Percolating paths through random points

David Aldous, Maxim Krikun

Percolating paths through random points

David J. Aldous[*]
Department of Statistics
367 Evans Hall # 3860
U.C. Berkeley CA 94720
aldous@stat.berkeley.edu

Maxim Krikun
Department of Statistics
367 Evans Hall # 3860
U.C. Berkeley CA 94720
krikun@stat.berkeley.edu

April 6, 2007

Abstract

We prove consistency of four different approaches to formalizing the idea of minimum average edge-length in a path linking some infinite subset of points of a Poisson process. The approaches are (i) shortest path from origin through some m distinct points; (ii) shortest average edge-length in paths across the diagonal of a large cube; (iii) shortest path through some specified proportion δ of points in a large cube; (iv) translation-invariant measures on paths in $\mathbb{R}^d$ which contain a proportion δ of the Poisson points. We develop basic properties of a normalized average length function $c(\delta)$ and pose challenging open problems.

Key words and phrases. Combinatorial optimization, continuum percolation, first passage percolation, Euclidean traveling salesman problem, Poisson process.

AMS 2000 subject classifications. 60K35, 60C05, 60G55.

1 Introduction

Fix dimension $d \geq 2$ and let (ξ_i) be a Poisson point process of rate 1 per unit volume in $\mathbb{R}^d$. "Volume" means d-dimensional Lebesgue measure.

To start with an analogy, one can informally describe the critical value for continuum percolation as the smallest c such that there exists some infinite sequence $\xi_{j_1}, \xi_{j_2}, \xi_{j_3}, \ldots$ of distinct points such that $\max_{i \geq 1} |\xi_{j_{i+1}} - \xi_{j_i}| \leq 2c$. Here $|\cdot|$ denotes Euclidean distance. What we study in this paper can analogously be described informally as the smallest c such that there exists some path through an infinite sequence $\xi_{j_1}, \xi_{j_2}, \xi_{j_3}, \ldots$ of distinct points whose *average edge-length* $\lim_{n \to \infty} n^{-1} \sum_{i=1}^{n} |\xi_{j_{i+1}} - \xi_{j_i}| \leq c$. One could formalize this directly by e.g. replacing *lim* by *lim sup*, requiring the property to hold almost surely and then taking the *inf* of such c. But such a definition seems neither elegant nor convenient. Our purpose in this paper is to study four indirect approaches to this formalization question and show that they lead to the same constant, which we call $c(0+)$. Along the way we introduce a function $c(\delta)$ which plays a role analogous to the percolation function. Our results are collected in Theorem 1.

Conceptually, this topic seems intermediate between *first passage percolation* and the random *traveling salesman problem* (TSP). Regarding the former, our $c(0+)$ differs (informally speaking) from a continuum analog of the time constant in first passage percolation [6, 8] because we use "distance along a path" in place of "end-to-end Euclidean distance" (nonetheless we use "continuum first passage percolation" as a descriptor of one of our approaches below). Regarding the latter, take n random points in a d-dimensional cube of volume n. Let $L_n(1)$ be the length of the shortest cycle through all n points, i.e. the length of the solution of the TSP. Almost 50 years ago, Beardwood-Halton-Hammersley [2] proved there exists a constant $c(1)$ such that $EL_n(1) \sim c(1)n$ as $n \to \infty$. Subsequent work on related problems is described in the monographs by Steele [13] and Yukich [14]. Another of our approaches modifies the TSP by considering cycles through some specified proportion δ of the Poisson points. Additional motivation for the current paper comes from work on such problems in the "mean-field" setting, described in section 2.2.

Challenging problems for future research are listed in section 2.1. One of our techniques – subadditive analysis of optimal cost/reward ratios – seems potentially applicable in other contexts, as briefly discussed in section 4.1.

2 The equivalence theorem

Approach 1 *Continuum first passage percolation from the origin.* For each $m \geq 1$ define a random variable T_m as the minimum, over choices $\{\xi_{j_1}, \ldots, \xi_{j_m}\}$ of m distinct points of the Poisson process on $\mathbb{R}^d$, of

$$|\xi_{j_1}| + \sum_{i=2}^{m} |\xi_{j_i} - \xi_{j_{i-1}}|.$$

Comment. One certainly expects that T_m/m should converge to a constant, but we see no easy argument. In particular we don't see how to apply subadditivity arguments directly to study T_m.

Approach 2 *Continuum first passage percolation across a diagonal.* For $s > 0$ define a random variable W_s as the minimum, over all $m \geq 1$ and all choices $\{\xi_{j_1}, \ldots, \xi_{j_{m-1}}\} \subset [0,s]^d$ of $m-1$ distinct points of the Poisson process, of

$$m^{-1} \sum_{i=1}^{m} |\xi_{j_i} - \xi_{j_{i-1}}| \tag{1}$$

where $\xi_{j_0} = (0, \ldots, 0)$ and $\xi_{j_m} = (s, \ldots, s)$.
Comment. Here we can attempt subadditivity analysis, based on splitting the cube of side $2s$ into 2^d subcubes of side s, though because of the "ratio" form of (1) we are not in the usual format for the subadditive ergodic theorem.

Approach 3 *TSP on sparse subsets of the cube.* Let $\mathbb{C}_n = [0, n^{1/d}]^d$ be the cube of volume n in $\mathbb{R}^d$. Put n random (independent, uniformly distributed) points (ζ_j) into $\mathbb{C}_n$. Fix $0 < \delta \leq 1$. Let $L_n(\delta)$ be the minimum, over all choices of cycles $(\zeta_{j_1}, \zeta_{j_2}, \ldots, \zeta_{j_m}, \zeta_{j_{m+1}} = \zeta_{j_1})$ through $m = \lceil \delta n \rceil$ disjoint choices from the random points, of the cycle length $\sum_{i=1}^{m} |\zeta_{j_{i+1}} - \zeta_{j_i}|$.
Comment. Here subadditivity can be applied in familiar ways. Note that, in contrast to continuum percolation where definitions are in terms of the process on infinite space $\mathbb{R}^d$, the approaches above all envisage taking limits over finite regions. For the record we give a final approach which does work directly on infinite space, as "the $n = \infty$ analog of Approach 3", though we admit it does not seem very useful.

Approach 4 *Translation invariant distributions on infinite paths through Poisson points.* Consider a locally finite set (x_i) of points in $\mathbb{R}^d$, together

3

with a set $\mathcal{E}$ of edges whose endpoints are in (x_i), where the edges form a collection of doubly-infinite paths, each point of (x_i) appearing either once or never in the paths. Write $\mathbf{S}$ for the space of such points-and-paths configurations. The Euclidean translation group acts naturally on $\mathbf{S}$, so one can define a probability distribution μ on $\mathbf{S}$ to be *invariant* if it is invariant under the action of the Euclidean translation group. Let $\mathcal{M}$ be the set of invariant distributions on $\mathbf{S}$ under which the distribution of the points (ξ_i) is the Poisson point process of rate 1. Informally, a $\mu \in \mathcal{M}$ is just a way of collecting some subsets of the Poisson points into paths using a rule which doesn't depend on the location of the origin. For $\mu \in \mathcal{M}$ there is a constant $\delta(\mu) \in [0,1]$ specified informally as "the proportion of points which are in some infinite path" and defined formally via the formula: for every cube $\mathbb{C} \subset \mathbb{R}^d$,

$E_\mu(\text{number of points } \xi_i \in \mathbb{C} \text{ which are in some infinite path}) = \delta(\mu) \, \text{volume}(\mathbb{C})$.

Similarly there is a constant $\ell(\mu)$ interpreted as "mean edge-length over all edges in the paths of μ" and formally via the formula: for every cube $\mathbb{C} \subset \mathbb{R}^d$,

$$E_\mu(\text{length of } \mathcal{E} \cap \mathbb{C}) = \delta(\mu)\ell(\mu) \, \text{volume}(\mathbb{C})$$

(here we regard $\mathcal{E}$ as a subset of $\mathbb{R}^d$). Finally define

$$\bar{c}(\delta) := \inf\{\ell(\mu) : \ \mu \in \mathcal{M}, \ \delta(\mu) = \delta\}.$$

Theorem 1

(a) *For $0 < \delta \leq 1$ there exists a constant $c(\delta)$ such that $\frac{L_n(\delta)}{\delta n} \to c(\delta)$ in L^2 as $n \to \infty$.*

(b) *The function $c(\delta)$ is non-decreasing and continuous on $(0,1]$, the function $\delta c(\delta)$ is convex, and the limit $c(0+) := \lim_{\delta \downarrow 0} c(\delta)$ is strictly positive.*

(c) *$W_s \to c(0+)$ a.s. as $s \to \infty$.*

(d) *$\bar{c}(\delta) = c(\delta)$, $0 < \delta \leq 1$.*

(e) *$m^{-1}T_m \to c(0+)$ in probability.*

From (a) we see $c(1)$ is the constant in the Beardwood-Halton-Hammersley [2] theorem. Monte Carlo simulations [7] give (for $d = 2$) $c(1) \approx 0.7119$ but no close rigorous bounds are known.

4

2.1 Discussion

The function $c(\delta)$, $0 < \delta \leq 1$ seems worthy of study as a analog of the classical percolation function from lattice percolation theory [5]:

$$f(p) := P(\text{origin is in some infinite component})$$

in bond percolation with edge-probability p. In particular one can ask whether there exists a *scaling exponent* $0 < \alpha < \infty$, that is whether

$$c(\delta) - c(0+) \asymp \delta^\alpha \text{ as } \delta \downarrow 0. \tag{2}$$

For the record we state the (probably very hard)

Open Problem 1 *Prove (2) holds for some α. Or give bounds on the possible values of α.*

The next question may be easier. The facts that $c(\delta)$ is nondecreasing and that $\delta c(\delta)$ is convex imply: either
(i) $c(\delta)$ is strictly increasing on $0 < \delta < 1$; or
(ii) $c(\delta)$ is constant on $0 < \delta < \delta_0$, for some $0 < \delta_0 \leq 1$.
But the latter seems implausible.

Open Problem 2 *Prove that $c(\delta)$ is strictly increasing on $0 < \delta < 1$.*

Getting reasonable bounds on the numerical value of $c(0+)$ seems difficult. Standard methods (comparison with branching random walk: section 3.3) give an explicit lower bound (8), which in $d = 2$ is $1/(2\pi e) \approx 0.0585$. But we don't see any simple way to get an interesting upper bound. Even Monte Carlo methods seem difficult to code convincingly; for the record we write

Open Problem 3 *In $d = 2$ study the numerical value $c(0+)$ and the presumed scaling exponent α via Monte Carlo methods.*

The variation in which (in $d = 2$) we restrict to "upward oriented" paths, that is edges $(x_i, y_i) \to (x_{i+1}, y_{i+1})$ are required to have $y_{i+1} > y_i$, is easier to study via simulation; our small-scale simulations suggest the analog of $c(0+)$ in this variation is ≈ 0.62, which would be an *a priori* upper bound for our original $c(0+)$.

Another question concerns variances. Take $d = 2$ here. For the basic traveling salesman problem, that is for $L_n(1)$, it is known that $\mathrm{var}(L_n(1))$ is order n; precisely,

$$\liminf_n n^{-1}\mathrm{var}(L_n(1)) > 0, \quad \limsup_n n^{-1}\mathrm{var}(L_n(1)) < \infty.$$

The upper bound is explicit in Steele [12] and the lower bound follows from a corresponding large deviation lower bound in Rhee [11]. On the other hand, for first-passage percolation it has long been conjectured [4] that variance grows as the $\frac{2}{3}$ power of expectation, though little has been proved rigorously [3].

Open Problem 4 *Prove* $\mathrm{var}(T_m) \asymp m^{2/3}$. *Or just prove* $\mathrm{var}(T_m) = o(m)$.

We will not even venture a conjecture for the asymptotic behavior of $\mathrm{var}(L_n(\delta))$.

As a small rewriting of the definition of $L_n(\delta)$, let $L(n,m)$ be the minimum, over all choices of cycles $(\zeta_{j_1}, \ldots, \zeta_{j_m}, \zeta_{j_{m+1}} = \zeta_{j_1})$ through some chosen m of the random points, of the cycle length $\sum_{i=1}^{m} |\zeta_{j_{i+1}} - \zeta_{j_i}|$. Theorem 1 implies that, if $m_n/n \to 0$ sufficiently slowly, then $L(n, m_n)/m_n \to c(0+)$. However it seems plausible this also holds for smaller values of m_n.

Open Problem 5 *Prove that* $L(n, m_n)/m_n \to c(0+)$ *in probability whenever* $m_n/n \to 0$ *and* $\frac{m_n}{\log^a n} \to \infty$ $\forall a < \infty$.

In other words, for fixed $c < c(0+)$ consider Poisson points in a cube of volume n; is it true that any cycle with average edge-length $\leq c$ can have at most poly-log(n) edges? This would be an analog of the fact that subcritical percolation cluster size distribution has a geometrically-decreasing tail [5].

2.2 The mean-field model

Instead of the Euclidean model in this paper, one can consider a "mean-field" model on n points for which the $\binom{n}{2}$ inter-point links are assumed to have independent random lengths with Exponential (mean n) distribution. Within this model one can define the function $\tilde{c}(\delta)$ analogous to $c(\delta)$. Mézard-Parisi [10] used the non-rigorous *replica method* of statistical physics to argue $n^{-1}EL_n(1) \to \tilde{c}(1)$, where $\tilde{c}(1) \approx 2.04$ is derived from numerical solution of a certain fixed point equation. Using probabilistic reformulations of these statistical physics ideas, Aldous [1] gave a (still non-rigorous) analysis of the whole function $\tilde{c}(\delta)$, exhibited in Figure 1 of [1], which suggests $\tilde{c}(\delta) - \tilde{c}(0+) \asymp \delta^{1/3}$ as $\delta \downarrow 0$. So one might conjecture that the scaling exponent $\frac{1}{3}$ also holds in the Euclidean case.

Note that the subadditivity arguments we use in the Euclidean case to prove Theorem 1 rest upon the "boundary effects are negligible" property of $\mathbb{R}^d$. In the mean-field model, the limit analog of the Poisson process is a certain infinite random tree, for which boundary effects are not negligible and subadditivity arguments cannot be used. Indeed, Approach 4 was developed in the mean-field setting as a substitute for subadditivity.

6

3 Proofs

We start with Approach 3 and prove part (a) of Theorem 1 in 3.1; then we prove part (b) of Theorem 1 in 3.2 and give a lower bound on $c(0+)$ in 3.3. Sections 3.4, 3.5, 3.6 contain the proofs of the remaining parts (c), (d), (e) of Theorem 1. Part (e) seems hardest, for reasons explained at the start of section 3.6.

3.1 TSP on sparse subsets

We rely on subadditivity arguments as in [13, 14]. These monographs develop general results for sub- or superadditive Euclidean functions satisfying regularity properties. Unfortunately functionals like $L_n(\delta)$ lack the *monotonicity property* ([13] equation (3.5)) and it is not clear whether the *smoothness property* ([14] section 3.3) is both valid and exploitable. We will use the inequality in (b) below as a substitute for monotonicity.

Write $A_1, A_2, \ldots$ for constants depending only on dimension $d \geq 2$. We start with a purely deterministic lemma (note that by scaling the case of general s is equivalent to the case $s = 1$).

Lemma 2 *Let $\{x_1, \ldots, x_n\}$ be arbitrary points in the cube $[0, s]^d$ and let $L(m)$ be the length of the shortest cycle through some $m \leq n$ of these points.*

(a) [**Uniform boundedness**]

$$L(n) \leq A_1 s n^{(d-1)/d}. \tag{3}$$

(b) For $1 \leq m_1 < m_2 \leq n$,

$$\frac{L(m_1)}{m_1} \leq \frac{L(m_2)}{m_2} + \frac{s d^{1/2}}{m_1}.$$

(c) [**Geometric subadditivity**] *Let $k \geq 2$ and let $(\mathbb{C}^j,\ 1 \leq j \leq k^d)$ be the natural partition of $[0, s]^d$ into k^d subcubes of side s/k. Suppose that, for each j, there exists a cycle of length l_j through some subset $S_j \subset \{x_i\} \cap \mathbb{C}^j$. Then there exists a cycle through $\cup_j S_j$ of length at most $\sum_j l_j + A_2 s k^{d-1}$.*

Proof. Parts (a) and (c) are standard ([13] sections 2.2 and 2.3). For (b), let $y_1, y_2, \ldots, y_{m_2}, y_{m_2+1} = y_1$ be a minimum-length cycle attaining $L(m_2)$. Then there exists k such that (interpreting $k + i$ modulo m_2)

$$\frac{1}{m_1} \sum_{i=1}^{m_1} |y_{k+i+1} - y_{k+i}| \leq \frac{L(m_2)}{m_2}$$

because the right side equals the *average* of the left side as k varies. To make a cycle on $\{y_k, \ldots, y_{k+m_1-1}\}$ replace edge (y_{k+m_1-1}, y_{k+m_1}) by edge (y_{k+m_1-1}, y_k), whose length is at most the diameter $sd^{1/2}$ of the cube $[0, s]^d$.

Remark. Lemma 2(a) implies that the worst-case cycle length is the same order of magnitude as the average-case lengths we will be studying. This has the pleasant consequence that events of probability tending to zero will make asymptotically negligible contributions to expectation of length, and so can be ignored: we use this *uniform boundedness property* several times later.

We start analysis of the probability model by making definitions to which subadditivity arguments can easily be applied. Recall (ξ_i) denotes a Poisson point process of rate 1 per unit volume in $\mathbb{R}^d$. Let $N(s)$ be the number of points of (ξ_i) in $[0, s]^d$. Define $L(s, \delta)$ as the length of the minimum-length cycle through at least some $\lceil \delta N(s) \rceil$ of the points of $(\xi_i) \cap [0, s]^d$. By the triangle inequality, replacing "at least $\lceil \delta N(s) \rceil$" by "exactly $\lceil \delta N(s) \rceil$ points" changes nothing. We will start by using the next lemma as a definition of $c(\delta)$, and later show this agrees with the limit in Theorem 1(a).

Lemma 3 *For fixed $0 < \delta \leq 1$ there exists a constant $c(\delta)$ such that*

$$\frac{L(s, \delta)}{\delta s^d} \to c(\delta) \ \text{in } L^2 \ \text{as } s \to \infty.$$

Proof. First note $s \to EL(s, \delta)$ is continuous, from the representation

$$EL(s, \delta) = \sum_{n=0}^{\infty} \frac{s^{nd}}{n!} e^{-s^d} s a_n(\delta),$$

where $a_n(\delta)$ is the expected length of the shortest cycle through some $\lceil \delta n \rceil$ of n uniform random points in the unit cube $[0, 1]^d$. Next, given $s > 0$, we can write any $x > 0$ as $x = ks + t$ with integer k and $t \in [0, s)$. Geometric subadditivity (Lemma 2(c)) then implies

$$EL(x, \delta) \leq k^d EL(s + t/k, \delta) + A_2(s + t/k)k^{d-1}.$$

Taking $x \to \infty$ while keeping s fixed

$$\limsup_{x \to \infty} x^{-d} EL(x, \delta) \leq \inf_{\varepsilon > 0} \sup_{\tau \in [0, \varepsilon]} s^{-d} EL(s + \tau, \delta).$$

So by the continuity property

$$\lim_{s \to \infty} s^{-d} EL(s, \delta) = \inf_{s > 0} s^{-d} EL(s, \delta) = \delta c(\delta), \ \text{say} \tag{4}$$

8

with $0 \le c(\delta) < \infty$.

Once again by geometric subadditivity, temporarily abbreviating $L(s, \delta)$ to L_s,

$$EL_{ks}^2 \le k^d EL_s^2 + k^d(k^d - 1)(EL_s)^2 + A_2 sk^{2d-1}EL_s + (A_2 sk^{d-1})^2.$$

By the uniform boundedness property (3)

$$EL_s^2 \le A_1^2 s^2 E\mathcal{P}_{(s^d)}^{\frac{2(d-1)}{d}} \le A_1^2 s^2 \cdot A_3 s^{2(d-1)} = A_1^2 A_3 s^6,$$

where $\mathcal{P}_{(s^d)}$ has Poisson(s^d) distribution and where the second inequality holds for some A_3 for all $s \ge 1$. Combining the two displayed inequalities and (4) gives

$$\limsup_x x^{-2d} EL_x^2 = (\delta c(\delta))^2,$$

and finally

$$s^{-d}L(s, \delta) \to \delta c(\delta) \text{ in } L^2.$$

Remark. More sophisticated modern proofs ([13] sec. 2.4; [14] sec. 4.1) of the TSP case ($\delta = 1$) use concentration inequalities to obtain almost sure convergence; we have not investigated concentration inequalities or a.s. convergence for $L(s, \delta)$.

Now we proceed with the proof of Theorem 1(a) with $c(\delta)$ defined by Lemma 3.

Proof of Theorem 1(a). Recall the definition of $L_n(\delta)$. Let $\mathbb{C}_n = [0, n^{1/d}]^d$ be the cube of volume n in $\mathbb{R}^d$. Put n random (independent, uniformly distributed) points (ζ_j) into $\mathbb{C}_n$. Fix $0 < \delta \le 1$. Let $L_n(\delta)$ be the minimum, over all choices of cycles $(\zeta_{j_1}, \zeta_{j_2}, \ldots, \zeta_{j_m}, \zeta_{j_{m+1}} = \zeta_{j_1})$ through any chosen $m = \lceil \delta n \rceil$ of the random points, of the cycle length $\sum_{i=1}^m |\zeta_{j_{i+1}} - \zeta_{j_i}|$. Again by the triangle inequality, this is the same as saying "any $m \ge \lceil \delta n \rceil$ of the random points".

Fix small $\varepsilon > 0$ and consider, in $\mathbb{C}_n$, a Poisson process of rate $1 - \varepsilon$ per unit volume. By standard properties of the Poisson process, for each n we can couple this to the process of n i.i.d. uniform points in $\mathbb{C}_n$ in such a way that, with probability $\to 1$ as $n \to \infty$, each point of the Poisson process is a point of the uniform process. Call this the *inclusion coupling*. There is a similar inclusion coupling between the uniform process on $\mathbb{C}_n$ and the Poisson process of rate $1 + \varepsilon$.

Now write $L_\lambda(s, \delta)$ to mean the quantity $L(s, \delta)$ applied to a Poisson process of rate λ; and write $N_\lambda(s)$ for the number of points of that Poisson

process in $[0, s]^d$. When the inclusion couplings hold and when

$$\frac{\delta}{1 + 2\varepsilon} N_{1+\varepsilon}(n^{1/d}) \leq \delta n \leq \frac{\delta}{1 - 2\varepsilon} N_{1-\varepsilon}(n^{1/d})$$

then we have

$$L_{1+\varepsilon}(n^{1/d}, \tfrac{\delta}{1+2\varepsilon}) \leq L_n(\delta) \leq L_{1-\varepsilon}(n^{1/d}, \tfrac{\delta}{1-2\varepsilon}). \tag{5}$$

This holds because, in informal language, for each $\leq$ we have more points to choose from, and a weaker constraint on minimum number of points in the cycle. Now by scaling

$$L_{1-\varepsilon}(n^{1/d}, \tfrac{\delta}{1-2\varepsilon}) = (1 - \varepsilon)^{-1/d} L((1 - \varepsilon)^{-1/d} n^{1/d}, \tfrac{\delta}{1-2\varepsilon}),$$

and similarly

$$L_{1+\varepsilon}(n^{1/d}, \tfrac{\delta}{1+2\varepsilon}) = (1 + \varepsilon)^{-1/d} L((1 + \varepsilon)^{-1/d} n^{1/d}, \tfrac{\delta}{1+2\varepsilon}).$$

Taking limits in (5), using Lemma 3 and the uniform boundedness property, we get

$$\limsup_n n^{-1} E L_n(\delta) \leq (1 - \varepsilon)^{-1-1/d} \tfrac{\delta}{1-2\varepsilon} c(\tfrac{\delta}{1-2\varepsilon})$$

and similarly

$$(1 + \varepsilon)^{-1-1/d} \tfrac{\delta}{1+2\varepsilon} c(\tfrac{\delta}{1+2\varepsilon}) \leq \liminf_n n^{-1} E L_n(\delta).$$

Letting $\varepsilon \downarrow 0$ and using continuity of $c(\delta)$ (which we prove independently in the next subsection) we see that for $0 < \delta < 1$

$$\frac{L_n(\delta)}{\delta n} \to c(\delta) \text{ in } L^1.$$

The case $\delta = 1$ is similar, but of course is already part of the usual proof [13] of the Beardwood-Halton-Hammersley theorem, so we omit it.

3.2 Properties of $c(\delta)$

Next we prove part (b) of Theorem 1.

Proposition 4 $c(\delta)$ *is non-decreasing on* $0 < \delta \leq 1$.

Proof. Fix $0 < \delta_1 < \delta_2 \le 1$. By Lemma 2(b)

$$\frac{L(s, \delta_1)}{\lceil \delta_1 N(s) \rceil} \le \frac{L(s, \delta_2)}{\lceil \delta_2 N(s) \rceil} + \frac{sd^{1/2}}{\lceil \delta_1 N(s) \rceil}$$

and so

$$\frac{L(s, \delta_1)}{\delta_1 s^d} \le R_s \frac{L(s, \delta_2)}{\delta_2 s^d} + \frac{sd^{1/2}}{\delta_1 s^d}$$

where

$$R_s = \frac{\lceil \delta_1 N(s) \rceil / \delta_1}{\lceil \delta_2 N(s) \rceil / \delta_2}.$$

Since R_s is uniformly bounded and $R_s \to 1$ as $s \to \infty$, using Lemma 3 we deduce $c(\delta_1) \le c(\delta_2)$.

Proposition 5 $\delta c(\delta)$ *is convex on* $0 < \delta \le 1$.

Proof. Fix $0 < \delta_1 < \delta_2 \le 1$ and $0 < \lambda < 1$. Let $(\mathbb{C}^j, \; 1 \le j \le k^d)$ be the partition of $[0, s]^d$ into k^d equal subcubes. Let N_j be the number of points of (ξ_i) in the j-th subcube and let $\ell_j(s)$ (resp. $\tilde{\ell}_j(s)$) be the length of the shortest cycle through some $\lceil \delta_1 N_j \rceil$ (resp. $\lceil \delta_2 N_j \rceil$) points in $\mathbb{C}^j$.

Take any $\delta < \lambda \delta_1 + (1 - \lambda)\delta_2$. The event

$$\sum_{j=1}^{\lceil \lambda k^d \rceil} \lceil \delta_1 N_j \rceil + \sum_{j=\lceil \lambda k^d \rceil + 1}^{k^d} \lceil \delta_2 N_j \rceil \; \ge \; \lceil \delta N(s) \rceil$$

has probability $\to 1$ as $s \to \infty$, and on this event we have by Lemma 2(c)

$$L(s, \delta) \le \sum_{j=1}^{\lceil \lambda k^d \rceil} \ell_j(s) + \sum_{j=\lceil \lambda k^d \rceil + 1}^{k^d} \tilde{\ell}_j(s) + A_2 s k^{d-1}.$$

Taking expectations, letting $s \to \infty$ and using Lemma 3 we obtain

$$\delta c(\delta) \le \frac{\lceil \lambda k^d \rceil}{k^d} \, \delta_1 c(\delta_1) + \frac{k^d - \lceil \lambda k^d \rceil}{k^d} \, \delta_2 c(\delta_2),$$

and letting $k \to \infty$

$$\delta c(\delta) \le \lambda \delta_1 c(\delta_1) + (1 - \lambda)\delta_2 c(\delta_2). \tag{6}$$

If $\delta = \lambda' \delta_1 + (1 - \lambda')\delta_2$ then (6) holds for all $\lambda < \lambda'$ and hence for $\lambda = \lambda'$, proving convexity.

Proposition 6 $c(\delta)$ *is continuous on* $(0, 1]$.

Proof. Convexity of $\delta c(\delta)$ implies continuity of $\delta c(\delta)$, and hence continuity of $c(\delta)$, on the open interval $0 < \delta < 1$. Continuity at $\delta = 1$ requires a separate argument.

Take $\delta < 1$. The cycle attaining $L_n(\delta)$ passes through $\lceil n\delta \rceil$ points in $[0, n^{1/d}]^d$. By Lemma 2(a) there exists a cycle through the remaining $q = n - \lceil n\delta \rceil$ points with length at most

$$A_1 n^{1/d} q^{(d-1)/d} \leq A_1 n (1 - \delta)^{(d-1)/d}.$$

By joining the two cycles we find

$$L_n(1) \leq L_n(\delta) + A_1 n (1 - \delta)^{(d-1)/d} + 2n^{1/d} d^{1/2}.$$

Letting $n \to \infty$

$$c(1) \leq \delta c(\delta) + A_1 (1 - \delta)^{(d-1)/d}$$

and this implies continuity as $\delta \uparrow 1$.

3.3 A lower bound on $c(0+)$

We start by noting a one-sided bound relating the T_k in Approach 1 to $c(0+)$. In the context of Theorem 1(a), the cycle attaining length $L_n(\delta)$ can be converted into a path from the origin by replacing some edge (ζ, ζ') by the edge from the origin to ζ'. It follows that

$$\lim_k P(k^{-1} T_k \leq c(\delta) + \varepsilon) = 1 \ \forall \varepsilon > 0$$

and thus

$$\lim_k P(k^{-1} T_k \leq c(0+) + \varepsilon) = 1 \ \forall \varepsilon > 0. \tag{7}$$

We can now use a standard argument. Consider branching random walk on $\mathbb{R}^+$, starting with one individual at the origin in generation 0, and where each individual in each generation (at position x, say) has children at positions $(x + |\xi_j|, \ j \geq 1)$ where (ξ_j) forms a Poisson point process of rate 1 in $\mathbb{R}^d$. Write $\theta_k(\cdot)$ for the mean measure for the positions of the generation-k individuals $(Y_{k,i}, i \geq 1)$:

$$\theta_k(\cdot) = \sum_i P(Y_{k,i} \in \cdot).$$

This is exactly the same measure as the mean measure for lengths of k-step paths from the origin through the Poisson points:

$$\theta_k(\cdot) = \sum_{(j_1,\ldots,j_k)} P\left(|\xi_{j_1}| + \sum_{i=2}^{k} |\xi_{j_i} - \xi_{j_{i-1}}| \in \cdot \right)$$

where the sum is over ordered distinct k-tuples. So for T_k as defined in Approach 1, and for $c > 0, \lambda > 0$

$$
\begin{aligned}
P(T_k \leq ck) & \leq & \theta_k[0, ck] \quad \text{(Markov's inequality)} \\
& \leq & e^{\lambda ck} \int_0^\infty e^{-\lambda x} \theta_k(dx) \quad \text{(large deviation inequality)} \\
& = & \left[e^{\lambda c} \int_0^\infty e^{-\lambda x} \theta_1(dx) \right]^k
\end{aligned}
$$

by the structure of branching random walk. Comparing with (7) we see

$$c(0+) \geq \sup\left\{ c : \inf_{\lambda > 0} e^{\lambda c} \int_{\mathbb{R}^d} e^{-\lambda |y|} \, dy \ < 1 \right\}.$$

Writing v_d for the volume of the unit ball in $\mathbb{R}^d$,

$$\int_{\mathbb{R}^d} e^{-\lambda |y|} \, dy = \frac{v_d \Gamma(d+1)}{\lambda^{d-1}}.$$

The $\inf_{\lambda > 0}(\cdot)$ is now attained at $\lambda = (d-1)/c$ and we finally find

$$c(0+) \geq e^{-1}(d-1)\left(v_d \Gamma(d+1) \right)^{-\frac{1}{d-1}} \tag{8}$$

and of course $v_d = \pi^{d/2}/\Gamma(1+d/2)$. In particular, for $d = 2$ we find $c(0+) \geq (2\pi e)^{-1}$.

3.4 The limit for W_s

We now turn to Approach 2. Here we consider the rate-1 Poisson process (ξ_i) on $\mathbb{R}^d$. Let $\mathbf{\Pi}_{0,s}$ be the set of paths π across the diagonal of $\mathbb{C}_s$; that is, of paths

$$(0,\ldots,0) = \xi_{j_0}, \xi_{j_1}, \ldots, \xi_{j_{m-1}}, \xi_{j_m} = (s,\ldots,s)$$

where $\{\xi_{j_1}, \ldots, \xi_{j_{m-1}}\}$ are distinct points of $\{\xi_i\} \cap [0,s]^d$. Write

$$\ell(\pi) = \sum_{i=1}^{m} |\xi_{j_i} - \xi_{j_{i-1}}|; \quad m(\pi) = m$$

13

for the length and number of edges in the path π and

$$w(\pi) = \ell(\pi)/m(\pi)$$

for the *average edge-length* of π.

By considering diagonally-adjacent unit cubes and picking (where possible) one point from each, we see there exists a path $\pi_{\text{greedy}} \in \mathbf{\Pi}_{0,s}$ such that

$$m(\pi_{\text{greedy}}) - 1 \text{ has Binomial } (\lfloor s \rfloor, 1 - e^{-1}) \text{ distribution;}$$

$$\ell(\pi_{\text{greedy}}) \leq A_4 \lceil s \rceil \tag{9}$$

for some constant A_4. Recall the definition

$$W_s := \min_{\pi \in \mathbf{\Pi}_{0,s}} \frac{\ell(\pi)}{m(\pi)}.$$

Applying to π_{greedy} and using (9) we see

$$\limsup_s W_s \leq A_4/(1 - e^{-1}) \text{ a.s.} \tag{10}$$

Similarly one can check that $s/m(\pi_{\text{greedy}})$ is uniformly integrable as $s \to \infty$ and so

$$(W_s, 0 < s < \infty) \text{ is uniformly integrable.} \tag{11}$$

While subadditivity is not applicable directly to W_s, the proof below is an easy indirect application.

Proposition 7 $W_s \to \beta$ *a.s. and in* L^1 *as* $s \to \infty$, *for some constant* $0 \leq \beta < \infty$.

Proof. For $c \geq 0$ define

$$X_{0,s}^{(c)} = \min_{\pi \in \mathbf{\Pi}_{0,s}} (\ell(\pi) - cm(\pi) + c).$$

We shall prove the Proposition for

$$\beta := \sup\{c : EX_{0,s}^{(c)} \geq 0 \; \forall s > 0\}.$$

For $s < t$ write $\mathbf{\Pi}_{s,t}$ for the set of paths across the diagonal of $[s,t]^d$ and define $X_{s,t}^{(c)}$ analogously to $X_{0,s}^{(c)}$. Given a path $\pi_1 \in \mathbf{\Pi}_{0,s}$ and a path $\pi_2 \in \mathbf{\Pi}_{s,t}$, their concatenation gives a path $\pi \in \mathbf{\Pi}_{0,t}$, for which

$$\ell(\pi) \leq \ell(\pi_1) + \ell(\pi_2); \quad m(\pi) = m(\pi_1) + m(\pi_2) - 1. \tag{12}$$

14

In such a concatenation, the last edge of π_1 and the first edge of π_2 are replaced by a single edge and the inequality for $\ell(\pi)$ arises only from the triangle inequality for this replacement.

Consider first a value c such that $EX_{0,s}^{(c)} < 0$ for some s. For this s let π be the random path in $\mathbf{\Pi}_{0,s}$ such that $E(\ell(\pi) - cm(\pi) + c) = EX_{0,s}^{(c)} < 0$. So by the concatenation property (12) and the strong law of large numbers we can construct random paths π_k in $\mathbf{\Pi}_{0,ks}$ such that

$$\limsup_{k\to\infty} \frac{\ell(\pi_k)}{m(\pi_k) - 1} \leq \frac{E\ell(\pi)}{Em(\pi) - 1} < c \text{ a.s.}$$

and it easily follows that

$$\limsup_{s\to\infty} W_s < c \text{ a.s.}$$

From choice of c and definition of β we deduce

$$\limsup_{s\to\infty} W_s \leq \beta \text{ a.s.}$$

On the other hand consider a value of c such that $EX_{0,s}^{(c)} \geq 0$ for all $s > 0$. The process $(X_{s,t}^{(c)})$ is subadditive by (12). A routine application of the subadditive ergodic theorem shows that there exists a constant $\gamma(c) \geq 0$ such that

$$\lim_{s\to\infty} s^{-1} X_{0,s}^{(c)} = \gamma(c) \text{ a.s.} \tag{13}$$

Now $W_s = \ell(\pi_s)/m(\pi_s)$ for some random π_s in $\mathbf{\Pi}_s$, and

$$X_{0,s}^{(c)} \leq \ell(\pi_s) - cm(\pi_s) + c,$$

implying

$$W_s - c \geq \frac{X_{0,s}^{(c)} - c}{m(\pi_s)}. \tag{14}$$

Combining (13) with (14) gives

$$\liminf_s (W_s - c) \geq \gamma(c) \cdot \liminf_s \frac{s}{m(\pi_s)} \geq 0 \text{ a.s.}$$

and consequently

$$\liminf_s W_s \geq c \text{ a.s.}$$

15

From choice of c and definition of β we deduce

$$\liminf_{s \to \infty} W_s \geq \beta \text{ a.s.}$$

completing the proof of a.s. convergence in Proposition 7. Finally, note $\beta < \infty$ by (10), and note that L^1 convergence follows from (11).

We use concatenation constructions based on (12) several times later, and here is a rather technical formulation of the results of such constructions, designed to replace the use of the expectations $E\ell(\pi_s)$ and $Em(\pi_s)$ by using truncation. By "adjacent cubes" we mean disjoint cubes in which we can choose diagonals to form a connected path (e.g. as in Figure 1 below).

Lemma 8 (Concatenation argument) *Let π_s be a random path across the diagonal of the cube $[0, s]^d$, and let $w_0 \in (0, s)$. Then there exist paths π_{sn} through the Poisson points in n adjacent cubes of side s each, such that*

$$\limsup_n w(\pi_{sn}) \leq w_0 + \frac{w_0 + sd^{1/2}P(w(\pi_s) > w_0)}{(sd^{1/2}/w_0 - 1)P(w(\pi_s) \leq w_0)} \quad a.s. \quad (15)$$

and

$$\lim_n \frac{m(\pi_{sn})}{n} \geq (sd^{1/2}/w_0 - 1)P(w(\pi_s) \leq w_0) \quad a.s.$$

Proof. Given a random path π_s across the diagonal of the cube $[0, s]^d$, consider the modified path

$$\tilde{\pi}_s = \begin{cases} \pi_s & \text{on the event } w(\pi_s) \leq w_0, \\ \nearrow_s & \text{otherwise,} \end{cases}$$

where $\nearrow_s$ is a shortcut path consisting of a single edge from $(0, \ldots, 0)$ to $(s, \ldots, s)$. By concatenating n independent copies of $\tilde{\pi}_s$ we get a path π_{ns} such that, using (12),

$$\begin{aligned}
w(\pi_{ns}) &\leq \frac{\sum\limits_{i=1}^n \ell(\tilde{\pi}_s^{(i)})}{\sum\limits_{i=1}^n m(\tilde{\pi}_s^{(i)}) - (n-1)} \leq \frac{w_0 \sum\limits_{i=1}^n m(\tilde{\pi}_s^{(i)}) + sd^{1/2}k}{\sum\limits_{i=1}^n m(\tilde{\pi}_s^{(i)}) - (n-1)} \\
&\leq w_0 + \frac{w_0(n-1) + sd^{1/2}k}{\sum\limits_{i=1}^n m(\tilde{\pi}_s^{(i)}) - (n-1)} \quad (16)
\end{aligned}$$

16

where $k = \#\{i : w(\pi_s^{(i)}) > w_0$. By definition of $\tilde{\pi}_s$ either $m(\tilde{\pi}_s) \geq sd^{1/2}/w_0$, or $m(\tilde{\pi}_s) = 1$, so that

$$m(\pi_{sn}) = \sum_{i=1}^{n} m(\tilde{\pi}_s^{(i)}) - (n-1) \quad \geq \quad (n-k)(sd^{1/2}/w_0) + k - (n-1)$$

$$\geq \quad (n-k)(sd^{1/2}/w_0 - 1). \qquad (17)$$

Since $k/n \to P(w(\pi_s) > w_0)$ a.s. as $n \to \infty$, combining (16) with (17) gives (15) and the second inequality of the lemma follows from (17).

Proposition 9 $\beta = c(0+)$.

Proof. Given a cycle through m points in $[0, s]^d$, one can make a path from $(0, \ldots, 0)$ to $(s, \ldots, s)$ through these m points using extra length at most $2sd^{1/2}$. So for any $\delta > 0$ we have, setting $s = n^{1/d}$ and $m = \lceil \delta n \rceil$,

$$W_{n^{1/d}} \leq \frac{L_n(\delta) + 2n^{1/d}d^{1/2}}{\lceil \delta n \rceil + 1}.$$

Letting $n \to \infty$ and using Proposition 7 and Theorem 1(a), we see $\beta \leq c(\delta)$. So $\beta \leq c(0+)$.

For the converse, consider the cube $[0, Ks]^d$ for even K, partitioned into K^d equal subcubes. In each subcube choose a diagonal path $\pi_s^{(i)}$ attaining the minimum W_s, so that concatenating these paths forms a cycle through the large cube (see Figure 1).

Fix $\varepsilon > 0$. By Proposition 7, for $s = s(\varepsilon)$ large enough the random paths $\pi_s^{(i)}$ satisfy

$$P(w(\pi_s^{(i)}) > \beta + \varepsilon) \leq \varepsilon.$$

Then by Lemma 8 applied with $w_0 = \beta + \varepsilon$ we can construct cycles π_{K^d} in $[0, Ks]^d$ such that

$$P\left(w(\pi_{K^d}) > \beta + \varepsilon + \frac{\beta + \varepsilon + sd^{1/2}\varepsilon}{(1-\varepsilon)(sd^{1/2}/(\beta+\varepsilon) - 1)} + \varepsilon\right) \to 0 \text{ as } K \to \infty.$$

$$(18)$$

By taking s sufficiently large this simplifies to

$$P(w(\pi_{K^d}) > \beta + 2\varepsilon + 2\beta\varepsilon) \to 0 \text{ as } K \to \infty. \qquad (19)$$

Also from Lemma 12 there exists $\delta_s > 0$ such that

$$P\left(\frac{m(\pi_{K^d})}{s^d K^d} \geq \delta_s\right) \to 1 \text{ as } K \to \infty.$$
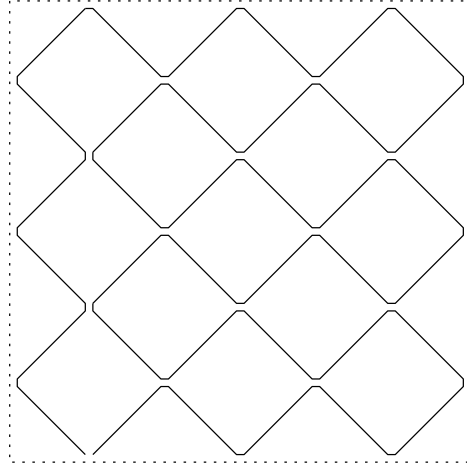
17

Figure 1: A cycle built from subcube diagonals in the 6x6 cube

On this event we have, for $\delta < \delta_s$,

$$L(Ks, \delta) \le w(\pi_{K^d}) m(\pi_{K^d}).$$

So as $K \to \infty$

$$P(L(Ks, \delta) \le (\beta + 2\varepsilon + 2\beta\varepsilon) s^d K^d \delta_s) \to 1$$

and then by Lemma 3

$$\delta c(\delta) \le (\beta + 2\varepsilon + 2\beta\varepsilon)\delta_s.$$

Letting $\delta \uparrow \delta_s$ then implies

$$c(0+) \le \beta + 2\varepsilon + 2\beta\varepsilon,$$

so that letting $\varepsilon \downarrow 0$ gives the desired inequality

$$c(0+) \le \beta.$$

Propositions 7 and 9 establish part (c) of Theorem 1. For later use we record a small variation.

Lemma 10 (linear diagonal percolation) *For $\eta > 0$ let*

$$W_s^{(\eta)} := \min_{\pi \in \mathbf{\Pi}_{0,s}: m(\pi) \le \eta s} \frac{\ell(\pi)}{m(\pi)}$$

18

be the minimum average edge-length over paths across the diagonal of $[0, s]^d$ through at most ηs points. Then there exists a function $\beta(\eta)$ such that

$$\lim_{s \to \infty} W_s^{(\eta)} = \beta(\eta) \quad a.s.,$$

$$\beta(\eta) \downarrow c(0+) \ as \ \eta \uparrow \infty.$$

Proof. The subadditivity argument used in Proposition 7 applies unchanged to $W_s^{(\eta)}$, giving the first limit. The function $\eta \to \beta(\eta)$ is *a priori* non-increasing, and

$$\lim_{\eta \to \infty} \beta(\eta) \geq \beta = c(0+)$$

by Proposition 9.

Given $s < \infty$ let

$$\pi_s \text{ be a path attaining } W_s$$

$$\pi_s' \text{ be a path attaining } W_s^{(\eta)} \text{ for } \eta = 2s^{d-1} \ .$$

Given $\varepsilon > 0$ we can choose $s = s(\varepsilon) < \infty$ sufficiently large that

$$P(\pi_s' \neq \pi_s) < \varepsilon$$

(because $\eta s = 2s^d$ will likely exceed the Poisson(s^d) number of points in $[0, s]^d$) and

$$P(w(\pi_s) > c(0+) + \varepsilon) < \varepsilon$$

by Propositions 7 and 9. So

$$P(w(\pi_s') > c(0+) + \varepsilon) < 2\varepsilon.$$

Applying Lemma 12 to π_s' and $w_0 = c(0+) + \varepsilon$ gives the upper bound in

$$\beta(2s^{d-1}) \leq \limsup_k w(\pi_{sk}') \leq c(0+) + \varepsilon + \frac{c(0+) + \varepsilon + sd^{1/2}2\varepsilon}{(\frac{sd^{1/2}}{c(0+)+\varepsilon} - 1)(1 - 2\varepsilon)}$$

and the lower bound follows from the construction in Lemma 12. Taking $\varepsilon \to 0$, so that $s = s(\varepsilon) \to \infty$,

$$\lim_{\eta \to 0} \beta(\eta) \leq c(0+).$$

19

3.5 Translation invariant distributions in the infinite model

Here we prove part (d) of Theorem 1 by relating Approach 4 to Approach 3. Fix $0 < \delta \leq 1$ and recall (ξ_i) denotes a Poisson point process of rate 1 per unit volume in $\mathbb{R}^d$. Recall from section 3.1 the definition of $L(s, \delta)$. Write $\mathcal{E}(s, \delta)$ and $\mathcal{V}(s, \delta)$ for the edge-set and the vertex-set of a cycle attaining length $L(s, \delta)$. For a cube $\mathbb{C}$, vertex-set $\mathcal{V}$ and edge-set $\mathcal{E}$, write $|\mathcal{V} \cap \mathbb{C}|$ for the number of vertices of $\mathcal{V}$ in $\mathbb{C}$, and write $\mathrm{len}(\mathcal{E} \cap \mathbb{C})$ for the total length of edges of $\mathcal{E}$ restricted to $\mathbb{C}$.

Write U_s for a uniform random position in $[0, s]^d$, and write $\tilde{\xi}_i^{(s)} = \xi_i - U_s$ for the positions of the Poisson points relative to the "random origin" U_s; then write $\widetilde{\mathcal{E}}(s, \delta)$ for the corresponding set of relative positions of edges of the tour attaining $L(s, \delta)$:

$$\widetilde{\mathcal{E}}(s, \delta) = \{(\tilde{\xi}_i^{(s)}, \tilde{\xi}_j^{(s)}) : \ (\xi_i, \xi_j) \in \mathcal{E}(s, \delta)\}.$$

Now the pair $\left((\tilde{\xi}_i^{(s)}), \widetilde{\mathcal{E}}(s, \delta) \right)$ takes values in the space $\mathbf{S}^*$ of point-sets and paths defined as the space $\mathbf{S}$ in section 2, except that for $\mathbf{S}^*$ we allow cycles in addition to doubly-infinite paths.

There is a natural metric topology on $\mathbf{S}^*$ obtained by regarding it as a space of marked point processes. For each s the point process $(\tilde{\xi}_i^{(s)})$ is exactly a Poisson process. By letting $s \to \infty$ through some subsequence we can define a limit

$$\left((\tilde{\xi}_i^{(s)}), \widetilde{\mathcal{E}}(s, \delta) \right) \ \xrightarrow{d} \ ((\xi_i), \mathcal{E}(\delta)) \ \text{ on } \mathbf{S}^*$$

where $\mathcal{E}(\delta)$ is an edge set on some subset $\mathcal{V}(\delta)$ of vertices of the Poisson point process (ξ_i). It is easy to check that the edge-set forms doubly-infinite paths (rather than finite cycles) and so the right side has some distribution μ on $\mathbf{S}$. From the uniform distribution for U_s it is easy to check that μ is translation invariant. We will show (in the notation of section 2)

$$\delta(\mu) = \delta; \quad \ell(\mu) \leq c(\delta)$$

which immediately implies $\bar{c}(\delta) \leq c(\delta)$. With a little extra effort we could prove $\ell(\mu) = c(\delta)$, but this doesn't help.

From the fact that $\mathcal{V}(s, \delta)$ contains exactly $\lceil \delta N(s) \rceil$ of the $N(s)$ Poisson points in $[0, s]^d$, it is clear that for fixed $r > 0$

$$\lim_{s \to \infty} E \, |\mathcal{V}(s, \delta) \cap (U_s + \mathbb{C}_r)| = \delta r^d$$

20

and therefore
$$E \, |\mathcal{V}(\delta) \cap \mathbb{C}_r| = \delta r^d,$$
the interchange of limits being justified by the fact that the total number of Poisson points in $U_s + \mathbb{C}_r$ has fixed distribution. This tells us that $\delta(\mu) = \delta$. Now fix $b < \infty$ and let $\mathcal{E}^b(\delta)$ and $\mathcal{E}^b(s, \delta)$ be the subsets of $\mathcal{E}(\delta)$ and $\mathcal{E}(s, \delta)$ obtained by taking only edges of length $\leq b$. By translation invariance
$$E \, \mathrm{len}(\mathcal{E}^b(\delta) \cap \mathbb{C}_r) = \delta G(b) r^d$$
for some $G(b)$, and then
$$\ell(\mu) = \delta \int_0^\infty b \, dG(b).$$
By weak convergence, letting $s \to \infty$ through a subsequence,
$$
\begin{aligned}
E \, \mathrm{len}(\mathcal{E}^b(\delta) \cap \mathbb{C}_1) &= \lim_s E \, \mathrm{len}(\mathcal{E}^b(s, \delta) \cap (U_s + \mathbb{C}_1)) \\
&= \lim_s s^{-d} E \, \mathrm{len}(\mathcal{E}^b(s, \delta)).
\end{aligned}
$$
Since $\delta c(\delta) = \lim_s s^{-d} E \, \mathrm{len}(\mathcal{E}(s, \delta))$, we can let $b \to \infty$ and use Fatou's lemma to conclude $\ell(\mu) \leq c(\delta)$.

For the converse, let μ be a translation invariant distribution attaining $\bar{c}(\delta)$. That is, μ specifies an edge-set $\mathcal{E}$ on a vertex-subset $\mathcal{V} \subset (\xi_i)$ such that
(i) $E|\mathcal{V} \cap \mathbb{C}_r| = \delta r^d \quad \forall r > 0$;
(ii) $E \, \mathrm{len}(\mathcal{E} \cap \mathbb{C}_r) = \delta \bar{c}(\delta) r^d \quad \forall r > 0$.
Now consider large s and small $\eta > 0$. The intersection of $\mathcal{E}$ and $\mathbb{C}_s$ consists of a set of paths, each of which enters $\mathbb{C}_s$ at some point on some face, and exits at some point on some face. Consider the subset of paths which intersect $[\eta s, (1-\eta)s]^d$, and write $\mathcal{E}^{s,\eta}$ for the edges in this subset, truncating an edge which crosses a face at the crossing point. We extend this edge-set into a cycle in $\mathbb{C}_s$ as follows. Suppose the number of crossing points on each face is some even number (otherwise add superfluous edges, making no asymptotic difference); let D be the total number of crossing points. On each face create a tour (in the face) of the crossing points; by Lemma 2(a) this has length at most $A_1 s D^{(d-2)/(d-1)}$. Within each such tour replace alternate edges by double edges. The collection of these within-face edges, and the paths through the interior of $\mathbb{C}_s$, form a connected graph where each vertex has even degree, so we can find a Eulerian cycle; write $\mathcal{F}^{s,\eta}$ for the edges in this cycle. So
$$E \, \mathrm{len}(\mathcal{F}^{s,\eta}) \leq \delta \bar{c}(\delta) s^d + (4d) A_1 s \, (ED)^{(d-2)/(d-1)} . \tag{20}$$

Because each crossing point is associated with some path-segment of length $\geq \eta s$ from the face of $\mathbb{C}_s$ to the face of $[\eta s, (1-\eta)s]^d$, we have

$$D \eta s \leq \mathrm{len}(\mathcal{E} \cap \mathbb{C}_s)$$

and so by (ii)

$$ED \leq (\eta s)^{-1} \delta \bar{c}(\delta) s^d.$$

Substituting into (20) gives

$$\limsup_s s^{-d} E \, \mathrm{len}(\mathcal{F}^{s,\eta}) \leq \delta \bar{c}(\delta). \qquad (21)$$

Now the cycle $\mathcal{F}^{s,\eta}$ goes through all the vertices $\mathcal{V} \cap [\eta s, (1-\eta)s]^d$. By (i)

$$E \left| \mathcal{V} \cap [\eta s, (1-\eta)s]^d \right| = \delta(1-2\eta)^d s^d$$

and by comparison with the total number of Poisson points in $\mathbb{C}_s$ we get a crude bound

$$\mathrm{var} \left| \mathcal{V} \cap [\eta s, (1-\eta)s]^d \right| \leq s^d(s^d + 1) \leq 2s^{2d} \quad (s \geq 1).$$

Now consider $k \geq 2$. Using geometric subadditivity (Lemma 2(c)) and independent copies of $\mathcal{F}^{s,\eta}$ on each of the k^d subcubes, we can construct a cycle $\mathcal{G}^{ks,\eta}$ in $[0, ks]^d$ such that (21) extends to

$$\limsup_s (ks)^{-d} E \, \mathrm{len}(\mathcal{G}^{ks,\eta}) \leq \delta \bar{c}(\delta). \qquad (22)$$

Chebyshev's inequality implies

$$\begin{aligned}
P(\text{ number of vertices in } \mathcal{G}^{ks,\eta} &\leq & \delta(ks)^d((1-2\eta)^d - \eta) \\
&\leq & \frac{2k^d s^{2d}}{(\delta k^d s^d \eta)^2} = 2\delta^{-2} \eta^{-2} k^{-d}.
\end{aligned}$$

Taking $k = k(s) \to \infty$ sufficiently slowly, (22) remains true, so by definition of $c(\cdot)$ and the uniform boundedness property

$$\delta((1-2\eta)^d - 2\eta) \; c(\delta((1-2\eta)^d - 2\eta)) \leq \delta \bar{c}(\delta).$$

Letting $\eta \to 0$ and using continuity of $c(\cdot)$ shows $c(\delta) \leq \bar{c}(\delta)$.

3.6 The limit for T_m

Recall the definition (Approach 1) of T_m. If we could prove a sublinear growth property, that the optimal path stays within a ball of radius $o(m)$, then part (e) of Theorem 1 would follow easily from the other parts. Because we cannot prove this property we use a more circuitous method which eventually (Lemma 14) compares a general path from the origin with either a cycle through the origin (Lemma 13) or a path between diagonals of a cube (Lemma 12).

Recall Lemma 10 on linear diagonal percolation. We need a stronger result. Given a realization of the Poisson point process in $[0, s]^d$, first remove some μs points, then look for the optimal diagonal path through at most ηs of the remaining points. We show that the same constant $\beta(\eta)$ appears in the limit.

Lemma 11 *Given the points of the Poisson process in $[0, s]^d$, for any $\eta > 0$, $\mu > 0$ let*

$$W_s^{(\eta,\mu)} := \max_\Delta \min_{\pi:\pi\cap\Delta=\emptyset} \frac{\ell(\pi)}{m(\pi)}$$

where the maximum is taken over all subsets Δ of points with size at most μs and the minimum over all paths $\pi \in \mathbf{\Pi}_{0,s}$ through at most ηs points that exclude Δ. Then

$$\lim_{s\to\infty} W_s^{(\eta,\mu)} = \beta(\eta) \ a.s.$$

Proof. We consider the two-dimensional case. The same proof works for dimension $d \geq 3$ without significant changes.

Consider the square $[0, k^2 s]^2$, partitioned into subsquares of side s. Consider the $2k + 1$ subsquares (S_i, $1 \leq i \leq 2k + 1$) whose lower left corners are $(k, 0), (k, 1), (k-1, 1), (k-1, 2), \ldots, (1, k), (0, k)$. For each i consider the "staircase" consisting of $k^2 - k$ diagonally-adjacent subsquares starting with S_i. Note that all these subsquares are distinct. See Figure 2.

For each such staircase consider a joint path $\pi_k^{(i)}$, obtained by concatenation of $k^2 - k$ paths each attaining $W_s^{(\eta)}$ in its own subcube.

Given $\eta > 0$ and $\varepsilon > 0$, using Lemma 8 and Lemma 10 we can choose large s so that

$$\limsup_k w(\pi_k^{(i)}) \leq \beta(\eta) + \varepsilon,$$

thus with probability $\to 1$ as $k \to \infty$ we can choose at least $3k/2$ staircases i satisfying

$$m(\pi_k^{(i)}) \leq \eta k^2 s, \qquad w(\pi_k^{(i)}) \leq \beta(\eta) + 2\varepsilon. \tag{23}$$
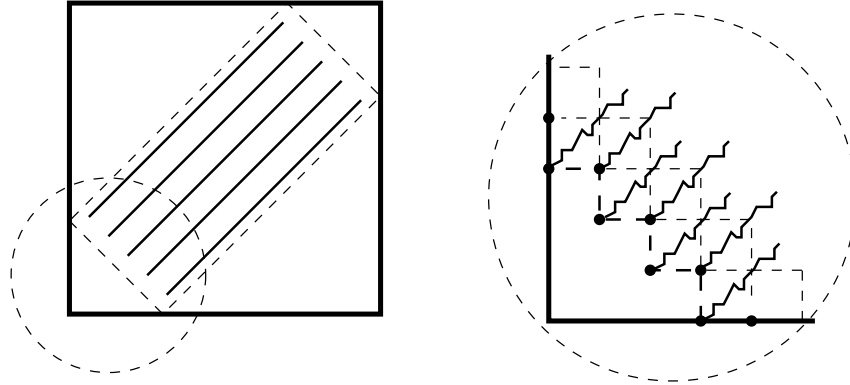
Figure 2: $2k + 1$ diagonal paths in a large square

For any subset of points Δ of size at most μk^2 there are at least k staircases containing no more than μk points of Δ each. Thus with probability 1, for all sufficiently large k and all such Δ we can find some $\pi_k^{(i)}$ satisfying (23) and

$$\pi_k^{(i)} \text{ contains } \leq \mu k \text{ points of } \Delta.$$

From such a path $\pi_k^{(i)}$ we create a path $\tilde{\pi}_k \in \mathbf{\Pi}_{0,k^2 s}$ by first deleting the points in Δ and then adding two edges from the path-ends to the corners to $[0, k^2 s]^d$. These changes have asymptotically negligible effect on average edge-length, so we have shown

$$\lim_{k \to \infty} W_{k^2 s}^{(\eta, \mu)} \leq \beta(\eta) + 2\varepsilon \text{ a.s.}$$

and the result follows easily.

Next we will use Lemma 11 to show that when one weakens conditions on $W_s^{(\eta)}$ by allowing paths through points outside the cube $[0, s]^d$, this doesn't affect the limit $\beta(\eta)$.

Lemma 12 (unbounded linear percolation) *Let $\tilde{W}_s^{(\eta)}$ be the minimal average step length over all paths from $(0, ...0)$ to $(s, ...s)$ through some points of a rate 1 Poisson point process in $\mathbb{R}^d$, with at most ηs steps. Then*

$$\lim_s \tilde{W}_s^{(\eta)} = \beta(\eta) \quad \text{in probability.}$$

Proof. Fix η and consider the possibility that, for some $\beta' < \beta(\eta)$

$$\limsup_s P(\tilde{W}_s^{(\eta)} < \beta') > 0. \tag{24}$$

It is enough to show this cannot happen.

Let π_s be the optimal path for $\tilde{W}_s^{(\eta)}$. There exists a constant ρ such that π_s is contained inside $[-\rho s, s + \rho s]^d$ a.s. for large s. In the cubes $[-\rho s, 0]^d$ and $[s, s + \rho s]^d$ take two diagonal paths π_s^1, π_s^2 with no more than $\eta\rho$ steps each, that attain minimal step length while avoiding the points of π_s. By Lemma 11 the average edge-length in π_s^1 and in π_s^2 is asymptotically $\beta(\eta)$. Then the concatenated path $\{\pi_s^1, \pi_s, \pi_s^2\}$ satisfies the constraints in Lemma 10 for a diagonal path across and within the large cube $[-\rho s, (\rho + 1)s]$ with at most $\eta(1 + 2\rho)s$ points. On the event in (24) this path has average step length less than some $\beta'' < \beta(\eta)$. So (24) contradicts the conclusion of Lemma 10.

Lemma 13 (free cycle) *Let $L_m(0)$ be the minimum length over all cycles on m points of the Poisson process that pass through the origin. Then*

$$\lim_{m\to\infty} L_m(0)/m \geq c(0+) \quad \text{in probability.}$$

Remarks. Such a cycle contains two edges from the origin to points of the Poisson process. Our results ultimately imply the inequality is really an equality, but we won't need to prove that now.

Proof. Assume, to get a contradiction, that for some $c < c(0+)$

$$\limsup_{m\to\infty} P(L_m(0)/m < c) > 0. \tag{25}$$

Let $\rho > 0$ be a constant such that the cycle π_m attaining length $L_m(0)$ is contained inside a box $[-\rho m, \rho m]^d$ a.s. for large m.

Fix $\eta > 0$. By Lemma 11 it's possible to choose two paths π_m^1 and π_m^2 across the diagonals of $[-\rho m, 0]^d$ and $[0, \rho m]^d$ respectively in such a way that π_m^1, π_m^2 have no common points with π_m except the origin, and

$$n_1 \leq \eta\rho m, \qquad l_1 = \beta(\eta)n_1 + o(m)$$

$$n_2 \leq \eta\rho m, \qquad l_2 = \beta(\eta)n_2 + o(m)$$

where n_1, n_2 and l_1, l_2 denote the number of edges and the length of π_m^1 and π_m^2.

One can adjust a few edges around the origin to get a concatenated path $\{\pi_m^1, \pi_m, \pi_m^2\}$ with length $L' \leq l_1 + l_2 + L_m(0)$ and with $n' = n_1 + n_2 + m - 2$ steps, across the diagonal of the whole cube $[-\rho m, \rho m]$. Observe

$$\frac{n_1 + n_2 + m}{2\rho m} \leq \eta + \frac{1}{2\rho}. \tag{26}$$

25

On the event $\{L_m(0)/m < c\}$ in (25) we have

$$\frac{l_1 + l_2 + L_m(0)}{n_1 + n_2 + m} - \beta(\eta) \leq \frac{o(m) + L_m(0) - \beta(\eta)m}{n_1 + n_2 + m}$$

$$\leq \frac{o(m) + (c - \beta(\eta))m}{n_1 + n_2 + m}$$

$$\leq \frac{c - \beta(\eta)}{1 + 2\eta\rho} + o(1) \leq \frac{c - c(0+)}{1 + 2\eta\rho} + o(1)$$

the final two inequalities because $c - \beta(\eta) \leq c - c(0+) < 0$. Because of (26) we can apply Lemma 10 and deduce

$$\beta(\eta + \tfrac{1}{2\rho}) \leq \beta(\eta) - \frac{c(0+) - c}{2\eta\rho + 1}.$$

But this implies $\lim_{\eta \to \infty} \beta(\eta) = -\infty$, which is impossible.

We can finally prove part (e) of Theorem 1.

Lemma 14 (free path) *Let T_m be the minimal length over all paths through m points of the Poisson process starting at the origin. Then*

$$\lim_m T_m/m = c(0+) \quad \text{in probability.}$$

Proof. The upper bound

$$\lim_k P(k^{-1}T_k \leq c(0+) + \varepsilon) = 1 \ \forall \varepsilon > 0$$

was given at (7). For the lower bound suppose, for some $c < c(0+)$

$$\limsup_m P(T_m/m < c) > 0. \tag{27}$$

Let π_m be a path attaining length T_m and let x_m be the position of its end-point. Consider

$$r_* = \sup\{r : \limsup_m P(|x_m| > rm, T_m < cm) > 0\}.$$

Clearly $r_* < \infty$. Suppose first $r_* = 0$. Then we can construct a free cycle from π_m by adding an edge from x_m to the origin, and such cycles have average length $\leq c + o(1)$, contradicting Lemma 13.

Alternatively suppose $r_* > 0$. Then using rotational invariance of the Poisson process, for arbitrary $\varepsilon > 0$

$$\limsup_m P(x_m \in [(r - \varepsilon)m, (r + \varepsilon)m]^d \text{ and } T_m < cm) > 0.$$

Adding an edge from x_m to $((r+\varepsilon)m, \ldots, (r+\varepsilon)m)$ gives a path between the diagonal corners of $[0, (r+\varepsilon)m]^d$ through m points with average edge length at most $c + 2\varepsilon d^{1/2}$, which contradicts Lemma 12 when ε is small (because $\beta(\eta) \geq c(0+)$). So (27) is false and we have proved $T_m/m \to c(0+)$ in probability.

4 Final Remarks

4.1 Subadditivity and cost-reward problems

The technique in section 3.4 seems applicable in many contexts where subadditivity is used. For instance, to modify the context of first-passage percolation on the lattice, suppose that for each edge e there is a "cost" $c(e)$ and a "reward" $r(e)$, so that for each path $\pi = (e_1. e_2, \ldots, e_n)$ there is a cost and a reward $(c(\pi), r(\pi)) = (\sum_i c(e_i), \sum_i r(e_i))$. Then one can study by the same technique

$$A_n := \min_{\pi} c(\pi)/r(\pi)$$

minimized over paths across the diagonal of $[0, n]^d$. Developing a general theorem which establishes limits $n^{-1} A_n \to a_0$ a.s. in such settings would be a natural research project.

References

[1] D.J. Aldous. Percolation-like scaling exponents for minimal paths and trees in the stochastic mean-field model. *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci*, 461:825–838, 2005.

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

[3] I. Benjamini, G. Kalai, and O. Schramm. First passage percolation has sublinear distance variance. *Ann. Probab.*, 31:1970–1978, 2003.

[4] R. Durrett. Harry Kesten's publications: a personal perspective. In M. Bramson and R. Durrett, editors, *Perplexing Problems in Probability*, pages 1–33. Birkhauser, 1999.

[5] G.R. Grimmett. *Percolation*. Springer-Verlag, Berlin, 2nd edition, 1999.

[6] C. D. Howard. Models of first-passage percolation. In H. Kesten, editor, *Probability on Discrete Structures*, volume 110 of *Encyclopaedia of Mathematical Sciences*, pages 125–173. Springer-Verlag, 2004.

[7] J.L. Jacobsen, N. Read, and H. Saleur. The traveling salesman problem, conformal invariance, and dense polymers. *Phys. Rev. Lett.*, 93:038701–038704, 2004. arXiv:cond-mat/0403277.

[8] H. Kesten. First-passage percolation. In *From Classical to Modern Probability*, number 54 in Progr. Probab., pages 93–143. Birkhauser, 2003.

[9] R. Meester and R. Roy. *Continuum Percolation*. Cambridge University Press, 1996. Cambridge Tracts in Math. 119.

[10] M. Mézard and G. Parisi. A replica analysis of the travelling salesman problem. *J. Physique*, 47:1285–1296, 1986.

[11] W. Rhee. On the fluctuations of the traveling salesperson problem. *Math. Oper. Res.*, 16:482–489, 1991.

[12] J.M. Steele. Complete convergence of short paths and Karp's algorithm for the TSP. *Math. Oper. Res.*, 6:374–378, 1981.

[13] J.M. Steele. *Probability Theory and Combinatorial Optimization*. Number 69 in CBMS-NSF Regional Conference Series in Applied Math. SIAM, 1997.

[14] J.E. Yukich. *Probability Theory of Classical Euclidean Optimization Problems*. Number 1675 in Lecture Notes in Math. Springer, 1998.