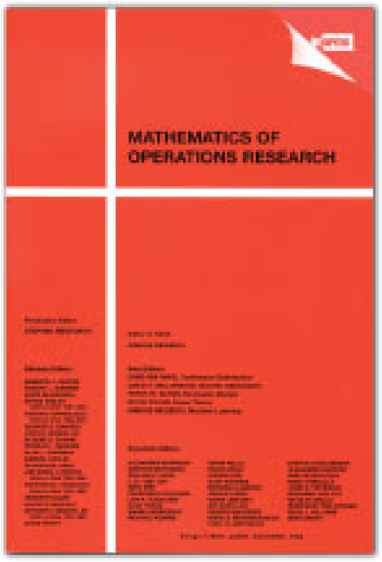

Mathematics of Operations Research

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Euclidean Networks with a Backbone and a Limit Theorem for Minimum Spanning Caterpillars

Petar Jevtić, J. Michael Steele

MATHEMATICS OF OPERATIONS RESEARCH
Vol. 40, No. 4, November 2015, pp. 992–1004
ISSN 0364-765X (print) | ISSN 1526-5471 (online)
http://dx.doi.org/10.1287/moor.2014.0706

Euclidean Networks with a Backbone and a Limit Theorem for Minimum Spanning Caterpillars

Petar Jevtić
Department of Mathematics and Statistics, McMaster University, Hamilton, Ontario L8S 4K1, Canada,
pjevtic@math.mcmaster.ca

J. Michael Steele
Wharton School, Department of Statistics, University of Pennsylvania, Philadelphia, Pennsylvania 19104,
steele@wharton.upenn.edu

A caterpillar network (or graph) G is a tree with the property that removal of the leaf edges of G leaves one with a path. Here we focus on minimum weight spanning caterpillars where the vertices are points in the Euclidean plane and the costs of the path edges and the leaf edges are multiples of their corresponding Euclidean lengths. The flexibility in choosing the weight for path edges versus the weight for leaf edges gives some useful flexibility in modeling. In particular, one can accommodate problems motivated by communications theory such as the "last mile problem." Geometric and probabilistic inequalities are developed that lead to a limit theorem that is analogous to the well-known Beardwood, Halton, and Hammersley theorem for the length of the shortest tour through a random sample, but the minimal spanning caterpillars fall outside the scope of the theory of subadditive Euclidean functionals.

Keywords: Euclidean networks; subadditive Euclidean functional; caterpillar graphs; shortest paths; traveling salesman problem; Beardwood, Halton, and Hammersley theorem; minimal spanning trees; Gutman graphs
MSC2000 subject classification: Primary: 60C05, 90C40; secondary: 60G42, 90C27, 90C39
OR/MS subject classification: Primary: networks/graphs; secondary: generalized networks, stochastic, theory
History: Received May 5, 2013; revised May 24, 2014. Published online in *Articles in Advance* June 17, 2015.

1. Introduction. To visualize a caterpillar graph, just draw a long path graph and then add a liberal sprinkling of additional vertices where each added vertex is connected to the path by a single edge. Such graphs (or networks) have a natural place in communication models where the path typically represents a fast backbone and the edges that come off the path represent local service connections.

Given a set $\chi = \{x_1, x_2, \ldots, x_n\}$ of n points in $\mathbb{R}^2$, we say that a graph G is a *spanning caterpillar* for χ if G is a caterpillar graph with vertex set χ. More formally, a spanning caterpillar G is determined by a triple $G = (\chi, E, \pi)$, with vertex set χ, edge set E, and a designated path graph π that is a subgraph of G. The graph G is connected and each vertex of G that is not a vertex of π is required to have degree one.

The main focus here is on *weighted spanning caterpillars*, where we differentiate between the costs of edges that are on the designated path π and those edges of G that are not on the path π. For each edge e in the edge set of G, we let $|e|$ denote its Euclidean length; that is, if $e = (x, y) \in E = E(G)$ then we have $|e| = |x - y|$. Now, given a fixed real value $\lambda > 0$, we define the weight $W(G)$ of the spanning caterpillar $G = (\chi, E, \pi)$ to be the sum over G of the weighted edge lengths:

$$W(G) = \lambda \sum_{e \in \pi} |e| + \sum_{e \notin \pi} |e|. \tag{1}$$

One motivation for this weighting scheme is the infamous "last mile" problem of communication network theory. In that context, the weight factor λ for the path edges would be smaller than one—possibly *much* smaller since communication along a network backbone may be very fast. Nevertheless, there is no mathematical reason to restrict the value of λ beyond requiring it to be positive; moreover, there are benefits to being flexible about the size of λ. For example, in a ground transportation model where the "drop off" cost is cheap, one would want to take λ to be larger than one.

Here we are concerned with the cost of the minimum weight spanning caterpillar under two distinct analytical situations. First, there are worst-case scenarios where the points are placed deterministically subject only to geometric constraints. Second, one wants to understand the more generic situations such as those with $\boldsymbol{\chi}_n = \{X_1, X_2, \ldots, X_n\}$, where the points X_i, $1 \le i \le n$ are independent and identically distributed random variables with values in $\mathbb{R}^2$. In this scenario, we let $\mathscr{C}(\boldsymbol{\chi}_n)$ denote the set of all spanning caterpillars of $\boldsymbol{\chi}_n$ and the random variable of primary interest is the weight of the *minimum spanning caterpillar* (MSCs):

$$M(\boldsymbol{\chi}_n) = M(X_1, X_2, \ldots, X_n) \stackrel{\text{def}}{=} \min\{W(G)\colon G \in \mathscr{C}(\boldsymbol{\chi}_n)\}.$$

Our main theorem is a strong law for $M(X_1, X_2, \ldots, X_n)$ that is of a kind that goes back to the limit theorem of Beardwood et al. [4] for the traveling salesman problem.

Theorem 1.1 (Strong Law for MSCs of Random Samples). *If the random points X_i, $i = 1, 2, \ldots$ are chosen independently and uniformly from the unit square, then there is a constant $\beta_{\mathrm{MSC}}(\lambda) > 0$ such that with probability one we have*

$$\lim_{n\to\infty} n^{-1/2} M(X_1, X_2, \ldots, X_n) = \beta_{\mathrm{MSC}}(\lambda). \tag{2}$$

More generally, if the independent random variables X_i, $i = 1, 2, \ldots$ have a density f on $\mathbb{R}^2$ with compact support, then we have with probability one that

$$\lim_{n\to\infty} n^{-1/2} M(X_1, X_2, \ldots, X_n) = \beta_{\mathrm{MSC}}(\lambda) \int_{\mathbb{R}^2} \sqrt{f(x)}\, dx, \tag{3}$$

where the constant $\beta_{\mathrm{MSC}}(\lambda)$ in (3) is the same as in (2).

Small values of λ favor path edges over leaf edges, so it is natural to ask if Theorem 1.1 might actually be a generalization of the Beardwood, Halton, and Hammersley theorem. As the theorem is framed and proved, it does not rigorously include the theorem. Still, as we outline in §8, one can give a theorem that covers both the behavior of minimum spanning caterpillars and the minimum cost traveling salesman paths.

The Beardwood, Halton, and Hammersley theorem and its extensions have an extensive literature (see, e.g., the monographs of Steele [31] and Yukich [33]), but, for several reason, the theory of the minimum spanning caterpillar falls outside of the scope of that literature. First, the MSC functional is not monotone, so it fails to be a subadditive Euclidean functional in the sense of Steele [28]. A corresponding lack of monotonicity was addressed for the minimal spanning tree in Steele [30] and for the minimal matching problem in Rhee [27], but the methods used there run into trouble here. One source of difficulty is that the vertex degrees of a minimum spanning caterpillar can be arbitrarily large. These and other distinctions are discussed as the proof of the theorem progresses (see, e.g., the remarks at the end of §§2 and 3 and the discussion of the minimal spanning tree (MST) in §8).

We begin the proof by developing the most basic geometric features of the MSC functional for general finite sets of points. In particular, Lemma 2.2 gives us crucial control of the number of edges incident to a vertex—provided that we constrain the lengths of those edges. Probability enters for the first time in §3, where we use concentration inequalities to show that with high probability the backbone path π will visit every element of a certain partition of $[0, 1]^2$ into subsquares. Sections 4, 5, and 6 complete the proof of Theorem 1.1.

Several relationships between the MSC, the traveling salesman problem (TSP), and the MST are then detailed in §§7, and 8 concludes with a brief discussion of extensions and refinements of the MSC limit theorem.

2. Geometric features of minimum spanning caterpillars. Several of our inferences about the structure of a minimal spanning caterpillar depend on estimates of the weight of a suboptimal spanning caterpillar. Some of these depend in turn on a classic bound for the length of the shortest path through a set of points in a square.

Lemma 2.1 (Short Path Bound). *For any $\{y_1, y_2, \ldots, y_m\} \subset [0, t]^2$, $t > 0$, there is a permutation $\sigma\colon \{1, 2, \ldots, m\} \to \{1, 2, \ldots, m\}$ such that*

$$\sum_{i=1}^{m-1} |y_{\sigma(i)} - y_{\sigma(i+1)}| \le 3t\sqrt{m}.$$

In other words, given a set of points $\{y_1, y_2, \ldots, y_m\}$ in a square of side t, we can always find a path through the points that is not longer than $3t\sqrt{m}$. Results of Few [11] are more precise (and still easily proved). For our purposes here, any explicit $O(t\sqrt{m})$ bound would suffice.

One of the challenging features of the minimum spanning caterpillar problem is that the minimal cost can go up or down as one adds points. For example, if $\chi = \{x_1, x_2, x_3, x_4\}$ is the set of corner points of the square $[0, 1]^2$, then with $\lambda = 1$ we have $M(\chi) = 3$, but if $x_5 = (2^{-1/2}, 2^{-1/2})$ and $\chi' = \chi \cup \{x_5\}$ then $M(\chi') = 2^{3/2} < 3$. Thus, as mentioned earlier, $M(\cdot)$ fails to be monotone so it is not a subadditive Euclidean functional in the sense of Steele [28] and [31].

The tools of this section help us to deal with this lack of monotonicity and several related geometric difficulties. The next lemma is the most critical of these, and it gives us a kind of local finiteness without which progress would be difficult. Here, for any graph G and any vertex y of G we let $N_G(y)$ be the set of the neighbors of y in G. Also, for any $y \in \mathbb{R}^2$ we have an associated family of annuli,

$$A(y, r) = \{x \in \mathbb{R}^2 \colon r/2 \le |x - y| \le r\} \quad 0 \le r < \infty.$$

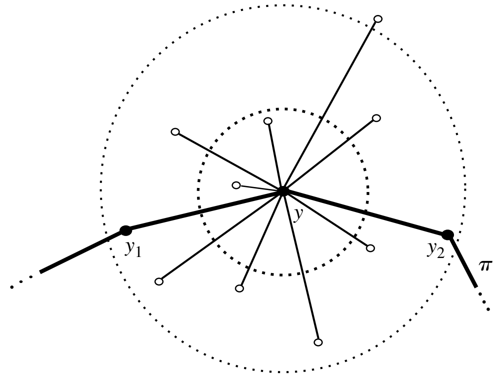

FIGURE 1. The annulus with the inner radius $r/2$ and the outer radius r centered at the path point y. Open dots denote nonpath points of G and the heavy line indicates the path π.

LEMMA 2.2 (NO CROWDED ANNULUS). *There exists a constant $\alpha = \alpha(\lambda) > 0$ such that for any set $\chi = \{x_1, x_2, \ldots, x_n\} \subset [0,1]^2$ and for any minimum spanning caterpillar $G = (\chi, E, \pi)$ we have for all $y \in \chi$ and all $r > 0$ that*

$$|N_G(y) \cap A(y,r)| \le \alpha.$$

PROOF. If y is a star point of G the assertion is trivial since y has just one neighbor. Hence we assume that $y \in \pi$ and—for the moment—we further assume that y is an interior vertex of π with neighbors y_1 and y_2 on π as shown in Figure 1. We now let $m = |N_G(y) \cap A(y,r)|$ and we assume without loss of generality that $m \ge 4$. We will now construct a new spanning graph G' of χ, as shown in Figure 2. The suboptimality of G' is used to get a bound on m.

First, delete all of the edges from y to the star points S of G in $N_G(y) \cap A(y,r)$. Thus, we delete at least $m-2$ edges, and the total cost of these edges is at least $(m-2)r/2$. Next we apply Lemma 2.1 to get a path π_0 through the points of S such that the Euclidean length of π_0 is not greater than $6rm^{1/2}$; here we use the observation that the annulus is contained in a box with side $2r$ and m is an upper bound on the number of points in S. We let z_1 and z_2 be the endpoints of the path π_0. These are distinct by our assumption that $m \ge 4$.

To complete the construction, we add the edge (y, z_1), delete the edge (y, y_2), and insert the edge (z_2, y_2). Consequently, for the path π' for $G' = (\chi, E', \pi')$ we can take the segment of π up to y, the edge (y, z_1), the path π_0 through S from z_1 to z_2, the edge (z_2, y_2), and then finally we take the remainder of the original path π that follows y_2.

By our construction we have

$$W(G') \le W(G) - (m-2)r/2 + 6\lambda r m^{1/2} + \lambda|y - z_1| - \lambda|y - y_2| + \lambda|z_2 - y_2|.$$

By the triangle inequality we have $|z_2 - y_2| - |y - y_2| \le r$ and trivially $|y - z_1| \le r$, so from $W(G) \le W(G')$ we have

$$m \le 12\lambda m^{1/2} + 4\lambda + 2.$$

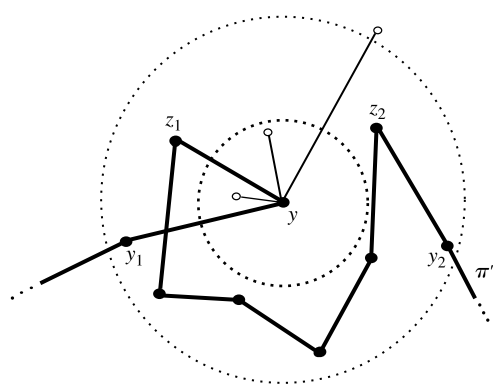

FIGURE 2. A view of the annulus of Figure 1 after surgery. In the new caterpillar G', all points of $\boldsymbol{\chi}$ that are in the annulus are now on the new path π'.

Therefore, for the case when y is an interior point of π we can take the generous bound $m \leq (14\lambda+2)^2$. The case when y is an end point of π is completely analogous—even a bit easier, so we omit the details for that case.

A basic consequence of the "no crowded annulus" lemma is that a vertex v of a MSC with a large number of neighbors must have some neighbor that is very close to v; in fact, it must be exponentially close.

Lemma 2.3 (Existence of an Exponentially Near Neighbor). *There exist constants $C>0$ and $0<\rho<1$ depending only on λ such that for any $\chi=\{x_1,x_2,\dots,x_n\}$, any spanning caterpillar $G=(\chi,E,\pi)$, $y_0\in\chi$, and $R\geq 0$, we have*

$$\min\{|y-y_0|\colon y\in N_G(y_0),\ |y-y_0|\leq R\}\leq C\rho^q R,$$

where $q=|N_G(y_0)\cap\{y\colon |y-y_0|\leq R\}|$.

Proof. The infinite set of annuli $A(y_0,R), A(y_0,R2^{-1}),\dots,A(y_0,R2^{-k}),\dots$ covers the punctured disk $\{y\colon 0<|y-y_0|\leq R\}$, and by Lemma 2.2 none of these annuli can contain more than α points of $N_G(y_0)$. Let k be the maximal integer for which one has $q\geq k\alpha$. One of the annuli $A(y_0,R2^{-j})$ with $j\geq k$ must then contain a point of $N_G(y_0)$; that is,

$$q\geq k\alpha \quad\text{implies}\quad \min\{|y-y_0|\colon y\in N_G(y_0)\}\leq 2^{-k}R,$$

and this is more than one needs for the lemma. Moreover, by review of the proof, one can check that $C=2$ and $\rho=2^{-1/\alpha}$ would suffice here.

Several of our arguments depend on decompositions of the unit square into subsquares, and the next lemma is typical. Here, by $\mathscr{B}(k)$ we denote the collection of all of the k^2 subsquares of $[0,1]^2$ that have the form

$$[a/k,(a+1)/k]\times[b/k,(b+1)/k],\quad \text{where } 0\leq a,b<k.$$

Lemma 2.4 (Cost to Drop One). *There is a constant $C=C(\lambda)$ such that for all $\chi=\{x_1,x_2,\dots,x_n\}$ and all minimal spanning caterpillars $G=(\chi,E,\pi)$, we have*

$$M(x_1,x_2,\dots,x_{n-1})\leq M(x_1,x_2,\dots,x_n)+C/k, \tag{4}$$

provided that every $B\in\mathscr{B}(k)$ contains a vertex of the path π.

Proof. We first set $\chi'=\{x_1,x_2,\dots,x_{n-1}\}$. If there is a minimum spanning caterpillar $G=(\chi,E,\pi)$ of χ where the point x_n is not a vertex of the path $\pi=\pi(G)$, then we can simply delete x_n and the edge incident to x_n to get a minimum spanning caterpillar of χ' that has weight less than $M(x_1,x_2,\dots,x_n)$. Similarly, if $x_n\in\pi$ but x_n has degree one in G, then we can just delete x_n and its edge to get a spanning caterpillar that has weight less than $M(x_1,x_2,\dots,x_n)$. Also, if x_n is on π and has degree equal to two, then we can delete x_n and the edges incident to x_n and add an edge connecting the neighbors of x_n on π. In all of these easy cases we get a spanning caterpillar for χ' that has weight less than $M(\chi)$.

Thus, we may assume that x_n is a vertex of π and that x_n has at least one neighbor in G that is not on the path π. As before, we have two cases to consider: (a) when x_n is an end point of the path π and (b) when x_n is an interior point of π as shown in Figure 3. The cases are similar, so we will only give the details for the second case.

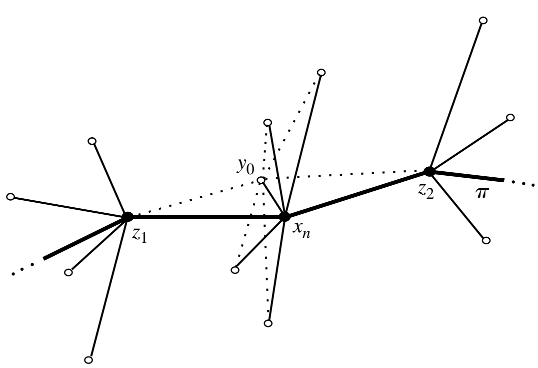

Figure 3. The caterpillar of Lemma 2.4 where dotted lines show the new caterpillar after x_n is dropped and y_0 is promoted to become a path vertex. None of the old edges incident to x_n are present in the new caterpillar.

Setting $m = |N_G(x_n)|$ we have $m \geq 3$ and there is at least one star vertex adjacent to x_n. Let μ be the smallest distance from x_n to a star vertex of G in $N_G(x_n)$, and let y_0 be a star vertex in $N_G(x_n)$ with $\mu = |y_0 - x_n|$. We also let z_1 and z_2 be the neighbors of x_n on the path π.

Now we construct a new spanning caterpillar $G' = (\chi', E', \pi')$. To define E' we take (a) all of the edges of E not incident to x_n, and (b) as new edges we add all of the edges (y_0, w) where $w \in N_G(x_n)\backslash\{y_0\}$. Since the set E' contains the edges (z_1, y_0) and (y_0, z_2), we define the path π' of the new spanning caterpillar G' by taking the old path π up to the vertex z_1, followed by the edge (z_1, y_0), (y_0, z_2) and then we follow the old path from z_2 to the end of π. This construction is illustrated by Figure 3.

To estimate the weight $W(G')$ of the spanning caterpillar that we have constructed, we repeat the construction with bookkeeping. By the triangle inequality and the definitions of m and μ, we have

$$\begin{aligned} W(G') &\leq M(x_1, x_2, \ldots, x_n) - \lambda|z_1 - x_n| - \lambda|z_2 - x_n| \\ &\quad + \lambda\{|z_1 - x_n| + |x_n - y_0|\} + \lambda\{|z_2 - x_n| + |x_n - y_0|\} \\ &\quad + (m-3)|y_0 - x_n| \\ &= M(x_1, x_2, \ldots, x_n) + (m - 3 + 2\lambda)\mu. \end{aligned} \tag{5}$$

The task now is to bound the last summand, and the plan is to exploit Lemma 2.3, which tells us that if m is large then μ must be small. We assume that each $B \in \mathscr{B}(k)$ contains a vertex of the path $\pi = \pi(G)$, so the optimality of G implies that the star edges incident to x_n cannot have length greater than $R = 2^{1/2}/k$. This gives us the lower bound

$$q \equiv |N_G(x_n) \cap \{y\colon |y - x_n| \leq R\}| \geq m - 2,$$

so, using Lemma 2.3 with $R = 2^{1/2}/k$ gives the bound $\mu \leq C\rho^{m-2}2^{1/2}/k$. Thus, we can generously bound last summand of (5) by

$$m\mu + 2\lambda\mu \leq C\rho^{-2}2^{1/2}\max_m\{m\rho^m\}/k + 2\lambda 2^{1/2}/k = O_\lambda(1/k),$$

which is all we need.

Lemma 2.5 (Cost to Add One). *There is a constant $C = C(\lambda)$ such that for all $\chi = \{x_1, x_2, \ldots, x_{n-1}\}$ and all minimal spanning caterpillars $G = (\chi, E, \pi)$, we have*

$$M(x_1, x_2, \ldots, x_n) \leq M(x_1, x_2, \ldots, x_{n-1}) + C/k, \tag{6}$$

provided that every $B \in \mathscr{B}(k)$ contains a vertex of the path π.

Proof. Unlike Lemma 2.4, this lemma is trivial. To get a spanning caterpillar of $\chi' = \chi \cup \{x_n\}$ we just join x_n to the nearest path point of the spanning caterpillar $G = (\chi, E, \pi)$. If $x_n \in B \in \mathscr{B}(k)$ there is a path point x' of G in B and we can joint x' to x_n at a cost not greater than $2^{1/2}/k$.

Remark. The results of this section underscore some distinctions between the MSC functional and the general theory of subadditive Euclidean functionals. Here the smoothness of M expressed by Lemmas 2.2 and 2.3 comes at a price. Constraints must be placed on the structure of the optimizing graph; specifically, one needs to know that backbone path π is well distributed throughout the square. This phenomenon is related in turn to the lack of uniform boundedness of the degrees of the MSC.

3. Stochastic features of the MSC's backbone. If the sample $\boldsymbol{\chi}_n = \{X_1, X_2, \ldots, X_n\}$ is independent with the uniform distribution on $[0, 1]^2$, and if $\lambda \neq 1$, then the minimal spanning caterpillar of $\boldsymbol{\chi}_n$ is unique with probability one, and it will be denoted by $G = (\boldsymbol{\chi}_n, E, \pi)$. If $\lambda = 1$ the minimal spanning caterpillar need not be unique since one typically has multiple choices for the backbone path π. Since these paths differ only in their first or last edges, we regain uniqueness in this case if we take $G = (\boldsymbol{\chi}_n, E, \pi)$ to be the minimum spanning caterpillar that has the smallest number of vertices on the path π.

The path $\pi = \pi(G)$ of the minimal spanning caterpillar is itself a graph, and we denote the set of vertices of π by $\pi_V(\boldsymbol{\chi}_n)$. The set of vertices of G that are not on the path will be denoted by $\pi_V^c(\boldsymbol{\chi}_n)$, and the elements of this set are called *star points*. Every element of $\boldsymbol{\chi}_n$ is thus either a star point or a path point.

Lemma 3.1 (Path Points in the Box). *There are two constants $\alpha = \alpha(\lambda) > 0$ and $C = C(\lambda)$ such that for all n, k, and $B \in \mathscr{B}(k)$ we have*

$$P(\pi_V(\boldsymbol{\chi}_n) \cap B = \varnothing) \leq Ck\exp(-\alpha n/k^3). \tag{7}$$

Proof. Let (G, E, π) be the minimum spanning caterpillar with vertex set $\boldsymbol{\chi}_n$; we observed earlier that for a uniform independent sample, the minimum spanning caterpillar is unique with probability one. To begin, we define $l = l(k)$ by setting

$$l = \lceil 3(\lambda + \sqrt{k\lambda}) \rceil + 1. \tag{8}$$

Now, for a given box $B \in \mathscr{B}(k)$, we let $\mathscr{H}(l, B)$ denote the set of l^2 squares of $\mathscr{B}(3kl)$ that are the middle ninth of B; explicitly, $\mathscr{H}(l, B)$ is the set of all squares $B' \in \mathscr{B}(3kl)$ for which we have

$$B' \subset [a/k + 1/(3k), a/k + 2/(3k)] \times [b/k + 1/(3k), b/k + 2/(3k)]. \tag{9}$$

Now we consider the two events

$$A = \{\omega\colon \pi_V(\boldsymbol{\chi}_n) \cap B = \varnothing\} \qquad \text{and} \qquad F = \left\{\omega\colon \min_{S \in \mathscr{H}(l, B)} |S \cap \boldsymbol{\chi}_n| > 0\right\};$$

that is, A is the event that there are no path points in the box B and F is the event that each subbox $S \in \mathscr{H}(l, B)$ contains at least one point of the vertex set $\boldsymbol{\chi}_n$.

If $A \cap F \neq \varnothing$, we take an $\omega \in A \cap F$ and then for $\boldsymbol{\chi}_n = \boldsymbol{\chi}_n(\omega)$ we construct a new spanning caterpillar $G' = (\boldsymbol{\chi}_n, E', \pi')$ as follows:

1. Since $\omega \in F$ we have $\boldsymbol{\chi}_n \cap S \neq \varnothing$ for each of the l^2 subsquares $S \in \mathscr{H}(l, B)$ and we select one point $v_S \in \boldsymbol{\chi}_n \cap S$ for each $S \in \mathscr{H}(l, B)$.
2. We let π_0 be a path through the set of l^2 points $\{v_S\colon S \in \mathscr{H}(l, B)\}$ that is of minimal Euclidean length.
3. Since $\omega \in A$, no vertices of $\pi_V(\boldsymbol{\chi}_n)$ are in B, so each element of the set $\{v_S\colon S \in \mathscr{H}(l, B)\}$ is a star point of G and each such v_S is connected to path point of caterpillar G that is in B^c. We call this edge e_S and we note that $|e_S| \geq 1/(3k)$ since the distance from a point of $S \in \mathscr{H}(l, B)$ to a point of B^c is at least $1/(3k)$.
4. To define the edge set E', we first take the edge set E and remove from E all of the set of edges $\{e_S\colon S \in \mathscr{H}(l, B)\}$. We then add to E' the edges of the path π_0 from step (2). Lastly, we add an edge e' that connects end point of π_0 to an end point of π. It does not matter how one makes the last choice from the four possibilities. The only control over the length of e' is that $|e'| \leq 2^{1/2}$.
5. To complete the specification of the spanning graph $G' = (\boldsymbol{\chi}_n, E', \pi')$, we take π' to be the path consisting of the edges of the old path π, the connecting edge e', and the path π_0 from step (2).

To estimate the weight of G' we recall $|e_S| \geq 1/(3k)$, bound the length of π_0 by Lemma 2.1 (with $t = 1/(3k)$), and use the generous bound $|e'| \leq 2$ to get

$$\begin{aligned} W(G') &\leq W(G) - \sum_{S \in \mathscr{H}(l,B)} |e_S| + \lambda \sum_{e \in \pi_0} |e| + \lambda |e'| \\ &\leq W(G) - l^2/(3k) + l\lambda/k + 2\lambda. \end{aligned} \tag{10}$$

By the minimality of the weight of the spanning caterpillar G we have that $W(G) \leq W(G')$, so by solving a quadratic equation we see that the bound (10) implies that

$$l \leq 3(\lambda + \sqrt{k\lambda}). \tag{11}$$

By our choice (8) of $l = l(k)$, the bound (11) does not hold, so we conclude that $A \cap F = \varnothing$.

Consequently, we have $A \subset F^c$ and since there are l^2 elements of $\mathscr{H}(l, B)$ we have

$$\mathrm{P}(A) \leq \mathrm{P}(F^c) \leq l^2 \left(1 - \frac{1}{9l^2k^2}\right)^n \leq l^2 \exp\left(-\frac{n}{9l^2k^2}\right). \tag{12}$$

Again, using the specification given by (8) of $l = l(k) = O(\sqrt{k})$, we have

$$0 < \inf_{k \geq 1} \left\{\frac{k}{9l^2(k)}\right\},$$

and we then take $\alpha = \alpha(\lambda)$ to be any constant less than this infimum. Again we observe by our choice (8) we have $l^2 = O(k)$, so we can then choose $C = C(\lambda)$ such that we have the bound (7) for all $n \geq 1$ and $k \geq 1$.

Remark. A novel and recurring feature of the MSC functional is that one has to attend to internal structures of the optimizing graph such as the set $\pi_V(\boldsymbol{\chi}_n)$ of points on the path. Lemma 3.1 tells us that with high probability every point in the square is reasonably close to one of these special points, and this is the kind one needs to make use of Lemmas 2.4 and 2.5.

4. Moments of change-one bounds and the variance.

LEMMA 4.1 (EXPECTED COST OF CHANGE). *For all $1 \le p < \infty$ and all $\epsilon > 0$ we have the bound*

$$\mathrm{E}[|M(\boldsymbol{\chi}_{n+1}) - M(\boldsymbol{\chi}_n)|^p] = O_{\lambda,\,p,\,\epsilon}(n^{\epsilon - p/3}) \quad \textit{for } n \to \infty. \tag{13}$$

PROOF. First we fix k and consider the set

$$F_n(k) = \{\omega\colon B \cap \pi_V(\boldsymbol{\chi}_n) \ne \varnothing \text{ for all } B \in \mathscr{B}(k)\}. \tag{14}$$

Since there are k^2 elements of $\mathscr{B}(k)$, Boole's inequality and Lemma 3.1 give us a bound on the complementary event,

$$P(F_n^c(k)) \le Ck^3 \exp(-\alpha n/k^3), \tag{15}$$

and for $F_{n-1}^c(k)$ we have the analogous bound.

Now, by using Lemmas 2.4 and 2.5 for $\omega \in F_n(k) \cap F_{n-1}(k)$ and applying Lemma 2.1 for $\omega \in F_n^c(k) \cup F_{n-1}^c(k)$ we have the pointwise bound

$$|M(\boldsymbol{\chi}_{n+1}) - M(\boldsymbol{\chi}_n)| \le C/k + Cn^{1/2}\mathbb{1}[F_n^c(k) \cup F_{n-1}^c(k)]. \tag{16}$$

Taking p'th powers and using $(a+b)^p \le 2^p(a^p + b^p)$ on the right side of (16), we find from the bound (15) that for a new constant $C = C(p)$ we have

$$\mathrm{E}[\,|M(\boldsymbol{\chi}_{n+1}) - M(\boldsymbol{\chi}_n)|^p\,] \le C/k^p + Cn^{p/2}k^3 \exp(-\alpha n/k^3).$$

Finally, taking $k = \lfloor n^{1/3-\epsilon/p} \rfloor$ gives us (13).

There is a general inequality for the variance that works nicely with "discrete continuity" like that provided in Lemma 4.1. To state the inequality, first consider a set of $2n$ independent random variables $\{X_i, X_i'\colon 1 \le i \le n\}$ that take values in $S = \mathbb{R}^d$. Next, given a Borel function $f\colon S^n \to \mathbb{R}$ we set $F = f(X_1, X_2, \ldots, X_n)$ and for $1 \le i \le n$ we set

$$F_i = f(X_1, \ldots, X_{i-1}, X_i', X_{i+1}, \ldots, X_n).$$

One then has the *jackknife* bound, see, e.g., Steele [29] or Boucheron et al. [5]:

$$\operatorname{Var} F \le \frac{1}{2}\,\mathrm{E}\sum_{i=1}^{n}(F - F_i)^2. \tag{17}$$

From this inequality and the estimate (13) with $p = 2$, one immediately gets a useful bound on the variance of $M(\boldsymbol{\chi}_n)$.

LEMMA 4.2 (VARIANCE ESTIMATE). *For all $\epsilon > 0$ we have*

$$\operatorname{Var} M(\boldsymbol{\chi}_n) = O_{\lambda,\,\epsilon}(n^{\epsilon+1/3}). \tag{18}$$

With the variance of $M(\boldsymbol{\chi}_n)$ under control, the proof of Theorem 1.1 will be in range once we determine the asymptotic behavior of $EM(\boldsymbol{\chi}_n)$.

5. Asymptotics of the mean. Let $\{N(t)\colon 0 \le t < \infty\}$ be a standard Poisson process with arrival rate one, and let $\{X_1, X_2, \ldots\}$ be an independent sequence of random vectors with the uniform distribution on $[0,1]^2$. We also assume that this sequence is independent of the process $\{N(t)\colon 0 \le t < \infty\}$. Next we set

$$\boldsymbol{\chi}_{N(t)} = \{X_1, X_2, \ldots, X_{N(t)}\} \qquad \text{and} \qquad \varphi(t) = \mathrm{E}[M(\boldsymbol{\chi}_{N(t)})].$$

The idea behind this construction is that $\varphi(t)$ is a smoothed version of the sequence of expected means, and we have the added benefit that for each $B \in \mathscr{B}(k)$ the cardinality of the set $\{X_1, X_2, \ldots, X_{N(t)}\} \cap B$ is Poisson with mean t/k^2. This observation leads to a simple inequality from which we can deduce the asymptotic behavior of $\varphi(t)$. The argument for Lemma 5.1 goes back to Beardwood et al. [4]. It is included here mainly for completeness, though it may help that some details are treated more explicitly than usual.

Lemma 5.1 (Poisson Averages of the Means). *For all $t > 0$ and all integer $k > 0$, we have*

$$\varphi(t) \le k\varphi(t/k^2) + 3k\lambda, \tag{19}$$

and there is a constant $\beta_{\mathrm{MSC}} = \beta_{\mathrm{MSC}}(\lambda) > 0$ such that

$$\lim_{t\to\infty} \varphi(t)/\sqrt{t} = \beta_{\mathrm{MSC}}. \tag{20}$$

Proof. For each $B \in \mathscr{B}(k)$, we let $S_B = B \cap \{X_1, X_2, \ldots, X_{N(t)}\}$. We then let $G_B = (S_B, E_B, \pi_B)$ be the unique minimum spanning caterpillar for S_B. Assume for the moment there is at least one edge of π_B for all B and let a_B and b_B be the two distinct end points of π_B. We then sew together all the paths π_B, $B \in \mathscr{B}(k)$. Specifically, we order the boxes of $\mathscr{B}(k)$ in boustrophedon (or plowman's) order; that is, we start on the upper left, go right along the top row, move down to the second row, then move left along the second row, etc. We connect b_B to $a_{B'}$, where B' is a successor of B. The Euclidean length of this stitching can be bound above (very crudely) by $3k$. This gives the bound

$$M(X_1, X_2, \ldots, X_{N(t)}) \le 3k\lambda + \sum_B W(G_B). \tag{21}$$

To remove the assumption that each S_B is not empty, we just note that if for some B we have $S_B = \varnothing$ we just can skip that box when we sew the small caterpillars together to get our spanning caterpillar $\boldsymbol{\chi}_n$. Similarly, if for some B the caterpillar G_B is a star with central vertex v we just take $a_B = b_B = v$. We can then go ahead with our sewing as before, and the bound (21) again applies. Finally, we take expectations in the bound (21). Scaling by both side and area, gives us $E[W(G_B)] = \varphi(t/k^2)k^{-1}$, and there are k^2 summands in sum, so in the end we have (19).

To prove (20), we replace t with $k^2 t$ in (19) and we divide by $\sqrt{k^2 t}$ to get the stabilized recursive inequality,

$$\frac{\varphi(k^2 t)}{\sqrt{k^2 t}} \le \frac{\varphi(t)}{\sqrt{t}} + \frac{3\lambda}{\sqrt{t}} \quad \text{for all } 1 \le k < \infty \text{ and } t > 0. \tag{22}$$

Now, given any $\epsilon > 0$ and any $T < \infty$, we can find by the continuity of φ an interval (a, b) such that $T < a < b$ and for which we have

$$\frac{\varphi(t)}{\sqrt{t}} \le \epsilon + \liminf_{s\to\infty} \frac{\varphi(s)}{\sqrt{s}} \equiv \epsilon + \gamma \quad \text{for all } t \in (a, b). \tag{23}$$

Choosing $T \ge (3\lambda/\epsilon)^2$, we then have by (22) and (23) that

$$\frac{\varphi(k^2 t)}{\sqrt{k^2 t}} \le \gamma + 2\epsilon \quad \text{for all } 1 \le k < \infty \text{ and all } t \in (a, b),$$

or, in other words, we have

$$\frac{\varphi(t)}{\sqrt{t}} \le \gamma + 2\epsilon \quad \text{for all } t \in \bigcup_{k=1}^{\infty} (k^2 a, k^2 b) \equiv S.$$

For $k \ge 3a(b-a)^{-1}$ the intervals $(k^2 a, k^2 b)$ and $((k+1)^2 a, (k+1)^2 b)$ overlap, so we have $(3^2 a^2 (b-a)^{-2} a, \infty) \subset S$, and consequently we have

$$\frac{\varphi(t)}{\sqrt{t}} \le \gamma + 2\epsilon \quad \text{for all } t > 3^2 a^3 (b-a)^{-2}.$$

Since $\epsilon > 0$ is arbitrary, this is more than we need for the proof.

Now we want to extract the asymptotic behavior of $\mathrm{E}[M(\boldsymbol{\chi}_n)]$ from what we have learned about $\varphi(t)$. One could appeal to the Tauberian theory for Borel means (cf. Korevaar [20, Chapter 6]), but $\mathrm{E}[M(\boldsymbol{\chi}_n)]$ is so well behaved that it is quicker to use bare hands.

Lemma 5.2 (Mean Increments and DePoissonization).

$$\lim_{n\to\infty} \mathrm{E}[M(\boldsymbol{\chi}_n)]/\sqrt{n} = \beta_{\mathrm{MSC}}(\lambda) > 0. \tag{24}$$

Proof. If we set $A(n) = \mathrm{E}[M(\boldsymbol{\chi}_n)]$ and take N_t to be a Poisson random variable with mean t, then by conditioning and the definition of φ we have

$$\varphi(t) = \mathrm{E}\, A(N_t) = \sum_{j=0}^{\infty} A(j) P(N_t = j) = \sum_{j=0}^{\infty} A(j) \frac{t^j}{j!} e^{-t}.$$

Fixing $0 < \epsilon < 1/6$, we then write

$$t_+(\epsilon) = t + t^{1/2+\epsilon} \qquad \text{and} \qquad t_-(\epsilon) = t - t^{1/2+\epsilon},$$

and, by repeated applications of (13) with $p = 1$, we get the relation

$$\sup_{j \in [t_-(\epsilon),\, t_+(\epsilon)]} |A(j) - A(\lfloor t \rfloor)| = O(t^{2\epsilon+1/6}). \tag{25}$$

Now we make estimations over the three ranges. On the midrange $[t_-(\epsilon), t_+(\epsilon)]$ we use (25), and on the outside ranges $[0, t_-(\epsilon)]$ and $[t_+(\epsilon), \infty)$ we use the bound from Lemma 2.1 that tells us $A(j) \le Cj^{1/2}$ for all j. Assembling the three pieces we have

$$\begin{aligned} |\varphi(t) - A(\lfloor t \rfloor)| &= \sum_{j=0}^{\infty} |\{A(j) - A(\lfloor t \rfloor)\}| e^{-t} t^j / j! \\ &= O(t^{1/2} P(N_t \le t_-(\epsilon))) + O(t^{2\epsilon+1/6}) + O(E(N_t^{1/2} \mathbb{1}(N_t \ge t_+(\epsilon)))) \\ &= o(t^{1/2}). \end{aligned}$$

To check this, note that first summand is $o(t^{1/2})$ because $P(N_t \le t_-(\epsilon)) = o(1)$, and the last summand is $o(t^{1/2})$ by the Cauchy-Schwarz inequality and the exponential estimate for the upper Poisson tail. From this bound and the limit (20) we have the limit (24). Finally, for the strict positivity of $\beta_{\mathrm{MSC}}(\lambda)$, one can look ahead to the comparison of $\beta_{\mathrm{MSC}}(\lambda)$ and β_{MST} given by the inequality (34) of §7.

6. Completion of the argument. The tools are in place to complete the proof of the first assertion of Theorem 1.1. We first note that if we set $n_j = j^2$, then by the variance bound (18) we have

$$\mathrm{Var}[n_j^{-1/2} M(\boldsymbol{\chi}_{n_j})] = O(j^{2\epsilon-4/3}).$$

Now, taking $0 < \epsilon < 1/6$, Chebyshev's inequality gives us for all $\delta > 0$ that

$$\sum_{j=1}^{\infty} \mathrm{P}(|n_j^{-1/2}(M(\boldsymbol{\chi}_{n_j}) - \mathrm{E}[M(\boldsymbol{\chi}_{n_j})])| \ge \delta) \le \delta^{-2} \sum_{j=1}^{\infty} \mathrm{Var}[n_j^{-1/2} M(\boldsymbol{\chi}_{n_j})] < \infty.$$

The Borel-Cantelli lemma, Lemma 5.2, and the arbitrariness of δ then tell us that

$$\lim_{j\to\infty} n_j^{-1/2} M(\boldsymbol{\chi}_{n_j}) = \beta_{\mathrm{MSC}} \quad \text{a.s.} \tag{26}$$

Next, fix k and recall the set $F_n(k) = \{\omega\colon B \cap \pi_V(\boldsymbol{\chi}_n) \ne \varnothing \text{ for all } B \in \mathcal{B}(k)\}$ that was introduced in the proof of Lemma 4.1. We observed there that we have the bound $P(F_n^c(k)) \le Ck^3 \exp(-\alpha n/k^3)$, so by another application of the Borel-Cantelli lemma we have $P(F_n^c(k)\ i.o.) = 0$.

Now, given n we define j by the relations $n_j \le n < n_{j+1}$. We can then apply Lemma 2.1 for $\omega \in F_{n_j}^c(k)$ and apply Lemma 2.5 for $\omega \in F_{n_j}(k)$ to get the bound

$$M(\boldsymbol{\chi}_n) \le M(\boldsymbol{\chi}_{n_j}) + C(n - n_j)k^{-1} + 3n^{1/2} \max(1, \lambda) \mathbb{1}(F_{n_j}^c(k)).$$

Dividing by $n^{1/2}$, taking the limsup, using (26), and recalling the definition of n_j, we find

$$\limsup_{n\to\infty} n^{-1/2} M(\boldsymbol{\chi}_n) \le \beta_{\mathrm{MSC}} \quad \text{with probability } 1,$$

and this proves half of the assertion of Theorem 1.1.

With the natural changes, one can prove the second half. This time we apply Lemma 2.1 for $\omega \in F_n^c(k)$ and use Lemma 2.5 for $\omega \in F_n(k)$ to get the bound

$$M(\boldsymbol{\chi}_{n_{j+1}}) \le M(\boldsymbol{\chi}_n) + C(n_{j+1} - n)k^{-1} + 3n_{j+1}^{1/2} \max(1, \lambda)\mathbb{1}(F_n^c(k)),$$

so, when we divide by $n^{1/2}$ and take the liminf on both sides, we find from (26) that

$$\beta_{\mathrm{MSC}} \le \liminf_{n\to\infty} n^{-1/2} M(\boldsymbol{\chi}_n) \quad \text{with probability } 1,$$

completing the proof of the first assertion of Theorem 1.1.

One can prove the second assertion of Theorem 1.1, by a variation of the approximation argument that Beardwood et al. [4] used in analysis of the traveling salesman problem. Here we use a maximal coupling that shortens that argument, but, even so, we just give a sketch.

First, we note that if the random variables X_i, $i = 1, 2, \ldots$ have a density f with compact support in $\mathbb{R}^2$, then by translation and scaling we can suppose without loss of generality that the support of f is contained in $[0, 1]^2$. Next we note that we can approximate f as well as we like by a density ϕ that is constant on each of the subsquares B in our decomposition $\mathscr{B}(k)$ of $[0, 1]^2$ into squares of side $1/k$. More precisely, for any $\epsilon > 0$ there is an integer k and there is a constant $\alpha(B) \ge 0$ for each $B \in \mathscr{B}(k)$ such that the weighted sum of indicator functions

$$\phi(x) = \sum_{B\in\mathscr{B}(k)} \alpha(B)\mathbb{1}_B(x)$$

is a density on $[0, 1]^2$ and

$$\int_{\mathbb{R}^2} |f(x) - \phi(x)|\, dx \le \epsilon.$$

Now, by the existence of a maximal coupling (see, e.g., Lindvall [22], p. 18), we can choose an independent sequence $Z_i = (X_i, Y_i)$, $i = 1, 2, \ldots$ such that for all i, X_i has density f, Y_i has density ϕ, and

$$P(X_i \ne Y_i) \le \epsilon. \tag{27}$$

Since we have $P(Y_i \in B) = \alpha_B/k^2$, we see by the law of large numbers that

$$|\{Y_1, Y_2, \ldots, Y_n\} \cap B| \sim n\alpha_B/k^2 \quad \text{with probability one,}$$

so, by scaling and the first part (2) of Theorem 1.1, we have with probability one for each $B \in \mathscr{B}(k)$ that

$$\begin{aligned}\lim_{n\to\infty} M(\{Y_1, Y_2, \ldots, Y_n\} \cap B)/\sqrt{n} &= \frac{1}{k}\beta_{\mathrm{MSC}}(\lambda)\sqrt{\alpha_B/k^2} \\ &= \beta_{\mathrm{MSC}}(\lambda)\int_B \sqrt{\phi(x)}\, dx \end{aligned}\tag{28}$$

since $\phi(x) = \alpha_B$ for all $x \in B$. The union of the minimum spanning caterpillars $\{Y_1, Y_2, \ldots, Y_n\} \cap B$ over all subsquares $B \in \mathscr{B}(k)$ has a length that differs from the length of the minimum spanning caterpillar of the whole sample $\{Y_1, Y_2, \ldots, Y_n\}$ by an amount that is bounded independently of n (for k fixed), so by summing the previous limit we have

$$\lim_{n\to\infty} M(\{Y_1, Y_2, \ldots, Y_n\})/\sqrt{n} = \beta_{\mathrm{MSC}}(\lambda)\int_{\mathbb{R}^2} \sqrt{\phi(x)}\, dx. \tag{29}$$

Finally, by the strong law of large numbers and the coupling bound (27) applied to the sum of the indicators $\mathbb{1}(X_i \ne Y_i)$, we see that the cardinality of the difference between the sets $\{Y_1, Y_2, \ldots, Y_n\}$ and $\{X_1, X_2, \ldots, X_n\}$ is almost surely $O(\epsilon n)$. Consequently we have that the difference between $M(Y_1, Y_2, \ldots, Y_n)$ and $M(X_1, X_2, \ldots, X_n)$ is almost surely $O(\sqrt{\epsilon n})$ so by (29) and the arbitrariness of $\epsilon > 0$, one has the second conclusion (3) of Theorem 1.1.

7. Connecting the MSC, MST, and TSP: Constants and complexity. The limiting constant $\beta_{\mathrm{MSC}}(\lambda)$ of Theorem 1.1 has a natural relationship to the corresponding limit constants β_{MST} and β_{TSP} for the minimal spanning tree problem and the traveling salesman problem. Although the values of β_{MST} and β_{TSP} are still not known exactly, they have been investigated repeatedly, see, e.g., Finch [12, pp. 497–500]. The known rigorous bounds on β_{MST} and β_{TSP} are relatively crude, but for the TSP there have been increasingly sophisticated simulations with bounded errors on the TSP calculations. Record holders Johnson et al. [18] give

$$\beta_{\mathrm{TSP}} = 0.7124 \pm 0.0002. \tag{30}$$

Moscato and Norman [25] also give a fractal, space-filling heuristic for which they determine the exact value $\beta_{\mathrm{MN*TSP}}$ of their limit constant,

$$\beta_{\mathrm{MN*TSP}} = \frac{4(1+2\sqrt{2})\sqrt{51}}{153} = 0.7147\ldots \tag{31}$$

This is certainly intriguing, but the space-filling model and the independent uniform model are not perfect matches.

In a remarkable paper Avram and Bertsimas [3] give an exact formula for β_{MST}, but it comes at the price of an infinite sum of integrals that are not easy to evaluate. The authors required an integration *tour de force* to give a rigorous proof (in their Theorem 9) that

$$\beta_{\mathrm{MST}} \geq 0.600822. \tag{32}$$

In a brief but interesting simulation study, Cortina-Borja and Robinson [6] estimated that

$$\beta_{\mathrm{MST}} = 0.6331 \pm 0.0013, \tag{33}$$

but this estimate is based on a relatively small sample size with n not bigger than $2^{15} = 32{,}768$. Since fast algorithms are known for the MST, it seems feasible to extend the simulations to much larger sample sizes.

Returning to spanning caterpillars, we note that for any set $\chi = \{x_1, x_2, \ldots, x_n\}$, we have the elementary bounds

$$\min(1,\lambda)\mathrm{MST}(\chi) \leq \mathrm{MSC}_\lambda(\chi) \leq \max(1,\lambda)\mathrm{TSP}(\chi), \tag{34}$$

where $\mathrm{TSP}(\chi)$ denotes the length of the shortest Euclidean path through χ, $\mathrm{MST}(\chi)$ denotes the Euclidean length of the minimal spanning tree, and $\mathrm{MSC}_\lambda(\chi)$ is the weight of the minimum spanning caterpillar with path edge weight factor λ. These bounds are crude, but they do show that $\beta_{\mathrm{MSC}}(\lambda)$ is strictly positive for all $\lambda > 0$. For $\lambda = 1$, the minimum spanning caterpillar constant $\beta_{\mathrm{MSC}}(1)$ also inherits the suggestive simulation bounds (30) and (33) as well as any rigorous bounds that are proved for β_{TSP} and β_{MST}.

For the exact optimum and for $(1+\epsilon)$ approximations, the computational complexity of the MSC is much closer to the notoriously hard TSP than to the notoriously easy MST. Even if some generous oracle were to identify the set of vertices that are on the path of an optimal MSC, one would still need to solve a traveling salesman problem to put those vertices into an optimal order. Although this observation may fall short of a formal proof that the Euclidean MSC problem is NP-hard, it cannot not fall short by much.

The question of an approximate solution of the Euclidean MSC problem is much more interesting. It seems inevitable that the approximation schemes of Arora [1] can be modified to provide a solution of the Euclidean MSC problem that is within in a factor of $1 + 1/c$ of the optimum and do so with a running time of $O(n(\log^{O(c)}(n))$. We have not pursued this point, but it does seem worth pursuing.

Even though our interest in minimum weight caterpillars comes mainly from communication networks, we should note that caterpillars have a long history in graph theory. Harary and Schwenk [17] used Pólya enumeration theory to show that the number of nonisomorphic caterpillars on $n+4$ vertices is given by the elegant formula $2^n + 2^{\lfloor n/2 \rfloor}$, and, in the same paper (p. 361), the authors credit A. Hobbs for introducing the term "caterpillar." Harary and Schwenk [15] and [16] had earlier investigated the connectivity properties of graph powers of caterpillars. More recently, Ortiz and Villanueva [26] studied independent sets in caterpillar graphs and found that the whole family of independent sets can be found in polynomial time. Caterpillars are the simplest graphs for which the graphical bandwidth problem is nontrivial, so caterpillars have a natural place in many bandwidth investigations, e.g., Assmann et al. [2], Miller [23], Monien [24], Haralambides et al. [14], Feige and Talwar [10], and Lin et al. [21].

Finally, we should note that in chemistry, caterpillars are also known as *Gutman trees*, the later name referring back to Gutman [13]; sometimes they are also called *benzenoid trees* because of their common presence in the structure of benzenoid hydrocarbons. Surveys of El-Basil [7, 8] and [9] detail these connections and give many further references.

8. Concluding observations. T. Tao's essay on the extended reals (Tao [32, pp. 38–56]) makes the case that the extended real numbers ${}^*\mathbb{R}$ can be of concrete benefit to almost any mathematician, pure or applied. Here, with a very light use of the extended reals, we can answer a question that was raised in the introduction; we just need to take the weight factor λ for the path edges to be a strictly positive infinitesimal.

Specifically, we fix λ to be any extended real number such that $0 < \lambda < x$ for all $x \in \mathbb{R}$ with $x > 0$. With this choice, the cost of any nonpath edge is prohibitively expensive. Consequently, one elects to run the backbone π through all of the points of $\boldsymbol{\chi}_n$, and the weight of the minimum spanning caterpillar is just λ times the (usual Euclidean) cost of the optimal traveling salesman path. Thus, if one reformulates Theorem 1.1 to permit an infinitesimal $\lambda \in {}^*\mathbb{R}$, we see that Theorem 1.1 is a strict generalization of the Beardwood, Halton, and Hammersley theorem. Moreover, a proof of Theorem 1.1 that allows for a strictly positive infinitesimal $\lambda \in {}^*\mathbb{R}$ is virtually identical to the proof we have given here.

It is fair to say that this use of the extended reals is just a matter of language. Nevertheless, we were genuinely uncertain at one point if one could view Theorem 1.1 as an honest generalization of the Beardwood, Halton, and Hammersley theorem. It eventually became clear that the formal introduction of the extended reals would make an affirmative answer easy. Although this may be a just matter of language, at the end of the day, it seems to be an instance of *useful* language.

Our second observation also concerns the flexibility of the free parameter λ. This parameter was introduced because of modeling motivations of the kind that were mentioned in the introduction. Nevertheless, after λ enters the game, it presents new mathematical possibilities—possibilities that are not present in problems like the TSP and MST. For example, $\beta_{\mathrm{MSC}}(\lambda)$ is differentiable, and $\beta'_{\mathrm{MSC}}(\lambda)$ provides a measure of the relative weight that is placed on the path edges in the limit. Thus, the free parameter λ offers a special handle on the asymptotic geometry of the MSC for which there are no direct analogs in the traditional theory of the TSP or MST.

Our final observations concern the possibility of refinements and extensions of Theorem 1.1. In particular, it is natural to ask if there is a central limit theorem to complement the strong law for minimum spanning caterpillars. After all, Kesten and Lee [19] established the central limit theorem for the minimal spanning tree under similar circumstances.

An important distinction between the MST and the MSC is that there is a variational characterization for the edge set of the MST. Specifically, if the edges between the elements of the vertex set V have distinct lengths, then e is an edge of the MST of V if and only if there exists a partition (A, A^c) of V such that e is the shortest edge between A and A^c. This property played a crucial role in the proof of Kesten and Lee [19], and there is no corresponding criterion for the TSP—or, *a fortiori*, for the MSC. Although it is perfectly plausible that there is a CLT for the MSC and the TSP, the proof of such a result seems very far from current capabilities. People have contemplated the possibility of a CLT for the TSP for at least 50 years.

On the other hand, one can be more sanguine about the extension of the MSC strong law to d-dimensions. Specifically, it is natural to expect that if the random variables X_i, $i = 1, 2, \ldots$ are independent and have a density f with compact support in $\mathbb{R}^d$, then we have with probability one that

$$\lim_{n\to\infty} n^{-(d-1)/d} M(X_1, X_2, \ldots, X_n) = \beta_{\mathrm{MSC},\,d}(\lambda) \int_{\mathbb{R}^d} f(x)^{1/d}\, dx, \tag{35}$$

where the constant $\beta_{\mathrm{MSC},\,d}(\lambda)$ depends only on the dimension d. A critical step in the proof of (35) would be to develop the appropriate d-dimensional analogs of Lemmas 2.2, 2.3, and 2.4. The surgeries we used to prove these lemmas were greatly simplified by working in $d = 2$, and, since we were motivated by communication network problems with distinctly geographic origins, we did not pursue this extension.

References

[1] Arora S (1998) Polynomial time approximation schemes for Euclidean traveling salesman and other geometric problems. *J. ACM* 45(5):753–782.

[2] Assmann SF, Peck GW, Sysło MM, Zak J (1981) The bandwidth of caterpillars with hairs of length 1 and 2. *SIAM J. Algebraic Discrete Methods* 2(4):387–393.

[3] Avram F, 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.

[4] Beardwood J, Halton JH, Hammersley JM (1959) The shortest path through many points. *Proc. Cambridge Philos. Soc.* 55:299–327.

[5] Boucheron S, Lugosi G, Bousquet O (2004) Concentration inequalities. Bousquet O, von Luxburg U, Rätsch G, eds. *Advanced Lectures on Machine Learning*, Vol. 3176, Lecture Notes in Comput. Sci. (Springer, Berlin), 208–240.

[6] Cortina-Borja M, Robinson T (2000) Estimating the asymptotic constants of the total length of Euclidean minimal spanning trees with power-weighted edges. *Statist. Probab. Lett.* 47(2):125–128.

[7] El-Basil S (1987) Applications of caterpillar trees in chemistry and physics. *J. Math. Chemistry* 1:153–174.

[8] El-Basil S (1990) Caterpillar (Gutman) trees in chemical graph theory. Gutman I, Cyvin S, eds. *Advances in the Theory of Benzenoid Hydrocarbons*, Vol. 153, Topics in Current Chemistry (Springer, Berlin), 273–289.
[9] El-Basil S (2008) Combinatorial properties of graphs and groups of physico-chemical interest. *Combinatorial Chemistry and High Throughput Screening* 11(9):707–722.
[10] Feige U, Talwar K (2005) Approximating the bandwidth of caterpillars. *Approximation, Randomization and Combinatorial Optimization*, Vol. 3624, Lecture Notes in Comput. Sci. (Springer, Berlin), 62–73.
[11] Few L (1955) The shortest path and the shortest road through n points in a region. *Mathematika* 2:141–144.
[12] Finch SR (2003) *Mathematical Constants*. *Encyclopedia of Mathematics and Its Applications*, Vol. 94 (Cambridge University Press, Cambridge, UK).
[13] Gutman I (1977) Topological properties of benzenoid systems. *Theoret Chemistry Accounts: Theory, Comput., Modeling* (*Theoretica Chimica Acta*) 45:309–315.
[14] Haralambides J, Makedon F, Monien B (1991) Bandwidth minimization: An approximation algorithm for caterpillars. *Math. Systems Theory* 24(3):169–177.
[15] Harary F, Schwenk A (1971) Trees with Hamiltonian square. *Mathematika* 18:138–140.
[16] Harary F, Schwenk A (1972) A new crossing number for bipartite graphs. *Utilitas Math.* 1:203–209.
[17] Harary F, Schwenk AJ (1973) The number of caterpillars. *Discrete Math.* 6:359–365.
[18] Johnson DS, McGeoch LA, Rothberg EE (1996) Asymptotic experimental analysis for the Held-Karp traveling salesman bound. Tardos E, ed. *Proc. Seventh Annual ACM-SIAM Sympos. Discrete Algorithms, Atlanta, GA* (ACM, New York), 341–350.
[19] Kesten H, Lee S (1996) The central limit theorem for weighted minimal spanning trees on random points. *Ann. Appl. Probab.* 6(2):495–527.
[20] Korevaar J (2004) Tauberian theory. *Grundlehren der Mathematischen Wissenschaften* [*Fundamental Principles of Mathematical Sciences*], Vol. 329 (Springer, Berlin).
[21] Lin M, Lin Z, Xu J (2006) Graph bandwidth of weighted caterpillars. *Theoret. Comput. Sci.* 363(3):266–277.
[22] Lindvall T (1992) *Lectures on the Coupling Method* (Dover Publications, Mineola, NY).
[23] Miller Z (1981) The bandwidth of caterpillar graphs. Hoffman F, Reid KB, Mullin RC, Stanton RG, eds., *Proc. Twelfth Southeastern Conf. Combinatorics, Graph Theory Comput., Vol. 2, Baton Rouge, LA*, Vol. 33 (Utilitas Mathematica Publications, Winnipeg, Canada), 235–252.
[24] Monien B (1986) The bandwidth minimization problem for caterpillars with hair length 3 is NP-complete. *SIAM J. Algebraic Discrete Methods* 7(4):505–512.
[25] Moscato P, Norman MG (1998) On the performance of heuristics on finite and infinite fractal instances of the Euclidean traveling salesman problem. *INFORMS J. Comput.* 10(2):121–132.
[26] Ortiz C, Villanueva M (2012) Maximal independent sets in caterpillar graphs. *Discrete Appl. Math.* 160(3):259–266.
[27] Rhee WT (1993) A matching problem and subadditive Euclidean functionals. *Ann. Appl. Probab.* 3(3):794–801.
[28] Steele JM (1981) Subadditive Euclidean functionals and nonlinear growth in geometric probability. *Ann. Probab.* 9(3):365–376.
[29] Steele JM (1986) An Efron-Stein inequality for nonsymmetric statistics. *Ann. Statist.* 14(2):753–758.
[30] Steele JM (1988) Growth rates of Euclidean minimal spanning trees with power weighted edges. *Ann. Probab.* 16(4):1767–1787.
[31] Steele JM (1997) *Probability Theory and Combinatorial Optimization*, Vol. 69, CBMS-NSF Regional Conference Series in Applied Mathematics (Society for Industrial and Applied Mathematics (SIAM), Philadelphia).
[32] Tao T (2008) *Structure and Randomness* (American Mathematical Society, Providence, RI).
[33] Yukich JE (1998) *Probability Theory of Classical Euclidean Optimization Problems*, Vol. 1675, Lecture Notes in Mathematics (Springer, Berlin).