Near-Minimal Spanning Trees: the Scaling Exponent in Probability Models

David J. Aldous*

University of California
Department of Statistics
367 Evans Hall # 3860
Berkeley CA 94720-3860

May 13, 2005

Abstract

xxx Not intended for publication in this form – I am looking for someone to prove the Ansatz!

*Research supported by N.S.F. Grant DMS0203062

1
1 Introduction

This paper gives details of one (rather simple) aspect of the following broad project [1]. Freshman calculus tells us how to find a minimum $x^*$ of a smooth function $f(x)$: set the derivative $f'(x^*) = 0$ and check $f''(x^*) > 0$. The related series expansion tells us, for points $x$ near to $x^*$, how the distance $\delta = |x - x^*|$ relates to the difference $\varepsilon = f(x) - f(x^*)$ in $f$-values: $\varepsilon$ scales as $\delta^2$. This scaling exponent 2 persists for functions $f : \mathbb{R}^d \to \mathbb{R}$: if $x^*$ is a local minimum and $\varepsilon(\delta) := \min \{f(x) - f(x^*) : |x - x^*| = \delta\}$, then $\varepsilon(\delta)$ scales as $\delta^2$ for a generic smooth function $f$.

Combinatorial optimization, exemplified by the traveling salesman problem (TSP), is traditionally viewed as a quite distinct subject, with theoretical analysis focussed on the number of steps that algorithms require to find the optimal solution. To make a connection with calculus, compare an arbitrary tour $x$ through $n$ points with the optimal (minimum-length) tour $x^*$ by considering the two quantities

\[
\begin{align*}
\delta_n(x) &= \{\text{number of edges in } x \text{ but not in } x^*\}/n \\
\varepsilon_n(x) &= \{\text{length difference between } x \text{ and } x^*\}/s(n)
\end{align*}
\]

where $s(n)$ is the length of the minimum length tour. Now define $\varepsilon_n(\delta)$ to be the minimum value of $\varepsilon_n(x)$ over all tours $x$ for which $\delta_n(x) \geq \delta$. Although the function $\varepsilon_n(\delta)$ will depend on $n$ and the problem instance, we anticipate that for typical instances drawn from a suitable probability model it will converge in the $n \to \infty$ limit to some deterministic function $\varepsilon(\delta)$. The universality paradigm from statistical physics [7] suggests there might be a scaling exponent $\alpha$ defined by

\[\varepsilon(\delta) \sim \delta^\alpha \text{ as } \delta \to 0\]

and that the exponent should be robust under model details.

There is fairly strong evidence [1] that for TSP the scaling exponent is 3. This is based on analytic methods in a mean-field model of interpoint distances (distances between pairs of points are random, independent for different pairs, thus ignoring geometric constraints) and on Monte Carlo simulations for random points in 2, 3 and 4 dimensional space. The analytic results build upon a recent probabilistic reinterpretation [2] of work of Krauth and Mézard [8] establishing the average length of mean-field TSP tours. But neither part of these TSP assertions is rigorous, and indeed rigorous proofs in $d$ dimensions seem far out of reach of current methodology. In contrast, for the minimum spanning tree (MST) problem, a standard
algorithmically easy problem, a simple heuristic argument repeated below strongly suggests that the scaling exponent is 2 for any reasonable probability model. The purpose of this paper is to think about the details.

Why study such scaling exponents? For a combinatorial optimization problem, a larger exponent means that there are more near-optimal solutions, suggesting that the algorithmic problem of finding the optimal solution is intrinsically harder. So scaling exponents may serve to separate combinatorial optimization problems of an appropriate type into a small set of classes of increasing difficulty. For instance, the minimum matching and minimum Steiner tree problems are expected to have scaling exponent 3, and thus be in the same class as TSP in a quantitative way, as distinct from their qualitative similarity as NP-complete problems under worst-case inputs. In contrast, algorithmically easy problems are expected to have scaling exponent 2, analogously to the “calculus” scaling exponent. Another such algorithmically easy problem (the NK model) will be mentioned in section 3.

1.1 Background

Steele [9] and Yukich [10] give general background concerning combinatorial optimization over random points.

A network is a graph whose edges e have positive real lengths len(e). Let G be a finite connected network. Recall the notion of a spanning tree (ST) T in G. Identifying T as a set of edges, write len(T) = ∑e∈T len(e). A minimal spanning tree (MST) is a ST of minimal length; such a tree always exists but may not be unique. The classical greedy algorithm (Kruskal’s algorithm [6]) for constructing a MST yields two fundamental properties which we record without proof in Lemma 1.

Let G_t be the subnetwork consisting of those edges e of G with len(e) < t. For arbitrary vertices v, w define

\[ \text{perc}(v, w) = \inf\{t : v \text{ and } w \text{ in same component of } G_t \} \]  \hspace{1cm} (1)

For an edge e = (v, w) of G write perc(e) = perc(v, w) ≤ len(e) and also define

\[ \text{perc}^-(v, w) = \inf\{t : v \text{ and } w \text{ in same component of } G_t \setminus \{e\} \} \]  \hspace{1cm} (2)

So perc(e) = min(perc^-(e), len(e)). Define the excess

\[ \text{exc}(e) = \text{len}(e) - \text{perc}(e) \geq 0. \]
Lemma 1. Suppose all the edge-lengths in $G$ are distinct.

(a) There is a unique MST, say $T$, and it is specified by the criterion $e \in T$ if and only if $\text{exc}(e) = 0$.

(b) For any vertices $v, w$

$$\text{perc}(v, w) = \max\{\text{len}(e) : e \text{ on path from } v \text{ to } w \text{ in } T\}.$$

1.2 The heuristic argument

Given a probability model for $n$ random points and their interpoint lengths, define a measure $\mu_n(\cdot)$ on $(0, \infty)$ in terms of the expectation

$$\mu_n(0, x) = \frac{1}{n} E|\{ \text{edges } e : 0 < \text{len}(e) - \text{perc}(e) < x \}|.$$

For any reasonable model with suitable scaling of edge-lengths we expect an $n \to \infty$ limit measure $\mu(\cdot)$, with a density $\nu(x) = d\mu/dx$ having a non-zero limit $\nu(0^+)$ as $x \downarrow 0$.

Now modify the MST by adding an edge $e$ with $\text{len}(e) - \text{perc}(e) = b$, for some small $b$, to create a cycle; then delete the longest edge $e' \neq e$ of that cycle, which necessarily has $\text{len}(e') = \text{perc}(e)$. This gives a spanning tree containing exactly one edge not in the MST and having length greater by $b$. Repeat this procedure with every edge $e$ for which $0 < \text{len}(e) - \text{perc}(e) < \beta$, for some small $\beta$. For large $n$, the number of such edges should be $n\mu_n(0, (\beta) \approx n\nu(0^+)\beta$ to first order in $\beta$, and assuming there is negligible overlap between cycles, each of the new edges will increase the tree length by $\sim \beta/2$ on average. So we expect

$$\delta(\beta) \sim \nu(0^+)\beta, \quad \varepsilon(\beta) \sim \nu(0^+)\beta^2/2.$$

This construction should yield essentially the minimum value of $\varepsilon$ for given $\delta$, so we expect

$$\varepsilon(\delta) \sim \frac{\delta^2}{2\nu(0^+)} \quad (3)$$

and in particular we expect the scaling exponent to be 2.

1.3 Results

Our goal is to formalize the argument above in the context of the following two probability models for $n$ random points. Fix dimension $d \geq 2$ (the case $d = 1$ is of course rather special).
**Model 1** The disordered lattice. Start with the discrete $d$-dimensional cube $C_d^m = [1, 2, \ldots, m]^d$, so there are $n = m^d$ vertices and there are $2d$ edges at each non-boundary vertex. Then take the edge-lengths to be i.i.d. random variables $\xi_e$, whose common distribution $\xi$ has finite mean and some bounded continuous density function $f_\xi(\cdot)$.

**Model 2** Random Euclidean. Take the continuum $d$-dimensional cube $[0, n^{1/d}]^d$ of volume $n$. Put down $n$ independent uniformly distributed random points in this cube. Take the complete graph on these $n$ vertices, with Euclidean distance as edge-lengths.

Each model is set up so that nearest-neighbor distances are order 1 and the MST $T_n$ has mean length of order $n$. To formalize the ideas in the introduction we define the random variable

$$\varepsilon_n(\delta) := \min\left\{\frac{\text{len}(T_n') - \text{len}(T_n)}{n} : |T_n' \setminus T_n| \geq \delta n\right\}$$

where the minimum is over spanning trees $T_n'$ and where $T_n' \setminus T_n$ is the set of edges in $T_n'$ but not in $T_n$.

**Theorem 2** (a) In either model,

$$\limsup_{\delta \downarrow 0} \delta^{-2} \limsup_n E\varepsilon_n(\delta) < \infty.$$

(b) In Model 1, and (assuming Ansatz 7) in Model 2,

$$\liminf_{\delta \downarrow 0} \delta^{-2} \liminf_n E\varepsilon_n(\delta) > 0.$$

### 1.4 Methodology

The heuristic argument suggests that the only model-dependent part of a proof should be the verification that $0 < \nu(0^+) < \infty$ in (3). In a first draft we wrote out arguments in an abstract setting of random size-$n$ networks converging, in the sense of *local weak convergence* [4], to a limit infinite random network. But the overhead in setting up these general abstract structures seemed to obscure rather than illuminate the essential arguments, so we reverted to discussing only the two specific models. In these models the relation in the infinite limit graph between minimal spanning forests (MSFs) and percolation is well understood by experts [5]. Conceptually we could first do calculations in the infinite graphs and then appeal to $n \to \infty$ limit arguments. But we have chosen the less sophisticated approach of doing what calculations we can in the finite models and only appealing to the MSF structure where needed.
2 Proofs

2.1 The upper bound: Model 1 with $d = 2$

We first consider Model 1 with $d = 2$ and then consider the other cases.

The upper bound rests upon a simple construction of near-minimal spanning trees, illustrated in Figure 1.

![Figure 1](image-url)

**Figure 1.** A special configuration on the $3 \times 3$ grid.

The figure illustrates a particular kind of configuration. There is a 4-cycle of edges $abcd$ where, for some $x$,

$$
\text{len}(a) = x, \text{len}(b) \in (x, x + \delta), \text{len}(c) < x, \text{len}(d) < x
$$

and where the eight other edges touching the cycle have lengths $> x + \delta$.

With such a configuration (within a larger configuration on $C_m^2$), edges $adc$ are in the MST, and edge $b$ is not. We can modify the minimal spanning tree by removing edge $d$ and adding edge $b$; this creates a new spanning tree whose extra length equals len$(b) - x$.

Thus given a realization of the edge-lengths on the $m \times m$ discrete square, partition the square into adjacent $3 \times 3$ regions; on each region where the configuration is as in Figure 1, make the modification above. This changes the MST $T_n$ into a certain near-minimal spanning tree $T'_n$. On each $3 \times 3$ square, the probability of seeing the Figure 1 configuration equals

$$
q(\delta) := \int_0^\infty f(x)(F(x + \delta) - F(x))F^2(x)(1 - F(x + \delta))^3 \, dx.
$$

Here $f$ and $F$ are the density and distribution functions of edge-lengths. And the (unconditioned) increase in edge-length of spanning tree caused by
the possible modification equals

\[ r(\delta) := \int_0^\infty f(x) \left( \int_x^{x+\delta} (y-x)f(y)dy \right) F^2(x)(1 - F(x+\delta))^8 \, dx. \]

Letting \( n \to \infty \) with fixed \( \delta \), and using the weak law of large numbers,

\[
\begin{align*}
&n^{-1}|T'_n - T_n| \xrightarrow{P} \frac{1}{9} q(\delta) \quad (5) \\
&n^{-1}(\text{len}(T'_n) - \text{len}(T_n)) \xrightarrow{P} \frac{1}{9} r(\delta). \quad (6)
\end{align*}
\]

Because we defined \( \varepsilon_n(\cdot) \) in terms of spanning trees which differ from the MST by a non-random proportion of edges, we need a detour to handle expectations over events of asymptotically zero probability. We defer the proof.

**Lemma 3** (a) For any sequence \( T^*_n \) of spanning trees, the sequence \( n^{-1}\text{len}(T^*_n) \) is uniformly integrable.

(b) There exist spanning trees \( T''_n \) such that

\[ |T''_n \setminus T_n| \geq a_n \]

where \( a_n/n \to 1/2 \).

Now consider the spanning tree \( T^*_n \) defined to be \( T'_n \) if \( n^{-1}|T'_n - T_n| \geq \frac{1}{10} q(\delta) \) and to be \( T''_n \) if not. It follows from (5,6) and Lemma 3 that

\[
\begin{align*}
&n^{-1}|T^*_n \setminus T_n| \geq \frac{1}{10} q(\delta) \quad \text{(for large } n) \\
&\limsup_n n^{-1} E(\text{len}(T^*_n) - \text{len}(T_n)) \leq \frac{1}{9} r(\delta).
\end{align*}
\]

Then from the definitions of \( q(\delta), r(\delta) \) and the assumption that \( f(\cdot) \) is bounded it is easy to check

\[ q(\delta) \sim c\delta, \quad r(\delta) \sim \frac{1}{2} \delta q(\delta) \text{ as } \delta \downarrow 0 \quad (7) \]

for a certain \( 0 < c < \infty \). This establishes the upper bound (a) in Theorem 2.

**Proof of Lemma 3.** Part (a) is automatic because, writing \( \sum_e \) for the sum over all edges of \( \mathbb{C}_m^2 \), the sequence \( n^{-1}\sum_e \xi_e \) is uniformly integrable. For (b), note that the cube \( \mathbb{C}_m^2 \) with \( 2m(m-1) \) edges can be regarded as a subgraph of the discrete torus \( \mathbb{Z}_m^2 \) with \( 2m^2 \) edges. Take a uniform random spanning tree \( \tilde{T}_n \) on \( \mathbb{Z}_m^2 \), delete edges not in \( \mathbb{C}_m^2 \) and add back boundary edges to make some (non-uniform) random spanning tree \( T_n \) on \( \mathbb{C}_m^2 \). By
symmetry of the torus we have $P(e \in \tilde{T}_n) = \frac{m^2 - 1}{2m^2}$ for each edge $e$ of the torus, and it follows that $P(e \in T_n) = \frac{m^2 - 1}{2m^2}$ for each non-boundary edge of the cube. Since there are $4(m - 1)$ boundary edges and $2(m - 1)(m - 2)$ non-boundary edges, for any spanning tree $t$ we have

$$E|T_n \cap t| \leq 4(m - 1) + (n - 1)(m^2 - 1)/(2m^2) = 4(n^{1/2} - 1) + (n - 1)^2/(2m^2).$$

So

$$E|T_n \setminus t| = (n - 1) - E|T_n \cap t| \geq a_n := (n - 1) - 4(n^{1/2} - 1) - (n - 1)^2/(2m^2).$$

So for any spanning tree $t$ there exists some spanning tree $t^*$ such that $|t^* \setminus t| \geq a_n$. Applying this fact to the MST gives (b).

2.2 Upper bound: other cases

The argument for Model 1 in the case $d \geq 3$ involves only very minor modifications of the proof above, so we turn to Model 2 with $d = 2$ (the case $d \geq 3$ is similar). Here it is natural to consider a different notion of special configuration.

![Figure 2. A special configuration on the 3 \times 3 square.](image)
Here there is a $3 \times 3$ square containing a concentric $1 \times 1$ square. There are three points within the larger square, all being inside the smaller square. In the triangle $abc$ formed by the three points, writing $x$ for the length of the second longest edge length, the length of the longest edge in $\in (x, x + \delta)$ and $x + \delta < 1$. For such a configuration (within a configuration on a $m \times m$ square containing the $3 \times 3$ square), edges $ac$ are in the MST, and edge $b$ is not. We can modify the minimal spanning tree by removing edge $a$ and adding edge $b$; this creates a new spanning tree whose extra length equals $\text{len}(b) - x$.

We now repeat the argument from the previous section, and the overall logic is the same. One gets different formulas for $q(\delta), r(\delta)$ but they have the same relationship (7). The weak law (5,6) is easily established. The only non-trivial difference is that we need to replace the technical Lemma 3 by the following technical lemma.

**Lemma 4** (a) There exists $c_1$ such that for any $n$ and any configuration on $n$ points in the square of area $n$, the MST $\hat{T}_n$ has $\text{len}(\hat{T}_n) \leq c_1 n$.

(b) For sufficiently large $n$, there exist spanning trees $T''_n$ such that $\text{len}(T''_n) \leq 12c_1 n$ and

\[
\frac{n^{-1}|T''_n \setminus T_n|}{n} \geq \frac{1}{2}.
\]

**Proof.** Part (a) follows from the analogous result for TSP – see [9] inequality (2.14). For (b), let $\xi_1, \ldots, \xi_n$ be the positions of the $n$ random points and recall that $T_n$ is their MST. Classify these points $\xi_i$ as “odd” or “even” according to whether the number of edges in the path inside $T_n$ from $\xi_i$ to $\xi_1$ is odd or even. Let $(\hat{\xi}_i)$ be a configuration obtained from $(\xi_i)$ by moving each “odd” point a distance $11c_1$ in some arbitrary direction. Let $\hat{T}_n$ be the MST on $(\hat{\xi}_i)$. Let $T''_n$ be the spanning tree on $(\xi_i)$ defined by

\[
(\xi_i, \xi_j) \in T''_n \text{ iff } (\hat{\xi}_i, \hat{\xi}_j) \in \hat{T}_n.
\]

Suppose $(\xi_i, \xi_j)$ is an edge of both $T_n$ and $T''_n$. Since one end-vertex is odd and the other is even, it is easy to see:

- either (i) $\text{len}(\xi_i, \xi_j) \geq 5c_1$; or (ii) $\text{len}(\hat{\xi}_i, \hat{\xi}_j) \geq 5c_1$.

But by part (a) there are at most $n/5$ edges satisfying (i), and similarly for (ii). So $|T_n \cap T''_n| \leq 2n/5$. Noting that

\[
\text{len}(T''_n) \leq 11c_1(n - 1) + \text{len}(\hat{T}_n) \leq 12c_1 n
\]

using (a), we have established (b).
2.3 The lower bound: a discrete lemma

The lower bound argument rests upon the following simple lemma.

**Lemma 5** Consider a finite connected network with distinct edge-lengths. If \( T \) is the MST and \( T' \) is any ST then
\[
\text{len}(T') - \text{len}(T) \geq \sum_{e' \in T' \setminus T} \text{exc}(e').
\]

**Proof.** Suppose \(|T' \setminus T| = k \geq 1\). It is enough to show that there exist \( e' \in T' \setminus T \) and \( e \in T \setminus T' \) such that
(i) \( T^* = T' \setminus \{e'\} \cup \{e\} \) is a ST;
(ii) \( |T^* \setminus T| = k - 1 \);
(iii) \( \text{len}(e') - \text{len}(e) \geq \text{exc}(e') \)
for then we can continue inductively.

To prove this we first choose an arbitrary \( e' \in T' \setminus T \). Consider \( T' \setminus \{e'\} \). This is a two-component forest; so the path in \( T \) linking the end-vertices of \( e' \) must contain some edge \( e \in T \setminus T' \) which links these two components. So choose some such edge \( e \). Properties (i) and (ii) are clear. Apply Lemma 3 to the end-vertices of \( e' \) and to \( T \) to see that \( \text{perc}(e') \geq \text{len}(e') \). So
\[
\text{len}(e') - \text{len}(e) \geq \text{len}(e') - \text{perc}(e') = \text{exc}(e')
\]
which is (iii).

Let us also record the following straightforward lemma.

**Lemma 6** Let \( \xi \) and \( W \) be independent real-valued random variables such that \( \xi \) has a density function bounded by a constant \( \bar{f} \). Then for any event \( A \subseteq \{\xi > W\} \) we have
\[
E(\xi - W)1_A \geq \frac{P^2(A)}{2\bar{f}}.
\]

2.4 The lower bound in Model 1

We treat the case \( d = 2 \), but \( d \geq 3 \) involves only minor changes. Recall \( \mathbb{C}_m^2 \) has \( c_n := 2(n - n^{1/2}) \) edges. Fix \( \delta > 0 \). Consider a pair \((T'_n, T_n)\) attaining the minimum in the definition (4) of \( \varepsilon_n(\delta) \). For a uniform random edge \( e_n \) of \( \mathbb{C}_m^2 \),
\[
P(e_n \in T'_n \setminus T_n) = \frac{E|T'_n \setminus T_n|}{c_n} \geq \frac{\delta}{2}
\]
(8)
and
\begin{align*}
E_{\varepsilon_n}(\delta) &= \frac{1}{n}E(\text{len}(T_n') - \text{len}(T_n)) \\
&\geq \frac{1}{n}E \sum_{e\in T_n'\setminus T_n} \text{exc}(e) \quad \text{by Lemma 5} \\
&= \frac{c_n}{n}E\text{exc}(e_n)1_{e_n\in T_n'\setminus T_n} \\
&\geq \frac{\delta}{2f} \quad \text{by (8)}
\end{align*}

For a fixed edge $e$ of $\mathbb{C}_m^2$, we can write
\[ \text{exc}(e) = (\xi^{(n)}(e) - W^{(n)}(e))^+ \]
where $\xi^{(n)}(e)$ is the edge-length of $e = (v, v^*)$ and where $W^{(n)}(e) = \text{perc}^{-}(e)$ as defined at (2). Note (and this is the key special feature that makes Model 1 easy to study) that $\xi^{(n)}(e)$ and $W^{(n)}(e)$ are independent. Since $\text{exc}(e) > 0$ on $\{e \in T_n' \setminus T_n\}$ we see that the quantity at (9) is of the form appearing in Lemma 6. So
\begin{align*}
\frac{n}{c_n}E_{\varepsilon_n}(\delta) &\geq E \left( \xi^{(n)}(e_n) - W^{(n)}(e_n) \right) 1_{e_n\in T_n'\setminus T_n} \quad \text{by (9)} \\
&\geq \frac{P^2(e_n \in T_n' \setminus T_n)}{2f} \quad \text{by Lemma 6} \\
&\geq \frac{\delta^2}{8f} \quad \text{by (8)}
\end{align*}
where $\bar{f}$ is the bound on the density of $\xi$. Because $c_n \sim 2n$ we have established part (b) of Theorem 2 in this case.

### 2.5 Minimum spanning forests

My purpose here is to state the Ansatz clearly; I don’t know whether it is "really hard" or not. The rest of the argument (next section) is only outlined, but I think it is really "not hard" to fill in the details.

One could rephrase the argument above in terms of the minimum spanning forest (MSF) on the infinite lattice $\mathbb{Z}^2$ with i.i.d. edge-lengths $\xi_e$. To discuss Model 2 it seems helpful to use the MSF explicitly, and to exploit theory [3, 4] of local weak convergence of spanning trees to the MSF. Here is a general definition, in the context of a countable-vertex network $G$ with distinct edge-lengths (see [5] for more detailed treatment). As in section 1.1 let $G_t$ be the subnetwork consisting of those edges $e$ of $G$ with $\text{len}(e) < t$. Define the MSF by:
an edge \((v, w)\) is in the MSF iff, for \(t = \text{len}(v, w)\), vertices \(v\) and \(w\) are in different components of \(G_t\) and at least one of these components is finite.

Now consider a Poisson point process \((\eta_i)\) of rate 1 in \(\mathbb{R}^d\), where \(d \geq 2\). Add an extra point \(0\) at the origin. Consider \(\{(\eta_i) \cup \{0\}\}\) as the vertices of a network (the complete graph with Euclidean edge-lengths). Now define \(\text{perc}(0, \eta_i)\) as the infimum of \(t\) such that:

- either \(0\) and \(\eta_i\) are in the same component of \(G_t\), or they are in different infinite components.

Define a measure \(\mu\) on \((0, \infty)\) by

\[
\mu(0, x) = E \sum_i 1_{(0 < \text{len}(0, \eta_i) - \text{perc}(0, \eta_i) < x)}.
\]

**Ansatz 7**

\[
\limsup_{x \downarrow 0} \frac{\mu(0, x)}{x} < \infty.
\]

This is of course a formalization on the heuristic idea \(\nu(0^+) < \infty\) from section 1.2. Analogous to Lemma 6 we state a straightforward lemma which describes how the Ansatz is used.

**Lemma 8** Let \((V_i, i \geq 1)\) be real-valued r.v.’s such that \(\mu(0, x) := E \sum_i 1_{(0 < V_i < x)}\) satisfies \(\limsup_{x \downarrow 0} \frac{\mu(0, x)}{x} < \infty\). Then there exists a function \(g(s) \sim \beta s^2\) as \(s \downarrow 0\), for some \(\beta > 0\), such that for any sequence of events \(A_i \subseteq \{V_i > 0\}\),

\[
E \sum_i V_i 1_{A_i} \geq g \left( \sum_i P(A_i) \right).
\]

### 2.6 The lower bound in Model 2

We start by copying and modifying the argument from section 2.4. Fix \(\delta > 0\). Let \(\xi_{U_n}\) be a uniform random vertex from \((\xi_i, 1 \leq i \leq n)\). Consider a pair \((T'_n, T_n)\) attaining the minimum in the definition (4) of \(\varepsilon_n(\delta)\). Then

\[
E \sum_i 1_{((\xi_{U_n}, \xi_i) \in T'_n \setminus T_n)} = \frac{2E|T'_n \setminus T_n|}{n} \geq 2\delta
\]
and
\[
E \varepsilon_n(\delta) = \frac{1}{n} E(\text{len}(T'_n) - \text{len}(T_n)) \\
\geq \frac{1}{n} E \sum_{e \in T'_n \setminus T_n} \text{exc}(e) \text{ by Lemma 5} \\
= \frac{1}{2} E \sum_i \text{exc}(\xi_{U_n}, \xi_i) 1(\langle \xi_{U_n}, \xi_i \rangle \in T'_n \setminus T_n)
\] (11)

Note that for \(0 < L < \infty\)
\[
E \sum_i 1(\text{len}(\xi_{U_n}, \xi_i) \geq L, \langle \xi_{U_n}, \xi_i \rangle \in T'_n) = \frac{2}{n} E|\{e \in T'_n : \text{len}(e) \geq L\}|
\leq \frac{2}{n} \frac{E \text{len}(T'_n)}{L}
\leq \frac{\delta}{2} \text{ for } L = L(\delta) \text{ sufficiently large}
\]

the last inequality because \(E \text{len}(T'_n) = O(n)\). So fixing such an \(L\), (10) implies
\[
E \sum_i 1(\langle \xi_{U_n}, \xi_i \rangle \in T'_n \setminus T_n, \text{len}(\xi_{U_n}, \xi_i) \leq L) \geq \delta
\] (12)

while (11) trivially implies
\[
2E \varepsilon_n(\delta) \geq E \sum_i \text{exc}(\xi_{U_n}, \xi_i) 1(\langle \xi_{U_n}, \xi_i \rangle \in T'_n \setminus T_n, \text{len}(\xi_{U_n}, \xi_i) \leq L).
\] (13)

The purpose of these representations is to exploit local weak convergence. For each \(n\) consider the \(n\) random points recentered at a uniform random point:
\[
\tilde{\xi}^{(n)}_{U_i} = \xi_i - \xi_{U_n}, \quad 1 \leq i \leq n.
\]

It is classical that the point process \((\tilde{\xi}^{(n)}_i)\) converges in distribution as \(n \to \infty\) to the point process \((\tilde{\eta}_i) = (\eta_i) \cup \{0\}\). Here convergence means convergence with respect to finite windows. That is, for arbitrary fixed \(L\) we consider only points within the cube \(C_L := [-L, L]^d\). As shown in Proposition 9 of [3], along with convergence of points we have convergence of the MST \(T_n\) to the MSF \(\mathcal{F}_{\infty}\) of \((\tilde{\eta}_i)\). (3 treated the case \(d = 2\) but \(d \geq 3\) is similar).

Here convergence of graphs is formalized via convergence of \(\{0, 1\}\)-valued labels attached to point-pairs. Moreover the arguments of [3] extend to show convergence of the perc\(^-\) functional. So symbolically
\[
\left(\tilde{\xi}^{(n)}_{i}, 1(\langle \tilde{\xi}^{(n)}_{i}, \tilde{\xi}^{(n)}_{j} \rangle \in T_n), \text{perc}_n(\tilde{\xi}^{(n)}_{i}, \tilde{\xi}^{(n)}_{j})\right)_{\tilde{\xi}^{(n)}_{i}, \tilde{\xi}^{(n)}_{j} \in C_L}
\]
For the near-minimal STs $T'_n$ appearing in (12,13), by a compactness argument and by passing to a subsequence of $n$ we may they converge to some forest $F'_\infty$ on $(\tilde{\eta}_i)$; that is, we may assume that (14) remains true when we append $1_{((\tilde{\xi}^{(n)}_i, \tilde{\xi}^{(n)}_j)) \in T'_n}$ to the left side and $1_{((\tilde{\eta}_i, \tilde{\eta}_j)) \in F'_\infty}$ to the right side. We can now take limits in (12) to deduce

$$\sum_i P((0, \eta_i) \in F'_\infty \setminus F_\infty, \ len(0, \eta_i) \leq L) \geq \delta.$$ 

And taking limits in (13) gives

$$2 \liminf_n E_{\varepsilon_n(\delta)} \geq E \sum_i (\text{len}(0, \eta_i) - \text{perc}^{-}(0, \eta_i)) 1_{((0, \eta_i) \in F'_\infty \setminus F_\infty, \ len(0, \eta_i) \leq L)}.$$ 

Writing

$$V_i = \text{len}(0, \eta_i) - \text{perc}^{-}(0, \eta_i)$$
$$A_i = \{(0, \eta_i) \in F'_\infty \setminus F_\infty, \ len(0, \eta_i) \leq L\},$$

we are precisely in the setting in which Ansatz 7 and Lemma 8 apply, and the conclusion is that the right side of (15) is $\geq (\beta - o(1))\delta^2$ for small $\delta$, implying the lower bound in Theorem 2.
References


3 The NK model

This is intended to attract the interest of readers of drafts; will be compressed or eliminated in any final version.

Another example of an “algorithmically easy” problem whose scaling exponent should be 2 is the Kauffman NK model. For our version of the model\(^1\) we fix \(K \geq 2\). We seek to minimize, over binary sequences \(x = (x_1, \ldots, x_N)\), the objective function

\[
H_N(x) = \sum_{i=1}^{N-K} W_i(x_i, x_{i+1}, \ldots, x_{i+K}),
\]

where the values \((W_i(b_0, b_1, \ldots, b_K) : i \geq 1, b \in \{0, 1\}^{K+1})\) are independent exponential(1) random variables. Write \(x^N\) for the minimizing sequence. By subadditivity there is a limit

\[
N^{-1}EH_N(x^N) \to c_K.
\]

To discuss scaling, for a general sequence \(y = y^N\) write

\[
\delta_N(y) = N^{-1}\#\{1 \leq i \leq N - K : (y_i, \ldots, y_{i+K}) \neq (x^N_i, \ldots, x^N_{i+K})\}
\]

\[
\varepsilon_N(y) = N^{-1}(H_N(y) - H_N(x^N))
\]

and then set

\[
\varepsilon_N(\delta) = \min\{\varepsilon_N(y) : \delta_N(y) \geq \delta\}
\]

\[
\varepsilon(\delta) = \lim_{N \to \infty} \varepsilon_N(\delta).
\]

Here is a very crude argument for \(\varepsilon(\delta) \propto \delta^2\). Given \(i\) and \(l \geq K+1\), consider the set of sequences \(y\) such that

\[
(y_j, \ldots, y_{j+K}) = (x^N_j, \ldots, x^N_{j+K}) \quad \forall j \notin [i + 1, i + l]
\]

\[
(y_j, \ldots, y_{j+K}) \neq (x^N_j, \ldots, x^N_{j+K}) \quad \forall j \in [i + 1, i + l]
\]

Over this set, let \(D_{i,l}\) be the minimum of \(H_N(y) - H_N(x^N)\) and let \(y^{(i,l)}\) be the minimizing sequence. The distribution of \(D_{i,l}\) essentially depends only on \(l\), not on \(i\) or \(N\); write \(f_l(0+)\) for its density at 0+. Let’s assume as an Ansatz

\[
\sum_{l \geq K+1} l^2 f_l(0+) = A < \infty. \quad (16)
\]

It is intuitively clear how to choose a sequence \(y\) which minimizes \(\varepsilon_N(y)\) for a given \(\delta\). Just fix a small \(\eta > 0\), and create a sequence of “excursions” away from \(x^N\) as follows. For each pair \((i, l)\) such that \(D_{i,l} < \eta l\), choose \(y\)

\(^1\) For compatibility with our other examples we treat this as a minimization problem, though it is usually taken as a maximization problem.
to equal $y^{(i,l)}$ on the sites $[i + K + 1, i + l]$; set $y = x^N$ elsewhere. With this scheme, $\delta$ will be the mean length of such an excursion starting from a given site, that is

$$\delta = \sum_{l \geq K+1} l \cdot \eta_l f_l(0+).$$

And $\epsilon$ is the mean increment of $H_N$ associated with excursions starting from a given site, that is

$$\epsilon = \sum_{l \geq K+1} (\eta l/2) \cdot \eta_l f_l(0+).$$

In other words $\delta \sim A\eta$, $\epsilon \sim A\eta^2/2$, giving $\epsilon \sim (2A)^{-1}\delta^2$.

Why should the Ansatz be true? Well, by subadditivity arguments we expect central limit behavior: $D_l \approx \text{Normal}(\mu_l, \sigma^2_l)$ for $\mu > 0$. This in turn suggests that $f_l(0+)$ should decrease geometrically fast in $l$.

Simulation with $K = 3$ confirms both the scaling exponent 2 and the prediction, implicit in the argument above, that as $\delta \to 0$ the near-optimal $y$ should have some limit distribution $L$ of excursion lengths with $P(L = l) \propto l f_l(0+)$ and hence with $EL < \infty$.

<table>
<thead>
<tr>
<th>$\lambda$</th>
<th>$\delta$</th>
<th>$\epsilon$</th>
<th>$\epsilon/\delta^2$</th>
<th>EL</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.002</td>
<td>0.0397</td>
<td>4.85 \cdot 10^{-5}</td>
<td>0.0308</td>
<td>10.9</td>
</tr>
<tr>
<td>0.004</td>
<td>0.0774</td>
<td>2.00 \cdot 10^{-4}</td>
<td>0.0334</td>
<td>11.0</td>
</tr>
<tr>
<td>0.008</td>
<td>0.147</td>
<td>7.69 \cdot 10^{-4}</td>
<td>0.0354</td>
<td>11.3</td>
</tr>
<tr>
<td>0.016</td>
<td>0.266</td>
<td>2.75 \cdot 10^{-3}</td>
<td>0.0388</td>
<td>11.8</td>
</tr>
</tbody>
</table>

**Table 1.** Monte Carlo simulations with $K = 3, N = 10,000$; 1000 repeats. These are exact optimizations done by introducing a Lagrange multiplier $\lambda$ which penalizes matching $(K + 1)$-tuples. We find $c_3 = 0.3065$. 

---

$^2$overlaps will be negligible