996$ (Finch, 2003, pp. 497–498).

The BHH theorem (2) is a strong law for independent identically distributed random variables, so it is natural to ask if there is an *analogous ergodic theorem* where one relaxes the hypotheses to those of the classic ergodic theorem for partial sums. Our main goal here is to answer this question. Specifically, we construct a stationary ergodic process with the uniform invariant measure on $[0,1]^2$ for which the length of the shortest path through n points is *not* asymptotic to $\beta\sqrt{n}$.

Theorem 1 (No Ergodic BHH). *There are constants $c_1 < c_2$ and a strictly stationary and ergodic process $\{X_t : 1 \le t < \infty\}$ such that each X_t, $1 \le t < \infty$, is uniformly distributed on $[0,1]^2$ and such that with probability one*

$$(3) \qquad \liminf_{n\to\infty} \frac{L(X_1, X_2, \ldots, X_n)}{\sqrt{n}} \le c_1 < c_2 \le \limsup_{n\to\infty} \frac{L(X_1, X_2, \ldots, X_n)}{\sqrt{n}}.$$

This theorem is obtained as a corollary of the next theorem where the condition of ergodicity is dropped. In this case, one can construct processes for which there is a more explicit control of the expected minimal path length.

Theorem 2 (Asymptotics of Expected Path Lengths). *There is a strictly stationary process $\{X_t^* : 1 \le t < \infty\}$ such that each X_t^*, $1 \le t < \infty$, is uniformly distributed on $[0,1]^2$ and such that*

$$(4) \quad \liminf_{n\to\infty} \frac{\mathbb{E}[L(X_1^*, X_2^*, \ldots, X_n^*)]}{\sqrt{n}} \le \beta/\sqrt{2} < \beta \le \limsup_{n\to\infty} \frac{\mathbb{E}[L(X_1^*, X_2^*, \ldots, X_n^*)]}{\sqrt{n}};$$

where β is the BHH constant.

Analogous constructions prove, more broadly, that there are no ergodic analogs for many other subadditive Euclidean functionals. We return to these and other general considerations in Section 10 where we also describe some open problems.

The Big Picture

The main idea of the proof of Theorem 2 is that one can construct a stationary process such that at successive (and ever larger) times N, the set of observations up to time N either looks very much like an independent uniformly distributed sample or else looks like a sample that has far too many "twin cities," i.e. points that are quite close together on a scale that depends on N. To make this idea precise, we use a sequence of parameterized transformation of stationary processes where each transformation adds a new epoch with too many twin cities — *at an appropriate scale*. Finally, we show that one can build a single stationary process with infinitely many such epochs.

This limit process provides the desired example of a stationary, uniform process for which the minimal length paths have expectations that behave much differently from those of Beardwood, Halton, and Hammersley. Finally, the ergodic process required by Theorem 1 is obtained by a brief extreme point argument that uses the representation of a stationary process as a mixture of stationary ergodic processes.

2. Two Transformations of a Stationary Sequence: the $H(\epsilon, N)$ and $T(\epsilon, N)$ Transformations

Our construction exploits an iterative process that transforms a given stationary process into another stationary process with additional properties. Given any doubly infinite sequence of $\mathbb{R}^d$-valued random variables $\mathcal{X} = \{\dots, X_{-1}, X_0, X_1, ...\}$ and any pair of integers a, b such that $a \le b$, we denote the $[a:b]$ segment of $\mathcal{X}$ by

$$\mathcal{X}[a:b] = (X_a, X_{a+1}, \dots, X_b).$$

Next, we say that the process $\mathcal{X}$ is *periodic in distribution* with period p if

$$\mathcal{X}[a:a+k] \stackrel{d}{=} \mathcal{X}[b:b+k] \quad \text{for all } k \ge 0 \text{ and all } b \text{ such that } b \equiv a \mod p. \tag{5}$$

This is certainly a weaker condition than strict stationarity, but, by an old randomization trick, one can transform a process that is periodic in distribution to a closely related process that is strictly stationarity. We will eventually apply this construction infinitely many times, so to fix ideas and notation, we first recall how it works in the simplest setting.

Lemma 3 (Passage from Periodicity in Distribution to Strict Stationarity)**.** *If the $\mathbb{R}^d$-valued doubly infinite sequence $\widehat{\mathcal{X}} = \{\dots, \widehat{X}_{-1}, \widehat{X}_0, \widehat{X}_1, ...\}$ is periodic in distribution with period p, and, if I is chosen independently and uniformly from $\{0, 1, 2, \dots, p-1\}$, then the doubly infinite sequence $\widetilde{\mathcal{X}} = \{\dots, \widetilde{X}_{-1}, \widetilde{X}_0, \widetilde{X}_1, ...\}$ defined by setting*

$$\widetilde{X}_t = \widehat{X}_{t+I} \quad \textit{for all } t \in \mathbb{Z}$$

is a strictly stationary process.

Proof. Fix $0 \le j < \infty$ and take Borel subsets $A_0, A_1, \dots, A_j$ of $\mathbb{R}^d$. By the definition of $\widetilde{\mathcal{X}}$ and by conditioning on I, one then has

$$\begin{aligned}
&\mathbb{P}(\widetilde{X}_t \in A_0, \widetilde{X}_{t+1} \in A_1, \dots, \widetilde{X}_{t+j} \in A_j) \\
&\qquad = \frac{1}{p} \sum_{i=0}^{p-1} \mathbb{P}(\widehat{X}_{t+i} \in A_1, \widehat{X}_{t+1+i} \in A_2, \dots, \widehat{X}_{t+j+i} \in A_j) \\
&\qquad = \frac{1}{p} \sum_{i=0}^{p-1} \mathbb{P}(\widehat{X}_{t+1+i} \in A_1, \widehat{X}_{t+2+i} \in A_2, \dots, \widehat{X}_{t+1+j+i} \in A_j) \qquad (6) \\
&\qquad = \mathbb{P}(\widetilde{X}_{t+1} \in A_0, \widetilde{X}_{t+2} \in A_1, \dots, \widetilde{X}_{t+1+j} \in A_j),
\end{aligned}$$

where the periodicity in distribution of $\widehat{\mathcal{X}}$ is used to obtain (6). Specifically, by periodicity in distribution, the last summand of (6) is equal to the first summand of the preceding sum. This tells us that $\mathbb{P}(\widetilde{X}_t \in A_0, \widetilde{X}_{t+1} \in A_1, \dots, \widetilde{X}_{t+j} \in A_j)$ does not depend on t, and, since j is arbitrary, we see that $\widetilde{\mathcal{X}}$ is stationary. $\square$

The rest of our analysis focuses on processes with values in the unit square, $[0,1]^2$, and it will be convenient to view these processes as taking values in the "flat torus" $\mathcal{T}$ that one gets by identifying the opposite sides of the $[0,1]^2$ while retaining the standard Riemannian metric. This point of view does no harm since the same asymptotic relation (2) holds regardless of whether the sequence $X_1, X_2, \dots, X_n$ is drawn from the unit square $[0,1]^2$ or the flat torus $\mathcal{T}$. The benefit of using the flat torus is that it has a natural additive group structure, and for $X = (x, y) \in \mathcal{T}$ and $0 < \epsilon < 1$, we can define the ϵ-translation $X(\epsilon)$ of X, by setting

$$X(\epsilon) = (\,(x + \epsilon) \bmod 1,\, y). \tag{7}$$

That is, we get $X(\epsilon)$ by shifting X by ϵ in just the first coordinate and the shift is taken modulo 1.

Incidentally, the torus model was considered earlier by Avram and Bertsimas (1992) for the minimum spanning tree problem. A little later, Jaillet (1993) exploited the asymptotic equivalence between the torus and the unit-square models for the minimum spanning tree, the travel salesman, and other combinatorial optimization problems.

We now consider a stationary process $\mathcal{X} = \{\ldots, X_{-1}, X_0, X_1, ...\}$ with values in $\mathcal{T}$. Given an $N \in \mathbb{N}$, we define "blocks" B_k, $k \in \mathbb{Z}$, of length $2N$ by setting

$$B_k = (X_{kN}, X_{kN+1}, \ldots, X_{(k+1)N-1}, X_{kN}(\epsilon), X_{kN+1}(\epsilon), \ldots, X_{(k+1)N-1}(\epsilon)), \tag{8}$$

where the translations $X_t(\epsilon)$, $kN \le t < (k+1)N$, are defined as in (7). We write the doubly infinite concatenation of these blocks

$$\{\ldots B_{-2}, B_{-1}, B_0, B_1, B_2, \ldots\},$$

and we note that this gives us a doubly infinite sequence of $\mathcal{T}$-valued random variables that we write as

$$\widehat{\mathcal{X}} = \{\ldots \widehat{X}_{-2}, \widehat{X}_{-1}, \widehat{X}_0, \widehat{X}_1, \widehat{X}_2, \ldots\}.$$

The blocks B_k, $k \in \mathbb{Z}$, can then be expressed in terms of segments of $\widehat{\mathcal{X}}$; specifically, for all $k \in \mathbb{Z}$, we also have

$$B_k = (\widehat{X}_{2kN}, \widehat{X}_{2kN+1}, \widehat{X}_{2kN+2}, \ldots, \widehat{X}_{(2k+1)N}, \ldots, \widehat{X}_{2(k+1)N-1}).$$

The process $\widehat{\mathcal{X}} = \{\widehat{X}_t : t \in \mathbb{Z}\}$ is called the *hat-process*, and the passage from $\mathcal{X}$ to $\widehat{\mathcal{X}}$ is called an $H(\epsilon, N)$-*transformation.* We denote this transformation by

$$\mathcal{X} \xrightarrow{H(\epsilon,N)} \widehat{\mathcal{X}}. \tag{9}$$

This hat-process $\widehat{\mathcal{X}} = \{\widehat{X}_t : t \in \mathbb{Z}\}$ is periodic in distribution with period $2N$, so one can use Lemma 3 to construct a closely related stationary sequence $\widetilde{\mathcal{X}} = \{\widetilde{X}_t : t \in \mathbb{Z}\}$. Specifically, we set

$$\widetilde{X}_t = \widehat{X}_{t+I} \quad \text{for all } t \in \mathbb{Z},$$

where the random index I has the uniform distribution on $\{0, 1, \ldots, 2N-1\}$ and I is independent of the sequence $\widehat{\mathcal{X}}$. The complete passage from $\mathcal{X}$ to $\widetilde{\mathcal{X}}$ is called a $T(\epsilon, N)$-*transformation*, and it is denoted by

$$\mathcal{X} \xrightarrow{T(\epsilon,N)} \widetilde{\mathcal{X}}. \tag{10}$$

We will make repeated use of the two-step nature of this construction, and it should be stressed that the hat-process $\widehat{\mathcal{X}}$ is more than a intermediate product. The properties of the hat-process $\widehat{\mathcal{X}}$ are the real guide to our constructions, and the stationary process $\widetilde{\mathcal{X}}$ is best viewed as a more polished version of $\widehat{\mathcal{X}}$.

Properties of the $T(\epsilon, N)$-transformation

The process $\widetilde{\mathcal{X}}$ that one obtains from $\mathcal{X}$ by a $T(\epsilon, N)$-transformation tends to retain much of the structure of $\mathcal{X}$. We begin with a simple example.

Lemma 4 (Preservation of Uniform Marginals)**.** *Let $\mathcal{X}$ be a doubly infinite $\mathcal{T}$-valued process. Given $0 < \epsilon < 1$ and $N \in \mathbb{N}$, let $\widetilde{\mathcal{X}}$ be the process defined by the transformation*

$$\mathcal{X} \xrightarrow{T(\epsilon,N)} \widetilde{\mathcal{X}}.$$

If X_t has the uniform distribution on $\mathcal{T}$ for each $t \in \mathbb{Z}$, then $\widetilde{X}_t$ has the uniform distribution on $\mathcal{T}$ for each $t \in \mathbb{Z}$.

Proof. The definition of the hat-process tells us that each $\widehat{X}_t$, $t \in \mathbb{Z}$, has the uniform distribution on $\mathcal{T}$. Thus, for each $s \in \mathbb{Z}$, the distribution of $\widetilde{X}_s$ is a mixture of uniform distributions, and hence the distribution of $\widetilde{X}_s$ is also uniform on $\mathcal{T}$. $\square$

Given a doubly infinite $\mathcal{T}$-valued sequence $\mathcal{X} = (\ldots, X_{-1}, X_0, X_1, \ldots)$ and a segment $\mathcal{X}[a:b] = (X_a, X_{a+1}, \ldots, X_b)$, we now consider the translation of the segment by $\delta > 0$:

$$\mathcal{X}[a:b](\delta) = (X_a(\delta), X_{a+1}(\delta), \ldots, X_b(\delta)).$$

As before, we have $X_t = (\xi_t, \xi_t') \in \mathcal{T}$ and $X_t(\delta) = (\xi_t + \delta, \xi_t')$ where the addition in the first coordinate is taken modulo one, and we say that $\mathcal{X}$ is *translation invariant* (or more precisely, δ-*translation invariant*) if

$$\mathcal{X}[a:b] \stackrel{d}{=} \mathcal{X}[a:b](\delta) \quad \text{for all } -\infty < a \le b < \infty \text{ and } \delta > 0.$$

This property is also preserved by a $T(\epsilon, N)$-transformation.

Lemma 5 (Preservation of Translation Invariance)**.** *If $\mathcal{X}$ is a doubly-infinite $\mathcal{T}$-valued process that is δ-translation invariant for some $\delta > 0$, then for each $0 < \epsilon < 1$ and $N \in \mathbb{N}$, the process $\widetilde{\mathcal{X}}$ defined by*

$$\mathcal{X} \xrightarrow{T(\epsilon,N)} \widetilde{\mathcal{X}}$$

is also δ-translation invariant.

Proof. By conditioning on I, it suffices to show that $\widehat{\mathcal{X}}$ is translation invariant, and, by the consistency of distributions of segments and subsegments, it suffices to show that

$$\widehat{\mathcal{X}}[0:2kN-1](\delta) \stackrel{\mathrm{d}}{=} \widehat{\mathcal{X}}[0:2kN-1] \stackrel{\mathrm{def}}{=} (B_0, B_1, \ldots, B_{k-1}), \quad \text{for all } k \ge 1.$$

For $k = 1$, one just needs to note that

$$\begin{aligned} B_0(\delta) &= (X_0(\delta), X_1(\delta), \ldots, X_{N-1}(\delta), X_0(\delta+\epsilon), X_1(\delta+\epsilon), \ldots, X_{N-1}(\delta+\epsilon)) \\ &\stackrel{\mathrm{d}}{=} (X_0, X_1, \ldots, X_{N-1}, X_0(\epsilon), X_1(\epsilon), \ldots, X_{N-1}(\epsilon)) \stackrel{\mathrm{def}}{=} B_0, \end{aligned}$$

where in the second line we used the translation invariance of $\mathcal{X}$. The case of compound blocks works in just the same way. $\square$

The process $\widetilde{\mathcal{X}}$ that one obtains from a doubly infinite stationary sequence $\mathcal{X}$ by a $T(\epsilon, N)$-transformation is typically singular with respect to $\mathcal{X}$. Nevertheless, on "short" segments the two processes are close in distribution. The next lemma makes this precise.

Lemma 6 (Closeness in Distribution)**.** *Let $\mathcal{X}$ be a translation-invariant doubly-infinite stationary sequence with values in the flat torus $\mathcal{T}$. For each $0 < \epsilon < 1$ and $N \in \mathbb{N}$, the process $\widetilde{\mathcal{X}}$ defined by $\mathcal{X} \xrightarrow{T(\epsilon,N)} \widetilde{\mathcal{X}}$ satisfies*

$$| \, \mathbb{P}(\widetilde{\mathcal{X}}[0:m] \in A) - \mathbb{P}(\mathcal{X}[0:m] \in A) \, | \le \frac{m}{N}, \tag{11}$$

for all Borel sets $A \subseteq \mathcal{T}^{m+1}$ and for all $m = 0, 1, 2, \ldots$.

Proof. Recalling the two-step construction that takes one from $\mathcal{X}$ to $\widetilde{\mathcal{X}}$, we first note that we can write $\widetilde{\mathcal{X}}[0:m]$ in terms of the hat-process $\widehat{\mathcal{X}}$ given by the construction (9); specifically, we have

$$\widetilde{\mathcal{X}}[0:m] = \widehat{\mathcal{X}}[I:I+m], \tag{12}$$

where the random variable I is independent of $\widehat{\mathcal{X}}$ and uniformly distributed on $\{0,1,\ldots,2N-1\}$. Now we condition on the value i of I. For any i such that $[i:i+m]\subseteq[0:N-1]$, the definition of the hat-process gives us the distributional identity

$$\widehat{\mathcal{X}}[i:i+m] = \mathcal{X}[i:i+m] \stackrel{d}{=} \mathcal{X}[0:m], \tag{13}$$

where in the last step we used the stationarity of $\mathcal{X}$. Similarly, for i such that $[i:i+m]\subseteq[N:2N-1]$, we have

$$\widehat{\mathcal{X}}[i:i+m] = \mathcal{X}[i-N:i-N+m](\epsilon) \stackrel{d}{=} \mathcal{X}[i-N:i-N+m] \stackrel{d}{=} \mathcal{X}[0:m], \tag{14}$$

where in the next-to-last step we use the translation invariance of $\mathcal{X}$, and in the last step we again used the stationarity of $\mathcal{X}$.

We now consider the "good set" of indices

$$G=\{i : 0\le i\le i+m<N \text{ or } N\le i\le i+m<2N\},$$

where the equalities (13) and (14) hold, and we also consider the complementary "bad set" of indices $G'=[0:2N-1]\backslash G$. We then condition on I and use (13) and (14) to obtain that

$$\begin{aligned}\mathbb{P}(\widetilde{\mathcal{X}}[0:m]\in A) &= \frac{1}{2N}\sum_{i\in G}\mathbb{P}(\widehat{\mathcal{X}}[i:i+m]\in A)+\frac{1}{2N}\sum_{i\in G'}\mathbb{P}(\widehat{\mathcal{X}}[i:i+m]\in A)\\ &= \frac{2N-2m}{2N}\,\mathbb{P}(\mathcal{X}[0:m]\in A)+\frac{1}{2N}\sum_{i\in G'}\mathbb{P}(\widehat{\mathcal{X}}[i:i+m]\in A),\end{aligned}$$

which is written more nicely as

$$\begin{aligned}\mathbb{P}(\widetilde{\mathcal{X}}[0:m]\in A)-\mathbb{P}(\mathcal{X}[0:m]\in A) = &-\frac{m}{N}\,\mathbb{P}(\mathcal{X}[0:m]\in A)\\ &+\frac{1}{2N}\sum_{i\in G'}\mathbb{P}(\widehat{\mathcal{X}}[i:i+m]\in A).\end{aligned}$$

The last sum has only $|G'|=2m$ terms, so we have the bounds

$$-\frac{m}{N}\le \mathbb{P}(\widetilde{\mathcal{X}}[0:m]\in A)-\mathbb{P}(\mathcal{X}[0:m]\in A)\le\frac{m}{N},$$

which complete the proof of the lemma. $\square$

3. Locally Uniform Processes: The Mean BHH Lemma

Our inductive construction requires a version of the Beardwood, Halton and Hammersley theorem for a certain class of dependent processes. We call these *locally uniform processes*, and to define them we need some notation.

First, we say that $Q\subseteq\mathcal{T}$ is a *sub-square of side length* α if it can be written as $[x,x+\alpha]\times[y,y+\alpha]$ where one makes the usual identifications of the points in the flat torus, and, for $0<\alpha<1$, we let $\mathcal{Q}(\alpha)$ denote the set of all sub-squares of $\mathcal{T}$ that have length less than or equal to α. Given a sequence of random variables $\mathcal{Y}=\{Y_t : t\in\mathbb{Z}\}$ and a set of indices $J\subseteq\mathbb{Z}$, we let $\mathcal{Y}[J]=\{Y_t : t\in J\}$, and, given a Borel set $A \subseteq \mathcal{T}$, we write $N(A, \mathcal{Y}[J])$ for the cardinality of the set $\{A \cap \mathcal{Y}[J]\}$. Finally, we use $\lambda(\cdot)$ to denote the Lebesgue measure on $\mathcal{T}$.

Definition 7 (Locally Uniform Processes)**.** If $\mathcal{Y} = \{Y_t : t \in \mathbb{Z}\}$ is a $\mathcal{T}$-valued process with uniform marginal distributions, we say that $\mathcal{Y}$ is (α, M)*-locally uniform* if for each interval of indices $[a : b]$ there exists a set of indices $J = J([a : b])$ such that

$$J \subseteq [a : b] \quad \text{and} \quad 0 \leq |[a : b]| - |J| \leq M,$$

and such that for each Borel set $A \subseteq \mathcal{Q}(\alpha)$ one has the variance of $N(A, \mathcal{Y}[J])$ bounded by its mean,

$$\operatorname{Var}(N(A, \mathcal{Y}[J])) \leq \mathbb{E}[N(A, \mathcal{Y}[J])], \tag{15}$$

as well as the conditional-uniformity condition

$$\{A \cap \mathcal{Y}[J]\} \stackrel{\text{dpp}}{=} \{U_s(A) : 1 \leq s \leq N(A, \mathcal{Y}[J])\}, \tag{16}$$

where the random variables $U_s(A)$, $1 \leq s < \infty$, are independent with the uniform distribution on A and where $\stackrel{\text{dpp}}{=}$ denotes equality in distribution as point processes.

The main task now is to show that locally uniform processes satisfy a relaxed version of the BHH theorem.

Lemma 8 (BHH in Mean for Locally Uniform Processes)**.** *If a $\mathcal{T}$-valued process $\mathcal{Y}$ is (α, M)-locally uniform for some $0 < \alpha < 1$ and some $M < \infty$, then*

$$\mathbb{E}[L(Y_1, Y_2, \ldots, Y_n)] \sim \beta\sqrt{n} \quad \text{as } n \to \infty, \tag{17}$$

where β is the BHH constant in (2)*.*

Proof. For any integer k such that $1/k \leq \alpha$ we consider the natural decomposition of the flat torus into k^2 sub-squares Q_i, $i = 1, 2, \ldots k^2$ of side length $1/k$. Given a set $\mathcal{S}_m \equiv \{y_1, y_2, \ldots, y_m\}$ of m points in $\mathcal{T}$, we let $L(\mathcal{S}_m) \equiv L(y_1, y_2, \ldots, y_m)$, and we note that, for each i, the optimal path through the subset $Q_i \cap \mathcal{S}_m$ has cost $L(Q_i \cap \mathcal{S}_m)$, $1 \leq i \leq k^2$.

We then stitch the k^2 together by considering the sub-squares Q_i, $1 \leq i \leq k^2$, in plowman's order — down one row then back the next. One can check that the stitching cost is less than $3k$, but all we need from these considerations is that there is a universal constant $C_1 > 0$ such that, for all $\mathcal{S}_m = \{y_1, y_2, \ldots, y_m\} \subseteq \mathcal{T}$, one has

$$L(\mathcal{S}_m) \leq C_1 k + \sum_{i=1}^{k^2} L(Q_i \cap \mathcal{S}_m). \tag{18}$$

More notably, one can also show that there is a universal constant $C_0 > 0$ for which one has

$$- C_0 k + \sum_{i=1}^{k^2} L(Q_i \cap \mathcal{S}_m) \leq L(\mathcal{S}_m). \tag{19}$$

This bound is due to Redmond and Yukich (1994), and it may be proved by noticing that the sum of the values $L(Q_i \cap \mathcal{S}_m)$ can be bounded by the length of the optimum path through $\mathcal{S}_m$ and the sum of lengths of the boundaries of the individual squares Q_i, $1 \leq i \leq k^2$. For the details of this argument — including extensions to $[0, 1]^d$, $d \geq 2$ — one can consult Yukich (1998, Chapter 3).

By the (α, M)-local uniformity of $\mathcal{Y}$, we know there is a set of indices J_n with cardinality $m = n + O(1)$ such that the variance and the conditional-uniformity conditions in Definition (7) hold for $\mathcal{Y}[J_n]$. Then, we immediately obtain for each $1 \leq i \leq k^2$ that

$$\mathbb{E}[L(Q_i \cap \mathcal{Y}[1:n])] \sim \mathbb{E}[L(Q_i \cap \mathcal{Y}[J_n])] \qquad \text{as } n \to \infty,$$

so, given the pointwise bounds (18) and (19), the lemma will follow once we show that

$$\mathbb{E}[L(Q_i \cap \mathcal{Y}[J_n])] \sim \beta k^{-2}\sqrt{n} \qquad \text{as } n \to \infty. \tag{20}$$

By the conditional-uniformity condition (16) the set $Q_i \cap \mathcal{Y}[J_n]$ has the same point process distribution as a uniform random sample from Q_i that has size equal to $N(n) \equiv N(Q_i, \mathcal{Y}[J_n])$. Hence, if $\{U_i : i = 1, 2, \ldots\}$ is a sequence of independent random variables with the uniform distribution on the flat torus $\mathcal{T}$, then we have by length-scaling of the shortest-path function, L, that

$$\begin{aligned}\mathbb{E}[L(Q_i \cap \mathcal{Y}[J_n])] &= k^{-1}\mathbb{E}[L(U_1, U_2, \ldots, U_{N(n)})]\} \\ &= k^{-1}\sum_{j=0}^{n} \ell(j)\mathbb{P}(N(n) = j),\end{aligned}$$

where, in the second step, we have conditioned on the value of $N(n)$ and we have set $\ell(j) = \mathbb{E}[L(U_1, U_2, \ldots, U_j)]$. The BHH relation (2) then tells us that we have $\ell(j) = \beta\sqrt{j} + o(\sqrt{j})$, so

$$\begin{aligned}\mathbb{E}[L(Q_i \cap \mathcal{Y}[J_n])] &= \beta k^{-1}\sum_{j=0}^{n} \sqrt{j}\,\mathbb{P}(N(n) = j) + o(\sqrt{n}) \\ &= \beta k^{-1}\mathbb{E}[\sqrt{N(n)}] + o(\sqrt{n}).\end{aligned} \tag{21}$$

Now, we just need to estimate $\mathbb{E}[\sqrt{N(n)}]$. We first note that one gets

$$\mathbb{E}[N(n)] = \lambda(Q_i)|\, J_n \,| \sim k^{-2}\sqrt{n} \qquad \text{as } n \to \infty, \tag{22}$$

from linearity of the expectation and the fact that each Y_t has the uniform distribution. From (22) and Jensen's inequality, we then obtain the upper bound

$$\mathbb{E}[\sqrt{N(n)}] \leq \sqrt{\mathbb{E}[N(n)]} \sim k^{-1}\sqrt{n} \qquad \text{as } n \to \infty. \tag{23}$$

To get a comparable lower bound, we take $0 < \theta < 1$ and note by Chebyshev's inequality, the variance condition (15), and the asymptotic relation (22) that

$$\mathbb{P}(N(n) < \theta\mathbb{E}[N(n)]) = O(1/n) \quad \text{as } n \to \infty,$$

so if we take expectations in the pointwise bound

$$\theta^{1/2}\{\mathbb{E}[N(n)]\}^{1/2}\mathbb{1}\{N(n) \geq \theta\mathbb{E}[N(n)]\} \leq N(n)^{1/2},$$

we obtain

$$\theta^{1/2}\{\mathbb{E}[N(n)]\}^{1/2}(1 - O(1/n)) \leq \mathbb{E}[\sqrt{N(n)}]. \tag{24}$$

From the relations (23), (24), and the arbitrariness of θ, we then have

$$\mathbb{E}[\sqrt{N(n)}] \sim k^{-1}\sqrt{n} \quad \text{as } n \to \infty,$$

and, together with (21), this completes the proof of (20) and the lemma. $\square$

4. Preservation of Local Uniformity

A major benefit of local uniformity is that it is preserved by the $H(\epsilon, N)$ and $T(\epsilon, N)$ transformations.

Proposition 9 (Local Uniformity)**.** *Consider a $\mathcal{T}$-valued process $\mathcal{X}$ with uniform marginal distributions such that $\mathcal{X}$ is (α, M)-locally uniform for some $0 < \alpha < 1$ and $M < \infty$. For any $0 < \epsilon < \alpha$ and $N < \infty$, if one has*

$$\mathcal{X} \xrightarrow{H(\epsilon,N)} \widehat{\mathcal{X}},$$

then $\widehat{\mathcal{X}}$ is $(\gamma, 2M + 4N)$-locally uniform for all $0 < \gamma \leq \min\{\epsilon, \alpha - \epsilon\}$.

The $H(\epsilon, N)$-transformation defined by (9) is a purely deterministic transformation of $\mathcal{X}$, and one can identify the elements of $\widehat{\mathcal{X}}$ with elements of $\mathcal{X}$ and their ϵ-shifts. Specifically, for each $t \in \mathbb{Z}$ there is a unique pair (s, k) with $0 \leq s \leq N - 1$ and $k \in \mathbb{Z}$ such that

$$\widehat{X}_t = \begin{cases} X_{kN+s} & \text{if } t = 2kN + s \\ X_{kN+s}(\epsilon) & \text{if } t = (2k + 1)N + s. \end{cases} \tag{25}$$

For any segment $[a : b]$ we consider the largest sequence of contiguous complete $2N$-blocks that is contained in $[a : b]$; that is, we let

$$k_a = \inf\{k : a \leq 2kN\} \quad \text{and} \quad k_b = \sup\{k : 2kN - 1 \leq b\}, \tag{26}$$

and we consider the interval $[2k_aN : 2k_bN - 1] \subseteq [a : b]$. Here, $2k_aN - a \leq 2N$ and $b - 2k_bN \leq 2N$, so we have

$$0 \leq |[a : b]| - |[2k_aN : 2k_bN - 1]| \leq 4N. \tag{27}$$

By the correspondence (25), the values in the segment $\widehat{\mathcal{X}}[2k_aN : 2k_bN - 1]$ are exactly the values one finds in $\mathcal{X}[k_aN : k_bN - 1]$ together with their ϵ-shifts; explicitly, we have

$$\{\widehat{X}_t : t \in [2k_aN : 2k_bN-1]\} = \{X_t : t \in [k_aN : k_bN-1]\} \cup \{X_t(\epsilon) : t \in [k_aN : k_bN-1]\}. \tag{28}$$

More generally, for any $J' \subseteq \mathbb{Z}$, the identification (25) provides us with a $J \subseteq \mathbb{Z}$ such that

$$\{\widehat{X}_t : t \in J\} = \{X_t : t \in J'\} \cup \{X_t(\epsilon) : t \in J'\}, \tag{29}$$

and, if one has $J' \subseteq [k_aN : k_bN - 1]$ for some $k_a < k_b$, then $J \subseteq [2k_aN : 2k_bN - 1]$.

The identifications (28) and (29) are the keys to the proof of Proposition 9. Now we just need to check that for any segment $[a : b]$ of indices of $\widehat{\mathcal{X}}$, we can find a subset $J \subseteq [a : b]$ that meets all of the requirements of Definition 7. The construction of J takes three steps.

First, we note that for k_a and k_b defined by (26) we have the inclusion

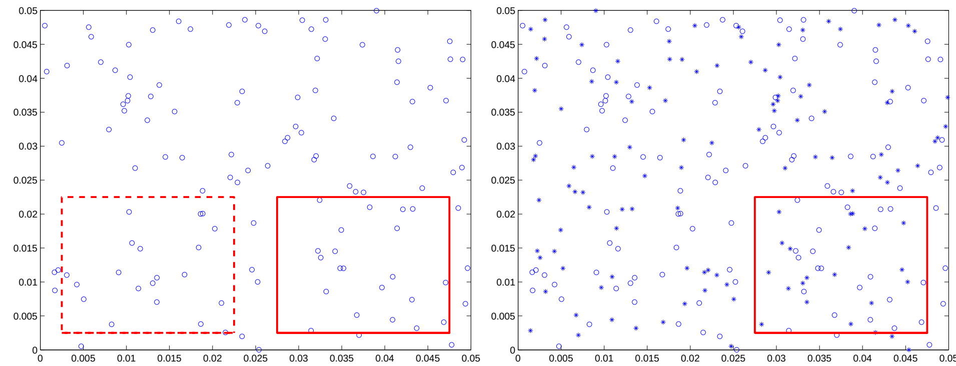
$$\{X_t : t \in [k_aN : k_bN - 1]\} \cup \{X_t(\epsilon) : t \in [k_aN : k_bN - 1]\} \subseteq \{\widehat{X}_t : t \in [a : b]\}.$$

Second, by (α, M)-local uniformity of $\mathcal{X}$, we know there is a set of indices J' for which we have

$$J' \subseteq [k_aN : k_bN - 1] \quad \text{with} \quad 0 \leq |\,[k_aN : k_bN - 1]\,| - |J'| \leq M; \tag{30}$$

moreover, $\mathcal{X}[J']$ satisfies the variance and uniformity conditions of Definition 7.

Figure 1. Preservation of Local Uniformity



The left plot shows a sub-square sample from a process $\mathcal{X}$ that is locally uniform on squares with side length $\alpha < .05$, and the right plot shows a sub-square sample from a process $\widehat{\mathcal{X}}$ obtained from $\mathcal{X}$ after an $H(0.02, 2880)$-transformation. The process $\widehat{\mathcal{X}}$ is locally uniform on squares with side length $\gamma < .02$. The two plots also give a graphical representation of the identity (31): the points that fall in the square in the right plot are the same as the points that fall in the two disjoint squares in the left plot.

Third, and finally, by the identification (29), the set J' determines a set J for which we have

$$|J| = 2|J'| \qquad \text{and} \qquad 0 \le |\,[a:b]\,| - |J| \le 2M + 4N,$$

where the cardinality bound follows from (27) and (30). This completes the choice of our candidate set $J \subseteq [a:b]$. Since J satisfies the cardinality condition required by Proposition 9, it only remains to confirm that for each $A \in \mathcal{Q}(\gamma)$ the point set $A \cap \{\widehat{X}_t : t \in J\}$ also meets our variance and uniformity conditions.

We first let A^ϵ be the set A shifted to the left by ϵ,

$$A^\epsilon = A - \epsilon = \{(x,y) \in \mathcal{T} : (x+\epsilon, y) \in A\},$$

and then we note that since $\epsilon < \alpha$ and $\gamma \le \min\{\epsilon, \alpha - \epsilon\}$ we have

$$A \cap A^\epsilon = \emptyset \qquad \text{and} \qquad A \cup A^\epsilon \subseteq \mathcal{Q}(\alpha).$$

The identification (29) then tells us that $\{\widehat{X}_t : t \in J\}$ contains only *paired* points, i.e. points that appear both with their original and their shifted value, so we have

$$A \cap \{\widehat{X}_t : t \in J\} = (A \cup A^\epsilon) \cap \{X_t : t \in J'\}. \tag{31}$$

This intuitive but hard-won identity is illustrated in Figure 1. It reduces the proof of Proposition 9 to two easy checks.

Check 1: Means Bound Variances. From (31) and $A \cap A^\epsilon = \emptyset$, we have

$$N(A, \widehat{\mathcal{X}}[J]) = N(A \cup A^\epsilon, \mathcal{X}[J']), \tag{32}$$

so, the variance of $N(A, \widehat{\mathcal{X}}[J])$ satisfies the equality

$$\mathrm{Var}(N(A, \widehat{\mathcal{X}}[J])) = \mathrm{Var}(N(A \cup A^\epsilon, \mathcal{X}[J'])).$$

From the variance condition (15) for $\mathcal{X}$ and the identity (31), we then obtain

$$\mathrm{Var}(N(A, \widehat{\mathcal{X}}[J])) \le \mathbb{E}[N(A \cup A^\epsilon, \mathcal{X}[J'])] = \mathbb{E}[N(A, \widehat{\mathcal{X}}[J])],$$

which is the variance condition (15) for $\widehat{\mathcal{X}}$.

Check 2: Conditional Uniformity. By (31) and the conditional-uniformity condition (16) for $\mathcal{X}$ we have

$$\{A \cap \widehat{\mathcal{X}}[J]\} \overset{\text{dpp}}{=} \{U_s(A \cup A^\epsilon) : 1 \leq s \leq N(A \cup A^\epsilon, \mathcal{X}[J'])\}, \tag{33}$$

where the random variables $U_s(A \cup A^\epsilon)$, $1 \leq s < \infty$, are independent and uniformly distributed on $A \cup A^\epsilon$. For each Borel set $B \subseteq \mathcal{T}$, we now introduce the shorthand $N_B \equiv N(B, \mathcal{X}[J'])$, and we note that

$$\begin{aligned} &\{U_s(A \cup A^\epsilon) : 1 \leq s \leq N_{A \cup A^\epsilon}\} \\ &\quad = \big\{A \cap \{U_s(A \cup A^\epsilon) : 1 \leq s \leq N_{A \cup A^\epsilon}\}\big\} \cup \big\{A^\epsilon \cap \{U_s(A \cup A^\epsilon) : 1 \leq s \leq N_{A \cup A^\epsilon}\}\big\} \\ &\quad \overset{\text{dpp}}{=} \{U_s(A) : 1 \leq s \leq N_A\} \cup \big\{U_s(A^\epsilon) : 1 \leq s \leq N_{A^\epsilon}\big\}, \end{aligned} \tag{34}$$

where the $U_s(A)$'s and the $U_s(A^\epsilon)$'s are independent from each other and uniformly distributed on A and A^ϵ respectively. We now note that the independent random variables $\epsilon + U_s(A^\epsilon)$, $1 \leq s \leq N_A$, are uniformly distributed on A, so we have

$$\big\{U_s(A) : 1 \leq s \leq N_A\big\} \cup \big\{U_s(A^\epsilon) : 1 \leq s \leq N_{A^\epsilon}\big\} \overset{\text{dpp}}{=} \big\{U_s(A) : 1 \leq s \leq N_A + N_{A^\epsilon}\big\}. \tag{35}$$

The equality (31) and the fact that $A \cap A^\epsilon = \emptyset$, then give us that

$$N(A, \widehat{\mathcal{X}}[J]) = N(A \cup A^\epsilon, \mathcal{X}[J']) = N(A, \mathcal{X}[J']) + N(A^\epsilon, \mathcal{X}[J']),$$

so, if we recall (33) and use (34) and (35), we obtain

$$\begin{aligned} \{A \cap \widehat{\mathcal{X}}[J]\} &\overset{\text{dpp}}{=} \{U_s(A) : 1 \leq s \leq N(A, \mathcal{X}[J']) + N(A^\epsilon, \mathcal{X}[J']\} \\ &= \{U_s(A) : 1 \leq s \leq N(A, \widehat{\mathcal{X}}[J])\}, \end{aligned}$$

which confirms the conditional-uniformity condition (16) for $\widehat{\mathcal{X}}$ and completes the proof of Proposition 9.

Local uniformity of the hat-process $\widehat{\mathcal{X}}$ transfers very easily to local uniformity of the stationary process $\widetilde{\mathcal{X}}$, and this is just what one needs to pace the parameter choices of our final construction.

Lemma 10 (Local Uniformity and $T(\epsilon, N)$-Transformations)**.** *Consider a $\mathcal{T}$-valued process $\mathcal{X}$ with uniform marginal distributions. If $\mathcal{X}$ is (α, M)-locally uniform for some $0 < \alpha < 1$ and $M < \infty$, and if for some $0 < \epsilon < \alpha$ and $N < \infty$ we have*

$$\mathcal{X} \xrightarrow{T(\epsilon,N)} \widetilde{\mathcal{X}},$$

then $\widetilde{\mathcal{X}}$ is $(\gamma, 2M + 4N)$-locally uniform for all $\gamma \leq \min\{\epsilon, \alpha - \epsilon\}$.

Proof. For any segment $[a : b]$ of $\widetilde{\mathcal{X}}$, we recall that $\widetilde{\mathcal{X}}[a : b] = \widehat{\mathcal{X}}[I + a : I + b]$, where I is an independent random variable uniform on $\{0, \ldots, 2N - 1\}$. Proposition 9 then tells us that $\widehat{\mathcal{X}}[i + a : i + b]$ is $(\gamma, 2M + 4N)$-locally uniform for each value of i. Moreover, for each $0 \leq i \leq 2N - 1$, one can choose the sets J_i given by Proposition 9 to have all the same cardinality. The $(\gamma, 2M + 4N)$-locally uniformity of $\widetilde{\mathcal{X}}$ then follows immediately after conditioning on the value of I. $\square$

5. Iterated $T(\epsilon, N)$-Transformations and a Limit Process

We now consider the construction of a process $\mathcal{X}^* = \{X_t^* : t \in \mathbb{Z}\}$ as limit of iterated $T(\epsilon, N)$-transformations. We first fix an increasing sequence of integers $1 \leq N_1 < N_2 < \cdots$ and a decreasing sequence of reals $1 > \epsilon_1 > \epsilon_2 > \cdots > 0$. Next, we let $\mathcal{X}^{(0)} = \{X_t^{(0)}, t \in \mathbb{Z}\}$ be the doubly infinite sequence of independent random variables with the uniform distribution on $\mathcal{T}$, and we consider the infinite sequence of doubly infinite stationary processes $\{\mathcal{X}^{(j)} : 0 \leq j < \infty\}$ that are obtained by successive applications of appropriate $T(\epsilon, N)$-transformations:

$$\mathcal{X}^{(0)} \xrightarrow{T(\epsilon_1, N_1)} \mathcal{X}^{(1)} \xrightarrow{T(\epsilon_2, N_2)} \mathcal{X}^{(2)} \xrightarrow{T(\epsilon_3, N_3)} \mathcal{X}^{(3)} \xrightarrow{T(\epsilon_4, N_4)} \cdots . \tag{36}$$

We denote the doubly infinite torus by $\mathcal{T}^{[-\infty:\infty]}$, and we let $\mathcal{B}(\mathcal{T}^{[-\infty:\infty]})$ be the set of all Borel subsets of $\mathcal{T}^{[-\infty:\infty]}$. We also let $\mathcal{M}$ be the set of all Borel measures on the doubly infinite torus, and we note that $\mathcal{M}$ becomes a complete metric space if we define the distance $\rho(\mu, \mu')$ between the Borel measures μ and μ' by setting

$$\rho(\mu, \mu') = \sum_{m=1}^{\infty} 2^{-m} \sup\{\, |\mu(A) - \mu'(A)| : \, A \in \mathcal{B}(\mathcal{T}^{[-m:m]})\, \}. \tag{37}$$

To show that the sequence $\{\mathcal{X}^{(j)} : 0 \leq j < \infty\}$ converges in distribution to a process $\mathcal{X}^*$, it thus suffices to show that the measures defined on $\mathcal{B}(\mathcal{T}^{[-\infty:\infty]})$ by

$$\mu_j(A) = P(\mathcal{X}^{(j)} \in A), \tag{38}$$

are a Cauchy sequence under the metric ρ. The Cauchy criterion can be checked under a very mild conditions on the sequence $\{N_j : j = 1, 2, \ldots\}$.

Lemma 11 (A Condition for Convergence)**.** *If the processes* $\{\mathcal{X}^{(j)}, 0 \leq j < \infty\}$ *are defined by the iterative* T*-transformations* (36) *and if*

$$\sum_{j=1}^{\infty} \frac{1}{N_j} < \infty, \tag{39}$$

then the sequence of processes $\{\mathcal{X}^{(j)}, 0 \leq j < \infty\}$ *converges in distribution to a stationary, translation invariant process*

$$\mathcal{X}^* = \{\ldots, X_{-1}^*, X_0^*, X_1^*, \ldots\}$$

such that X_t^* *is uniformly distributed on* $\mathcal{T}$ *for each* $t \in \mathbb{Z}$.

Proof. By the closeness inequality (11) and the definition (38) of μ_j, we have for all $m = 1, 2, ...$ that

$$\sup\{|\mu_j(A) - \mu_{j+1}(A)| : \, A \subseteq \mathcal{B}(\mathcal{T}^{[-m:m]}\} \leq \frac{2m}{N_{j+1}}. \tag{40}$$

The definition (37) of the metric ρ and a simple summation then give us

$$\rho(\mu_j, \mu_{j+1}) \leq \frac{4}{N_{j+1}},$$

so, by the completeness of the metric space $(\mathcal{M}, \rho)$, the sequence of processes $\{\mathcal{X}^{(j)}, 0 \leq j < \infty\}$ converges in distribution to a process $\mathcal{X}^*$. By Lemmas 3, 4, and 5, we know that each of the processes $\mathcal{X}^{(j)}$ is stationary, translation invariant, and has uniform marginal distributions. The process $\mathcal{X}^*$ then inherits these properties by convergence in distribution. $\square$

6. Path Lengths for the Limit Process

The next lemma expresses a kind of Lipschitz property for the TSP functional. Specifically, it bounds the absolute difference in the expected value of $L(Z)$ and $L(\widetilde{Z})$, where Z and $\widetilde{Z}$ are arbitrary n-dimensional random vectors with values in $\mathcal{T}^n$. The lemma is stated and proved for general Z and $\widetilde{Z}$, but our typical choice will be $Z = \mathcal{X}[0 : n-1]$ and $\widetilde{Z} = \widetilde{\mathcal{X}}[0 : n-1]$. To state the lemma, we recall that if $\mathcal{B}(\mathcal{T}^n)$ is the set of all Borel subsets of $\mathcal{T}^n$, then the *total variation distance* between Z and $\widetilde{Z}$ is given by

$$d_{\mathrm{TV}}(Z, \widetilde{Z}) = \sup\{|\mathbb{P}(Z \in A) - \mathbb{P}(\widetilde{Z} \in A)\,| : A \in \mathcal{B}(\mathcal{T}^n)\}.$$

We also recall that the function $(z_1, z_2, \ldots, z_n) \mapsto L(z_1, z_2, \ldots, z_n)/\sqrt{n}$ is uniformly bounded; in fact, by early work of Few (1955), this ratio is bounded by 3.

Lemma 12. *For all random vectors Z and $\widetilde{Z}$ with values in $\mathcal{T}^n$ we have*

$$|\mathbb{E}[L(Z)] - \mathbb{E}[L(\widetilde{Z})]| \le 3n^{1/2} d_{\mathrm{TV}}(Z, \widetilde{Z}). \tag{41}$$

Proof. By the maximal coupling theorem (Lindvall, 2002, Theorem 5.2) there exist a probability space and a random pair $(Z', \widetilde{Z}')$ such that $Z' \stackrel{d}{=} Z$, $\widetilde{Z}' \stackrel{d}{=} \widetilde{Z}$ and

$$\mathbb{P}(Z' \ne \widetilde{Z}') = d_{\mathrm{TV}}(Z, \widetilde{Z}).$$

Now, if we set $M_n = \max\{L(z_1, z_2, \ldots, z_n) : z_t \in \mathcal{T}, 1 \le t \le n\}$, then we have

$$\begin{aligned} |\mathbb{E}[L(Z)] - \mathbb{E}[L(\widetilde{Z})]| &= |\mathbb{E}[L(Z')] - \mathbb{E}[L(\widetilde{Z}')]| \\ &\le \mathbb{P}(Z' \ne \widetilde{Z}') M_n \\ &\le 3n^{1/2} d_{\mathrm{TV}}(Z, \widetilde{Z}), \end{aligned}$$

where in the last line we used the classic bound $M_n \le 3n^{1/2}$ from Few (1955). □

The immediate benefit of Lemma 12 is that it gives us a way to estimate the cost of minimal paths through the points of $\mathcal{X}^*[0 : n-1]$ for all $n \ge 1$.

Lemma 13 (TSP Differences in the Limit)**.** *For all $0 \le j < \infty$ and all $n \ge 1$ we have*

$$|\mathbb{E}[L(\mathcal{X}^{(j)}[0 : n-1])] - \mathbb{E}[L(\mathcal{X}^*[0 : n-1])]| \le 3\, n^{3/2} \sum_{k=j}^{\infty} \frac{1}{N_{k+1}} \tag{42}$$

Proof. Using the shorthand $\mathcal{X}_n^{(j)} \equiv \mathcal{X}^{(j)}[0 : n-1]$, one has by the triangle inequality that

$$|\mathbb{E}[L(\mathcal{X}_n^{(j)})] - \mathbb{E}[L(\mathcal{X}_n^*)]| \le \sum_{k=j}^{\infty} |\mathbb{E}[L(\mathcal{X}_n^{(k+1)})] - \mathbb{E}[L(\mathcal{X}_n^{(k)})]|. \tag{43}$$

Lemma 12 then tells us that

$$|\mathbb{E}[L(\mathcal{X}_n^{(k+1)})] - \mathbb{E}[L(\mathcal{X}_n^{(k)})]| \le 3\, n^{1/2} d_{\mathrm{TV}}(\mathcal{X}_n^{(k+1)}, \mathcal{X}_n^{(k)}), \tag{44}$$

and Lemma 6 implies that

$$d_{\mathrm{TV}}(\mathcal{X}_n^{(k+1)}, \mathcal{X}_n^{(k)}) \le \frac{n}{N_{k+1}}, \tag{45}$$

so using (44) and (45) in the sum (43) completes the proof. □

7. Parameter Choices

To pass from the general iterative construction (36) to the process required by Theorem 2, we need to make parameter choices that go beyond those required by Lemma 11 on sufficient condition for convergence.

First we fix η_j, $1 \le j < \infty$, to be any sequence of values in $(0,1)$ that decrease monotonically to zero as $j \to \infty$; these values just serve to provide us with a measure of smallness of scale. We then inductively define the values N_j and ϵ_j through which we finally define $\mathcal{X}^*$ by the sequence (36) of transformations $T(\epsilon_j, N_j)$, $j = 1, 2, \ldots$. To begin the construction, we can take any $\epsilon_1 \in (0,1)$ and any integer $N_1 \ge 2$. Subsequent values are determined by two rules:

Rule i. We choose an integer N_j such that $N_j > j^2 N_{j-1}$ and such that

$$|\mathbb{E}[L(\mathcal{X}^{(j-1)}[0:n-1])] - \beta\sqrt{n}| \le \eta_j \sqrt{n} \qquad \text{for all } n \ge \lfloor j^{-1} N_j \rfloor. \tag{46}$$

Rule ii. We choose an $\epsilon_j \in (0, \epsilon_{j-1})$ such that

$$\epsilon_j j^{1/2} N_j^{1/2} \le \eta_j. \tag{47}$$

The existence of the N_j required by the first rule follows from Lemma 10 which tells us that each of the processes $\mathcal{X}^{(j)}$, $1 \le j < \infty$, is locally uniform and from Lemma 8 which tells us that locally uniform processes satisfy the BHH condition (17).

8. Estimation of Expected Path Lengths: Proof of Theorem 2

At the j'th step, our construction (36) takes the form

$$\mathcal{X}^{(j-1)} \xrightarrow{H(\epsilon_j, N_j)} \widehat{\mathcal{X}}^{(j-1)} \quad \text{and} \quad \mathcal{X}^{(j-1)} \xrightarrow{T(\epsilon_j, N_j)} \widetilde{\mathcal{X}}^{(j-1)} \stackrel{\text{def}}{=} \mathcal{X}^{(j)},$$

so, from the bound on the randomization index I and the monotonicity of the TSP functional L, we obtain

$$\begin{aligned} L(\mathcal{X}^{(j)}[0:2jN_j-1]) &\le L(\widehat{\mathcal{X}}^{(j-1)}[0:2(j+1)N_j-1]) \\ &\le L(\mathcal{X}^{(j-1)}[0:(j+1)N_j-1]) + 2\epsilon_j (j+1) N_j. \end{aligned} \tag{48}$$

Here, in the first inequality, we used the fact that the set

$$\mathcal{S} = \{\, \widehat{X}_t^{(j-1)} : 0 \le t \le 2(j+1)N_j - 1 \,\}$$

is a superset of $\{\widehat{X}_t^{(j-1)} : I \le t \le I + 2jN_j - 1\} \equiv \{X_t^{(j)} : 0 \le t \le 2jN_j - 1\}$ for all values of the randomization index I, and in the second inequality, we estimate the tour that first visits each point of $\{X_t^{(j-1)} : 0 \le t \le (j+1)N_j - 1\}$ in an optimal order and then builds the suboptimal path through the points of $\mathcal{S}$ by adding a loop from $X_t^{(j-1)}$ to its twin location $X_t^{(j-1)}(\epsilon_j)$ for each $t \in [0:(j+1)N_j - 1]$.

When we take expectations in (48) and use the two rules (46) and (47), we obtain

$$\begin{aligned} (2jN_j)^{-1/2}\mathbb{E}[L(\mathcal{X}^{(j)}[0:2jN_j-1])] &\le (\beta+\eta_j)(1+j^{-1/2})2^{-1/2} + 2\epsilon_j (j^{1/2}+j^{-1/2}) N_j^{1/2} \\ &\le (\beta+\eta_j)(1+j^{-1/2})2^{-1/2} + 4\eta_j. \end{aligned}$$

From the bound (42) we then have

$$(2jN_j)^{-1/2}\mathbb{E}[L(\mathcal{X}^*[0:2jN_j-1])] \leq (\beta+\eta_j)(1+j^{-1/2})2^{-1/2}+4\eta_j+6jN_j\sum_{k=j}^{\infty}\frac{1}{N_{k+1}}$$
$$\leq (\beta+\eta_j)(1+j^{-1/2})2^{-1/2}+4\eta_j+12/j,$$

where, in the second inequality, we estimate the sum using $j^2N_j < N_{j+1}$, which follows from our first parameter formation rule. This last bound is more than one needs to complete the proof of the first inequality (4) required in Theorem 2.

The second inequality (4) of Theorem 2 is easier. If we take $n = \lfloor j^{-1}N_j \rfloor$ in the equation (46) defining N_j we have

$$\beta - \eta_j \leq \lfloor j^{-1}N_j \rfloor^{-1/2}\mathbb{E}[L(\mathcal{X}^{(j-1)}[0:\lfloor j^{-1}N_j\rfloor - 1])].$$

If we again us the bound (42) and estimate the infinite sum as before, then we have

$$\beta - \eta_j - 12/j \leq \lfloor j^{-1}N_j \rfloor^{-1/2}\mathbb{E}[L(\mathcal{X}^*[0:\lfloor j^{-1}N_j\rfloor - 1])],$$

and this suffices to complete the proof of second inequality (4) of Theorem 2.

9. Theorem 2 Implies Theorem 1

We now show that Theorem 1 is an easy corollary of Theorem 2. We first observe that the process $\mathcal{X}^*$ given by Theorem 2 determines a canonical measure μ on the set $\Omega = \mathcal{T}^{[-\infty:\infty]}$ of doubly infinite strings $\omega = (\ldots x_{-1}, x_0, x_1, \ldots)$ with $x_t \in \mathcal{T}$ for each $t \in \mathbb{Z}$; one simply lets $\mu(A) = P(\mathcal{X}^* \in A)$ for each Borel set $A \subseteq \Omega$. Furthermore, if S denotes the shift transformation on Ω, so that $S(\omega)_t = \omega_{t+1}$, then the stationarity of $\mathcal{X}^*$ corresponds to the invariance of μ under the shift transformation S; i.e. $\mu(S^{-1}(A)) = \mu(A)$ for all Borel sets $A \subseteq \Omega$.

Now, for $\omega = (\ldots x_{-1}, x_0, x_1, \ldots) \in \Omega$ we let $\omega[1:n] = (x_1, x_2, \ldots, x_n)$, and, for $c_1 < c_2$, we define the set $\Delta[c_1, c_2]$ by

$$\Delta[c_1,c_2] = \big\{\omega \in \Omega : \liminf_{n\to\infty} L(\omega[1:n])/\sqrt{n} \leq c_1 \text{ and}$$
$$c_2 \leq \limsup_{n\to\infty} L(\omega[1:n])/\sqrt{n}\,\big\}. \tag{49}$$

If $\mu(\Delta[c_1,c_2]) = 0$ for all $c_1 < c_2$, then $L(\omega[1:n])/\sqrt{n}$ converges with μ-probability one, and, since this ratio is bounded, the dominated convergence theorem implies the convergence of the expectations $\mathbb{E}[L(\omega[1:n])]/\sqrt{n}$. Since the convergence of these expectations is impossible by Theorem 2, we conclude that there is some pair $c_1 < c_2$ for which we have $0 < \mu(\Delta[c_1,c_2])$.

Now, if $\mathcal{M}_e$ denotes the set of all ergodic shift-invariant measures on the doubly-infinite torus Ω, then the ergodic decomposition theorem (cf. Dynkin, 1978) tells us there is a probability measure D_μ on $\mathcal{M}_e$ such that

$$\mu(A) = \int_{\mathcal{M}_e} \nu(A)\, D_\mu(d\nu) \qquad \text{for every Borel set } A \subseteq \Omega. \tag{50}$$

In particular, we have

$$0 < \mu(\Delta[c_1,c_2]) = \int_{\mathcal{M}_e} \nu(\Delta[c_1,c_2])\, D_\mu(d\nu),$$

so there is an $\nu \in \mathcal{M}_e$ such that $0 < \nu(\Delta[c_1,c_2])$.

Since the sets $\Delta[c_1,c_2]$ and $S^{-1}(\Delta[c_1,c_2])$ are identical, $\Delta[c_1,c_2]$ is an invariant set for the measure ν and, since ν is ergodic, we obtain that $\nu(\Delta[c_1,c_2]) = 1$. Finally, we take $\mathcal{X}$ to be the stationary process determined by the shift transformation S and the ergodic measure ν. By construction, the process $\mathcal{X}$ is stationary and ergodic with the uniform marginal distribution, so, by the definition (49) of $\Delta[c_1, c_2]$, we see that $\mathcal{X}$ has all of the features required by Theorem 1.

10. Extensions, Refinements, and Problems

There are easily proved analogs of Theorems 1 and 2 for many functionals of combinatorial optimization for which one has the analog of the Beardwood-Halton-Hammersley theorem. In particular, one can show that the analogs of Theorems 1 and 2 hold for the minimal spanning tree (MST) problem studied in Steele (1988) and the minimal matching problem studied in Rhee (1993). The construction of the processes in Theorems 1 and 2 needs no alteration, but it is necessary to check that one has an analog of Lemma 8; a check that is quite easy in these cases. Finally, one needs to make minor modifications to the arguments of Section 8, but these should also follow easily.

Still, there are interesting functionals for which it is not clear how to adapt the proof of Theorems 1 and 2. One engaging example is the sum of the edge lengths of the Voronoi tessellation. In this case, the analog of the BHH theorem was developed by Miles (1970) for Poisson sample sizes, and later by McGivney and Yukich (1999) for fixed sample sizes and with complete convergence. A second, interesting but much different example, is the length of the path that one obtains by running the Karp-Held algorithm for the TSP. The usual optimality properties are absent for this functional, but, nevertheless, Goemans and Bertsimas (1991) obtained the analog of the BHH theorem.

These two functionals are "less local" than the TSP, MST, or minimal matching functionals; in particular, they are not amenable to simple suboptimal patching heuristics like those we used in Section 8. Nevertheless, they are sufficiently local to allow for analogs of the BHH theorem, so it seems quite probable that the natural analogs of Theorems 1 and 2 would hold as well.

We should also note that there is a log-scale version of the BHH theorem that applies even for some non-random sequences of points in $[0,1]^2$. In particular, it applies to point sequences $\{x_1, x_2, \ldots\}$ in $[0,1]^2$ for which one has a strong control of the "rectangle discrepancy" that is given by

$$D_n = D(x_1, x_2, \ldots, x_n) = \sup_{Q \in \mathcal{Q}} \left| \frac{1}{n} \sum_{t=1}^{n} \mathbb{1}_Q(x_t) - \lambda(Q) \right|,$$

where $\mathcal{Q}$ is the set of all axis-aligned rectangles $Q \subseteq [0,1]^2$. The inequalities of Steele (1980), recently refined by Steinerberger (2010), suffice to show that if for all $r \in (0,1)$ one has $nD_n = o(n^r)$ as $n \to \infty$, then one also has the pointwise limit

$$\lim_{n\to\infty} \frac{\log L(x_1, x_2, \ldots, x_n)}{\log n} = \frac{1}{2}. \tag{51}$$

The leading example of such a sequence is $x_n = (n\phi \mod 1,\, n\psi \mod 1)$ where ϕ and ψ are algebraic irrationals that are linearly independent over the rationals. A deep theorem of Schmidt (1964) tells us that the discrepancy of this sequence satisfies the remarkable estimate $nD_n = O((\log n)^{3+r})$ for all $r > 0$. This is certainly more than one needs for the discrepancy criterion to justify the limit (51).

While it may seem inevitable that the processes defined by Theorems 1 and 2 would satisfy the limit (51), this cannot be proved just by consideration of the discrepancy. Even a sequence of independent uniformly distributed points fail to satisfy $nD_n = o(n^{1/2})$, and the sequences of Theorems 1 and 2 cannot be expected to be more uniformly distributed than a sequence of independent uniformly distributed points.

There are two further points worth noting. First, at the cost of using more complicated versions of the $H(\epsilon, N)$ and $T(\epsilon, N)$ transformations, one can replace the infimum bound $\beta/\sqrt{2}$ of Theorem 2 with a smaller constant. The method of Section 9 shows that any infimum bound less that β suffices to prove Theorem 1, so we did not pursue the issue of a minimal infimum bound.

Finally, it is feasible that the process $\{X_t^* : t \in \mathbb{Z}\}$ determined by Theorems 2 is itself ergodic, or even mixing. If this could be established, then one would not need the ergodic decomposition argument of Section 9. Unfortunately, it does not seem easy to prove that $\{X_t^* : t \in \mathbb{Z}\}$ is ergodic, even though this is somewhat intuitive.

References

Applegate, D. L., Bixby, R. E., Chvátal, V. and Cook, W. J. (2006), *The traveling salesman problem*, Princeton Series in Applied Mathematics, Princeton University Press, Princeton, NJ. A computational study.

Avram, F. and Bertsimas, D. (1992), 'The minimum spanning tree constant in geometrical probability and under the independent model: a unified approach', *Ann. Appl. Probab.* **2**(1), 113–130.

Beardwood, J., Halton, J. H. and Hammersley, J. M. (1959), 'The shortest path through many points', *Proc. Cambridge Philos. Soc.* **55**, 299–327.

Dynkin, E. B. (1978), 'Sufficient statistics and extreme points', *Ann. Probab.* **6**(5), 705–730.

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

Finch, S. R. (2003), *Mathematical constants*, Vol. 94 of *Encyclopedia of Mathematics and its Applications*, Cambridge University Press, Cambridge.

Goemans, M. X. and Bertsimas, D. J. (1991), 'Probabilistic analysis of the Held and Karp lower bound for the Euclidean traveling salesman problem', *Math. Oper. Res.* **16**(1), 72–89.

Jaillet, P. (1993), 'Cube versus torus models and the Euclidean minimum spanning tree constant', *Ann. Appl. Probab.* **3**(2), 582–592.

Lindvall, T. (2002), *Lectures on the coupling method*, Dover Publications Inc., Mineola, NY. Corrected reprint of the 1992 original.

McGivney, K. and Yukich, J. E. (1999), 'Asymptotics for Voronoi tessellations on random samples', *Stochastic Process. Appl.* **83**(2), 273–288.

Miles, R. E. (1970), 'On the homogeneous planar Poisson point process', *Math. Biosci.* **6**, 85–127.

Redmond, C. and Yukich, J. E. (1994), 'Limit theorems and rates of convergence for Euclidean functionals', *Ann. Appl. Probab.* **4**(4), 1057–1073.

Rhee, W. T. (1993), 'A matching problem and subadditive Euclidean functionals', *Ann. Appl. Probab.* **3**(3), 794–801.

Schmidt, W. M. (1964), 'Metrical theorems on fractional parts of sequences', *Trans. Amer. Math. Soc.* **110**, 493–518.

Steele, J. M. (1980), 'Shortest paths through pseudorandom points in the d-cube', *Proc. Amer. Math. Soc.* **80**(1), 130–134.

Steele, J. M. (1988), 'Growth rates of Euclidean minimal spanning trees with power weighted edges', *Ann. Probab.* **16**(4), 1767–1787.

Steinerberger, S. (2010), 'A new lower bound for the geometric traveling salesman problem in terms of discrepancy', *Oper. Res. Lett.* **38**(4), 318–319.

Yukich, J. E. (1998), *Probability theory of classical Euclidean optimization problems*, Vol. 1675 of *Lecture Notes in Mathematics*, Springer-Verlag, Berlin.