# On the Size of Some Trees Embedded in Rd

Pedro Machado Manhães de Castro, Olivier Devillers

▶ **To cite this version:**

**HAL Id: inria-00448335**

**https://hal.inria.fr/inria-00448335**

Submitted on 18 Jan 2010

INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

# On the Size of Some Trees Embedded in $\mathbb{R}^d$

Pedro Machado Manhães de Castro — Olivier Devillers

**N° 7179**

Janvier 2010

Rapport de recherche

# On the Size of Some Trees Embedded in $\mathbb{R}^d$

Pedro Machado Manhães de Castro , Olivier Devillers

Thème : Computational Geometry, Geometric Graphs
Équipe-Projet Geometrica

**Abstract:** This paper extends the result of Steele [6, 5] on the worst-case length of the Euclidean minimum spanning tree $EMST$ and the Euclidean minimum insertion tree $EMIT$ of a set of $n$ points $S \subset \mathbb{R}^d$. More precisely, we show that, if the weight $w$ of an edge $e$ is its Euclidean length to the power of $\alpha$, the following quantities $\sum_{e \in EMST} w(e)$ and $\sum_{e \in EMIT} w(e)$ are both worst-case $O(n^{1-\alpha/d})$, where $d$ is the dimension and $\alpha$, $0 < \alpha < d$, is the weight. Also, we analyze and compare the value of $\sum_{e \in T} w(e)$ for some trees $T$ embedded in $\mathbb{R}^d$ which are of interest in (but not limited to) the point location problem [2].

**Key-words:**  Computational Geometry, Geometric Graphs

# Sûr la Taille de Quelques Arbres Plongées dans $\mathbb{R}^d$

**Résumé :** Ce papier étend le résultat de Steele [6, 5] sûr la taille au pire des cas de la plus petite arbre de couverture minimal Euclidienne et l'arbre d'insertion minimal d'un ensemble de $n$ points $S \subset \mathbb{R}^d$. Plus précisément, nous démontrons que si le poids $w$ d'une arête $e$ est sa longueur Euclidienne à la puissance $\alpha$, les quantités suivantes $\sum_{e \in EMST} w(E)$ et $\sum_{e \in EMIT} w(E)$ valent au pire des cas $O(n^{1-\alpha/d})$, où $d$ est la dimension et $\alpha$, $0 < \alpha < d$, est le poid. Nous déterminons et comparons aussi la valeur de $\sum_{e \in T} w(E)$ pour des arbres $T$ plongées dans $\mathbb{R}^d$, qui sont d'intèrêt au problème de la localisation des points [2].

**Mots-clés :** Géométrie Algorithmique, Graphes Géométriques

# 1 Introduction

This paper extends the result of Steele [6, 5] on the worst-case length of the Euclidean minimum spanning tree $EMST$ and the Euclidean minimum insertion tree $EMIT$ of a set of $n$ points $S \subset \mathbb{R}^d$. More precisely, we show that, if the weight $w$ of an edge $e$ is its Euclidean length to the power $\alpha$, the following quantities $\sum_{e \in EMST} w(e)$ and $\sum_{e \in EMIT} w(e)$ are both worst-case $O(n^{1-\alpha/d})$, where $d$ is the dimension and $\alpha$, $0 < \alpha < d$, is the weight. Also, we analyze and compare the value of $\sum_{e \in T} w(e)$ for some trees $T$ embedded in $\mathbb{R}^d$ which are of interest in (but not limited to) the point location problem [2].

Let $S = \{p_i, 1 \le i \le n\}$ be a set of points in $\mathbb{R}^d$ and $G = (V, E)$ be the complete graph such that the vertex $v_i \in V$ is embedded on the point $p_i \in S$; the edge $e_{ij} \in E$ linking two vertices $v_i$ and $v_j$ is weighted by $|p_i - p_j|^\alpha$, its Euclidean length to the power of $\alpha$. $G$ is usually referred to as the *geometric graph* of $S$. We will denote the sum of the weight of the edges of $G$ by $|G|_\alpha$ (or $|G|$ if $\alpha = 1$). We will also refer to $|G|_\alpha$ as the *weighted-length* of $G$.

Consider the following trees:
(i) A *star* is a tree having one vertex that is linked to all others (see Figure 1a).
(ii) A *path* is a tree having all vertices of degree 2 but two with degree 1 (see Figure 1b).
(iii) Among all the trees spanning $S$, a tree with the minimal length is called an *Euclidean minimum spanning tree* of $S$ and denoted $EMST(S)$ (see Figure 1c).
(iv) Consider that an ordering is given by a permutation $\sigma$, vertices are inserted in the order $v_{\sigma(1)}, v_{\sigma(2)}, \dots, v_{\sigma(n)}$. We build incrementally a spanning tree $T_i$ for $S_i = \{p_{\sigma(j)} \in S, i \le j\}$ with $T_1 = \{v_{\sigma(1)}\}$, $T_i = T_{i-1} \cup \{v_{\sigma(i)} v_{\sigma(j)}\}$ and a fixed $k$, with $1 \le k < n$, such that $v_{\sigma(i)} v_{\sigma(j)}$ has the shortest length for any $\max(1, i - k) \le j < i$. This tree is called the *Euclidean minimum $k$-insertion tree*, and will be denoted by $EMIT_k(S)$ (see Figure 1e); when $k = n - 1$, we will write $EMIT(S)$ (see Figure 1d). $|EMIT(S)|$ depends on $\sigma$ and for some permutations it coincides with $|EMST(S)|$.

It is noteworthy that both the combinatorics of $EMST(S)$ and of $EMIT(S)$ are invariant to $\alpha$ (since $f(\lambda) = \lambda^\alpha$, is monotonically increasing for positive $\alpha$ and $\lambda$). What changes is the sum of the weights associated with the edges. Also, $EMST$ assumes that the position of all the points in $S$ is available (static) whereas $EMIT$ clearly does not (dynamic).



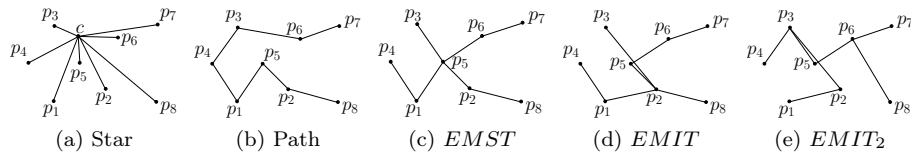(a) Star    (b) Path    (c) $EMST$    (d) $EMIT$    (e) $EMIT_2$

Figure 1: **Trees embedded in the plane.**

Steele proves [4] that if $p_i$ are i.i.d. random variables with compact support, then $|EMST(S)|_\alpha = O(n^{1-\alpha/d})$ with probability 1. For the extreme case of $\alpha = d$, Aldous and Steele [1] show that $|EMST(S)|_d = O(1)$ if the variables

above are evenly distributed in the unit cube. Without any dependence on probabilistic hypotheses, Steele proves [6] that the complexity of $|EMST(S)|$ is bounded by $O(n^{1-1/d})$ in the worst case. Finally, the asymptotic length of $|EMIT(S)|$ is shown to be the same as the one of $|EMST(S)|$ [5]. In other words, $|EMIT(S)| = O(n^{1-1/d})$. This result is surprising because it means that *a priori* knowledge of $S$ does not affect the asymptotic length of trees following the greedy strategy of an $EMST$. This fact has application in the dynamic point location problem [2].

This work is presented as follow: First, in Section 2, we extend the result of Steele [5] stating that $|EMIT|$ and $|EMST|$ have the same asymptotical behavior to the case of $|EMIT|_\alpha$ and $|EMST|_\alpha$ for $0 < \alpha < d$ as well. Second, in Section 3, we obtain the expected weighted-length of some stars of interest inside the unit ball. Then in Section 4 we obtain some ratios between the expected weighted-length of such stars and random paths in the unit ball. Finally, in Section 5 we obtain bounds on the expected weighted-length of $EMIT_k$.

## 2   Weighted Euclidean Minimum Insertion Tree

We extend the result of Steele [5] for $|EMIT|_\alpha$ (and consequently $|EMST|_\alpha$) with the following theorem:

**Theorem 1.** *Let $S$ be a sequence of $n$ points in $[0,1]^d$, then $|EMST(S)|_\alpha \leq |EMIT(S)|_\alpha \leq \gamma_{d,\alpha} n^{1-\alpha/d}$, with $d \geq 2$ and $0 < \alpha < d$. Where, $\gamma_{d,\alpha} = 1 + \frac{2^{4d}d^{d/2}}{(2^\alpha-1)(d/\alpha-1)}$.*

The proof of Theorem 1 follows exactly the same line as Steele [5], and starts with the two lemmas below. Given a fixed sequence $S = \{p_1, p_2, \ldots, p_n\}$ of points in $\mathbb{R}^d$, then we can build a spanning tree for $S$ by sequentially joining $x_i$ to the tree formed by $\{p_1, p_2, \ldots, p_{i-1}\}$ for $1 < i \leq n$. Let $w_i \in \mathbb{R}$ be defined as follow:

$$w_i = \min_{1 \leq j < i} |p_i - p_j|^\alpha, \tag{1}$$

then $w_i$ is the minimal cost of joining $p_i$ to a vertex of a spanning tree of $\{p_1, p_2, \ldots, p_{i-1}\}$. Now, we have that $|EMIT(S)|_\alpha = \sum_{1 < i \leq n} w_i$.

**Lemma 2.** *If $\{p_1, p_2, \ldots, p_n\} \subset [0,1]^d$ and $w_i = \min_{1 \leq j < i} |p_i - p_j|^\alpha$, for $1 < i \leq n$, $d \geq 2$ and $0 < \alpha < d$, then for any $0 < \lambda < \infty$ we have*

$$\sum_{\lambda \leq w_i < 2^\alpha \lambda} w_i^{d/\alpha} \leq 8^d d^{d/2}. \tag{2}$$

*Proof.* Let $C = \{i : \lambda \leq w_i < 2^\alpha \lambda\}$ and for each $i \in C$ let $B_i$ be a ball of radius $r_i = \frac{1}{4} w_i^{1/\alpha}$ with center $p_i$. We will argue by contradiction that $B_i \cap B_j = \emptyset$ for all $i < j$. If $B_i \cap B_j \neq \emptyset$, then the bounds $r_i \leq 2^\alpha \lambda$ and $r_j \leq 2^\alpha \lambda$ gives us

$$|p_i - p_j| \leq \frac{1}{4}(w_i^{1/\alpha} + w_j^{1/\alpha}) < \lambda^{1/\alpha}. \tag{3}$$

But, by definition of $w_j$ we have $|p_i - p_j|^\alpha \geq w_j$ for all $i < j$, which implies $|p_i - p_j| \geq w_j^{1/\alpha}$ for all $i < j$; and, by the lower bound on the summands in Eq.(2)

we have $\lambda \leq w_j$, which means $\lambda^{1/\alpha} \leq w_j^{1/\alpha}$, so we also see $|p_i - p_j| \geq \lambda^{1/\alpha}$. Since $|p_i - p_j| \geq \lambda^{1/\alpha}$ contradicts Eq.(3), we have $B_i \cap B_j = \emptyset$.

Now, since all of the balls $B_i$ are disjoint and contained in a sphere with radius $2\sqrt{d}$, the sum of their volumes is bounded by the volume of the sphere of radius $2\sqrt{d}$. Thus, if $\omega_d$ denotes the volume of the unit ball in $\mathbb{R}^d$, we have the bound $\sum_{i \in C} \omega_d w_i^{d/\alpha} 4^{-d} \leq \omega_d 2^d d^{d/2}$ from which Eq.(2) follows. $\qquad\square$

**Lemma 3.** *Let* $\Psi$ *be a positive and non-increasing function on the interval* $\left(0, \sqrt{d}\right]$, *then for any* $0 < a < b \leq \sqrt{d}$, *with* $d \geq 2$ *and* $0 < \alpha < d$,

$$\sum_{a \leq w_i \leq b} w_i^{(d/\alpha)+1} \Psi(w_i) \leq \frac{2^\alpha}{2^\alpha - 1} \cdot 8^d d^{d/2} \int_{a/2^\alpha}^b \Psi(\lambda) \mathrm{d}\lambda. \qquad (4)$$

*Proof.* By Lemma 2 we have for any $0 < \lambda < \infty$,

$$\sum_{a \leq w_i < b} w_i^{d/\alpha} I(\lambda \leq w_i < 2^\alpha \lambda) \leq 8^d d^{d/2},$$

where

$$I(\lambda \leq w_i < 2^\alpha \lambda) = I(\frac{1}{2^\alpha} w_i \leq \lambda < w_i)$$

is the indicator function. If we multiply by $\Psi(\lambda)$ and integrate over $[\frac{1}{2^\alpha} a, b]$, we find

$$\sum_{a \leq w_i \leq b} w_i^{(d/\alpha)} \int_{w_i/2^\alpha}^{w_i} \Psi(\lambda) \mathrm{d}\lambda \leq 8^d d^{d/2} \int_{a/2^\alpha}^b \Psi(\lambda) \mathrm{d}\lambda. \qquad (5)$$

Since $\Psi$ is non-increasing, the integrand on the left-hand side of Eq.(5) is bounded from below by $\Psi(w_i)$, so $\Psi(w_i) w_i (1 - \frac{1}{2^\alpha}) \leq \int_{w_i/2^\alpha}^{w_i} \Psi(\lambda) \mathrm{d}\lambda$, and Eq.(4) follows from Eq(5). $\qquad\square$

*Proof of Theorem 1.* Divide the set $\{w_2, w_3, \ldots, w_n\}$ in two sets $R_1 = \{w_i : w_i \leq n^{-\alpha/d}\}$ and $R_2 = \{w_i : w_i > n^{-\alpha/d}\}$. We have the trivial bound $\sum_{w_i \in R_1} w_i = \sum_{w_i \leq n^{-\alpha/d}} w_i \leq n \cdot n^{-\alpha/d} = n^{1-\alpha/d}$. Now, let $[a,b] = \left[n^{-\alpha/d}, \sqrt{d}\right]$, $\Psi(\lambda) = \lambda^{-d/\alpha}$ in Eq.(4), then we have:

$$\sum_{w_i \geq n^{-\alpha/d}} w_i \leq \frac{2^\alpha}{2^\alpha - 1} \cdot 8^d d^{d/2} (d/\alpha - 1)^{-1} \cdot \left(2^{d-\alpha} n^{1-\alpha/d} - d^{-(d/\alpha-1)/2}\right).$$

And hence, $\sum_{w_i \in R_2} w_i \leq \frac{2^{4d} d^{d/2}}{(2^\alpha - 1)(d/\alpha - 1)} \cdot n^{1-\alpha/d}$. Now, we have that

$$\sum_{i=2}^n w_i = \sum_{w_i \in R_1} w_i + \sum_{w_i \in R_2} w_i \leq \left(1 + \frac{2^{4d} d^{d/2}}{(2^\alpha - 1)(d/\alpha - 1)}\right) \cdot n^{1-\alpha/d},$$

from which Theorem 1 follows. The constant is $\gamma_{d,\alpha} = 1 + \frac{2^{4d} d^{d/2}}{(2^\alpha-1)(d/\alpha-1)}$. $\qquad\square$

# 3   Two Stars Embedded in $\mathbb{R}^d$

Assume a set of points $S = \{p_i, 1 \leq i \leq n\}$, evenly distributed inside the unit ball $\mathcal{B}$. Let $c$ be a point in $\mathcal{B}$, and $E(|cp|^\alpha, p \in \mathcal{B})$ be the expected value of the distance between $c$ and a point inside the unit ball to the power of $\alpha$, with $\alpha > 0$, as $n \to \infty$. The following theorem distinguish two stars of particular interest.

**Theorem 4.** *The shortest and largest expected weighted-length stars inside the ball are respectively: the star centered at $O$, the center of the ball; and the star centered at $\Omega$, a point on the boundary of the ball, denoted by $\mathcal{H}$.*

We have that

$$\lim_{n \to \infty} \sum_{i=1}^n \frac{|p_i - c|^\alpha}{n} = E(|cp|^\alpha, p \in \mathcal{B}).$$

In other words, $\lim_{n \to \infty} E(|cp|^\alpha, p \in \mathcal{B})$ is the weighted-length of an edge of the star centered in $c$ as $n \to \infty$. Now, let us turn the attention to the so-called *weighted Steiner star* of $S$, the star with the smallest weighted-length we can imagine of $S$. Consider that the weighted Steiner star of the first $n$ points is centered at $c_n^*$, then we have

$$\sum_{i=1}^n |p_i - c_n^*|^\alpha \leq \sum_{i=1}^n |p_i - c|^\alpha, \tag{6}$$

for any $n > 0$ and $c \in \mathcal{B}$. Let $A_n(c) = \sum_{i=1}^n |p_i - c|^\alpha / n$. The proof of Theorem 4 is divided into several lemmas. First, we show that $A_n(c_n^*)$ converges. Then we show that there is a point $\bar{c} \in \mathcal{B}$ such that $A_n(\bar{c})$ and $A_n(c_n^*)$ converge to the same limit. Finally, we argue that $\bar{c}$ must be the center of the ball, the point $O$.

**Lemma 5.** *Let $A_n(c) = \sum_{i=1}^n |p_i - c|^\alpha / n$ and $c_n^*$ be the center of the weighted Steiner star of the first $n$ points $A_n(c_n^*)$. Then $A_n(c_n^*)$ converges.*

*Proof.* As points are evenly distributed in $\mathcal{B}$, $c_n^*$ is a non-constant sequence with probability 1, and thus we can assume that $0 < A_n(c_n^*) < 2^\alpha$ for any $n > 1$ and $d \geq 1$. We have from Eq.(6)

$$
\begin{aligned}
A_{n+1}(c_{n+1}^*) &\leq& A_{n+1}(c_n^*) = \sum_{i=1}^n \frac{|p_i - c_n^*|^\alpha}{n+1} + \frac{|p_{n+1} - c_n^*|^\alpha}{n+1} \\
&\leq& \sum_{i=1}^n \frac{|p_i - c_n^*|^\alpha}{n} + \frac{|p_{n+1} - c_n^*|^\alpha}{n}.
\end{aligned}
$$

But $|p_{n+1} - c_n^*| \leq 2$, and thus $A_{n+1}(c_{n+1}^*) \leq A_n(c_n^*) + 2^\alpha/n$, which means that $\lim_{n \to \infty} \frac{A_{n+1}(c_{n+1}^*)}{A_n(c_n^*)} \leq 1$. Analogously, we have, from the fact that $|p_{n+1} - c_{n+1}^*| \leq 2$ and Eq.(6), $A_n(c_n^*) \leq \left(\frac{n+1}{n}\right) A_{n+1}(c_{n+1}^*) - \frac{2^\alpha}{n}$. This means that $\lim_{n \to \infty} \frac{A_{n+1}(c_{n+1}^*)}{A_n(c_n^*)} \geq 1$, and thus $\lim_{n \to \infty} \frac{A_{n+1}(c_{n+1}^*)}{A_n(c_n^*)} = 1$. As $0 < A_n(c_n^*) < 2^\alpha$ for any $n > 1$ and $d \geq 1$, $A_n(c_n^*)$ converges. $\qquad \square$

Now, we need the following easy lemma.

**Lemma 6.** *Assume $a, b \in \mathbb{R}$, $0 \le a, b \le 2$, for each $\alpha > 0$ there is a function $f_\alpha(b)$ such that $(a + b)^\alpha \le a^\alpha + f_\alpha(b)$ and $\lim_{b \to 0} f_\alpha(b) = 0$.*

*Proof.* For $\alpha \le 1$, we have $f_\alpha(b) = b^\alpha$, as $(a + b)^\alpha \le a^\alpha + b^\alpha$.
For $\alpha \ge 2$ and $\alpha \in \mathbb{N}$, we have $f_\alpha(b) = 2^{2\alpha-1}b$, as

$$(a + b)^\alpha = a^\alpha + b \sum_{i=1}^{\alpha} \binom{\alpha}{i} a^{\alpha-i} b^{i-1} \le a^\alpha + 2^{2\alpha-1}b.$$

Let, $\{x\}$ signify $x - \lfloor x \rfloor$, then for $\alpha > 1$ and $\alpha \notin \mathbb{N}$, we have $f_\alpha(b) = 2^{2\lfloor \alpha \rfloor} b^{\{\alpha\}} + 2^{2\alpha}b$, as

$$
\begin{aligned}
(a + b)^\alpha &= (a+b)^{\lfloor \alpha \rfloor}(a+b)^{\{\alpha\}} \\
&\le \left( a^{\lfloor \alpha \rfloor} + 2^{2\lfloor \alpha \rfloor - 1}b \right) \left( a^{\{\alpha\}} + b^{\{\alpha\}} \right) \\
&\le a^\alpha + 2^{2\lfloor \alpha \rfloor} b^{\{\alpha\}} + 2^{2\alpha}b.
\end{aligned}
$$

$\square$

**Lemma 7.** *There is a point $\bar{c} \in \mathcal{B}$ such that $A_n(\bar{c})$ and $A_n(c_n^*)$ converge to the same limit.*

*Proof.* If a topological space $X$ is compact, then every infinite subset of $X$ has an accumulation point. Assume $\bar{c}$ is an accumulation point of the sequence $\{c_i^*\}_{i=1,2,\ldots,\infty}$, then we have a subsequence of indices $\zeta : \mathbb{N}^* \to \mathbb{N}^*$ such that $\{c_{\zeta(i)}^*\}_{i=1,2,\ldots,\infty}$ converges to $\bar{c}$. Because of the triangulation inequality and by direct application of Lemma 6, for any $i > 0$, we have that

$$A_{\zeta(i)}(\bar{c}) \le A_{\zeta(i)}(c_{\zeta(i)}^*) + f_\alpha(|\bar{c}c_{\zeta(i)}^*|).$$

As $|\bar{c}c_{\zeta(i)}^*|$ converges to 0, then $f_\alpha(|\bar{c}c_{\zeta(i)}^*|)$ also converges to 0 (see Lemma 6). And thus $\lim_{i \to \infty} A_{\zeta(i)}(\bar{c}) = \lim_{i \to \infty} A_{\zeta(i)}(c_{\zeta(i)}^*)$. Therefore, as $A_n(c_n^*)$ converges, $\lim_{i \to \infty} A_{\zeta(i)}(c_{\zeta(i)}^*) = \lim_{n \to \infty} A_n(c_n^*)$ and $\lim_{n \to \infty} A_n(\bar{c}) = \lim_{i \to \infty} A_{\zeta(i)}(\bar{c}) = \lim_{n \to \infty} A_n(c_n^*)$.
$\square$

*Proof of Theorem 4.* By symmetry, $\lim_{n \to \infty} A_n(O) \le \lim_{n \to \infty} A_n(c)$ for any $c \in \mathcal{B}$, and from Eq.(6) and Lemma 7, we have that $\lim_{n \to \infty} A_n(O) \ge \lim_{n \to \infty} A_n(c_n^*) = \lim_{n \to \infty} A_n(\bar{c}) \ge \lim_{n \to \infty} A_n(O)$. Therefore, at the limit, we have that the length of the weighted Steiner star is equivalent to the length of the star centered at $O$.

With analogous arguments, we have that the largest weighted-star and a star centered at the boundary of the ball have equivalent length. $\square$

Denote the shortest and largest expected weighted-length stars inside the ball by $\mathcal{S}$ and $\mathcal{H}$ respectively. Let $E(|Op|^\alpha, p \in \mathcal{B})$ and $E(|\Omega p|^\alpha, p \in \mathcal{B})$ be the expected value of an edge of $\mathcal{S}$ and $\mathcal{H}$ respectively, then as $n \to \infty$ the size of $\mathcal{S}$ and $\mathcal{H}$ are given accordingly by $n \cdot E(|Op|^\alpha, p \in \mathcal{B})$ and $n \cdot E(|\Omega p|^\alpha, p \in \mathcal{B})$.

We analyze in the sequel the values of $E(|Op|^\alpha, p \in \mathcal{B})$ and $E(|\Omega p|^\alpha, p \in \mathcal{B})$.

**Theorem 8.** *When points are uniformly i.i.d in a ball, the expected size* $E(|Op|^\alpha, p \in \mathcal{B})$ *of and edge of the star centered at the center of the unit ball, with positive $\alpha$, is given by:*

$$\left(\frac{d}{d+\alpha}\right).$$

*Proof.* Let $\mathcal{B}_l$ be a ball with radius $l$ centered at the origin, we have

$$
\begin{aligned}
E(|Op|^\alpha, p \in \mathcal{B}) &= \int_0^1 l^\alpha Prob(p \in \mathcal{B}_{l+\mathrm{d}l} \setminus \mathcal{B}_l) = \int_0^1 l^\alpha \frac{(\mathrm{d}V_d(l)/l)\mathrm{d}l}{V_d(1)} \\
&= \int_0^1 dl^{d-1+\alpha}\mathrm{d}l = \frac{d}{d+\alpha},
\end{aligned}
$$

where $V_d(l)$ is the volume of a ball of radius $l$ (and $\mathrm{d}V_d(l)/l$ is its area). ☐

**Theorem 9.** *When points are uniformly i.i.d in a ball, the expected size* $E(|\Omega p|^\alpha, p \in \mathcal{B})$ *of and edge of the star centered at the boundary of the unit ball, with positive $\alpha$, is given by:*

$$2^{d+\alpha}\left(\frac{2d+\alpha}{2d+2\alpha}\right)\frac{B\left(\frac{d}{2}+\frac{1}{2}, \frac{d}{2}+\frac{1}{2}+\frac{\alpha}{2}\right)}{B\left(\frac{d}{2}+\frac{1}{2}, \frac{1}{2}\right)},$$

*where $B(x,y) = \int_0^1 \lambda^{x-1}(1-\lambda)^{y-1}\mathrm{d}\lambda$ is the so-called Beta function.*

The computation of the average is more involved than in Theorem 8, and we split the computation into several lemmas.

**Lemma 10.** *Consider the spherical cap $\mathcal{H}_h$ formed by crossing a ball $\mathcal{B}_R$ with radius $R$ centered at the origin, with the plane $x = R - h$. Denote $h$ the height of the cap. The volume of $\mathcal{H}_h$ is the volume of the intersection between the half-space $x \geq R - h$ and $\mathcal{B}_R$. This volume is given by:*

$$R^d \frac{\pi^{\frac{d-1}{2}}}{\Gamma(\frac{d+1}{2})} \int_0^{\arccos\left(\frac{R-h}{R}\right)} \sin^d(\lambda)\mathrm{d}\lambda. \tag{7}$$

*Proof.* The volume $V_d(r)$ of a ball with radius $r$ in dimension $d$ is given by $r^d \cdot \pi^{\frac{d}{2}}/\Gamma(1+\frac{d}{2})$. Each cross-section $x = R - h + \delta$, $0 \leq \delta \leq h$ is a $(d-1)$-dimensional ball. If we integrate all of those balls along the $x$ axis, we have $\int_{R-h}^R V_{d-1}(\sqrt{R^2 - t^2})\mathrm{d}t$. Eq.(7) follows from replacing $t$ by $\lambda = R\cos(t)$. ☐

**Lemma 11.** *Let $\Omega$ be a point on the boundary of the unit ball $\mathcal{B}_{unit}$, and $P_{\mathcal{H}}(l) = Prob(|\Omega p| \leq l; p \in \mathcal{B}_{unit})$ be the cumulative distribution function of distances between an uniformly distributed random point inside $\mathcal{B}_{unit}$ and $\Omega$, then*

$$P_{\mathcal{H}}(l) = \frac{1}{B\left(\frac{d}{2}+\frac{1}{2}, \frac{1}{2}\right)}\left(\int_0^{\arccos\left(1-l^2/2\right)} \sin^d(\lambda)\mathrm{d}\lambda + l^d \int_0^{\arccos\left(l/2\right)} \sin^d(\lambda)\mathrm{d}\lambda\right),$$

*where $B(x,y) = \int_0^1 \lambda^{x-1}(1-\lambda)^{y-1}\mathrm{d}\lambda$ is the Beta function.*

*Proof.* If we denote $\mathcal{B}_l$ the ball of radius $l$ centered in $\Omega$, the desired probability is clearly $volume(\mathcal{B}_l \cap \mathcal{B}_{unit})/volume(\mathcal{B}_{unit})$. $\mathcal{B}_l \cap \mathcal{B}_{unit}$ is the union of two spherical caps limited by the plane $x = 1 - l^2/2$ which can be computed using Lemma 10. ☐

*Proof of Theorem 9.* The theorem follows from:

$$E(|\Omega p|^\alpha, p \in \mathcal{B}) = \int_0^2 l^\alpha P'_{\mathcal{H}}(l)\mathrm{d}l$$

$$= \int_0^2 l^\alpha \left( \frac{\frac{1}{2}l^d \left(1 - \frac{l^2}{4}\right)^{\frac{d-1}{2}}}{B\left(\frac{d}{2} + \frac{1}{2}, \frac{1}{2}\right)} + dl^{d-1} \frac{\int_0^{\arccos(l/2)} \sin^d(\lambda)\mathrm{d}\lambda}{B\left(\frac{d}{2} + \frac{1}{2}, \frac{1}{2}\right)} \right) \mathrm{d}l$$

$$= \frac{1}{2} \int_0^2 \frac{l^{d+\alpha}\left(1 - \frac{l^2}{4}\right)^{\frac{d-1}{2}} \mathrm{d}l}{B\left(\frac{d}{2} + \frac{1}{2}, \frac{1}{2}\right)} + \frac{1}{2}\int_0^2 2dl^{d-1+\alpha} \frac{\int_0^{\arccos(l/2)} \sin^d(\lambda)\mathrm{d}\lambda}{B\left(\frac{d}{2} + \frac{1}{2}, \frac{1}{2}\right)}\mathrm{d}l$$

The right part of the expression above corresponds exactly to the expected value of $l^\alpha$ where $l$ is the length of a random segment determined by two evenly distributed points in the unit ball [3, 7]. Its value is given by:

$$\int_0^2 2dl^{d-1+\alpha} \frac{\int_0^{\arccos(l/2)} \sin^d(\lambda)d\lambda}{B\left(\frac{d}{2} + \frac{1}{2}, \frac{1}{2}\right)}\mathrm{d}l = 2^{d+\alpha}\left(\frac{d}{d+\alpha}\right)\frac{B\left(\frac{d}{2} + \frac{1}{2}, \frac{d}{2} + \frac{1}{2} + \frac{\alpha}{2}\right)}{B\left(\frac{d}{2} + \frac{1}{2}, \frac{1}{2}\right)}.$$

The left part of the expression can be obtained as follows:

$$\int_0^2 \frac{l^{d+\alpha}\left(1 - \frac{l^2}{4}\right)^{\frac{d-1}{2}} \mathrm{d}l}{B\left(\frac{d}{2} + \frac{1}{2}, \frac{1}{2}\right)} = 2\int_0^1 \frac{2^{d+\alpha}y^{d+\alpha}\left(1 - y^2\right)^{\frac{d-1}{2}} \mathrm{d}y}{B\left(\frac{d}{2} + \frac{1}{2}, \frac{1}{2}\right)}$$

$$= \int_0^1 \frac{2^{d+\alpha}z^{\frac{d}{2} + \frac{\alpha}{2} - \frac{1}{2}}\left(1 - z\right)^{\frac{d-1}{2}} \mathrm{d}z}{B\left(\frac{d}{2} + \frac{1}{2}, \frac{1}{2}\right)}$$

$$= 2^{d+\alpha}\frac{B\left(\frac{d}{2} + \frac{1}{2}, \frac{d}{2} + \frac{1}{2} + \frac{\alpha}{2}\right)}{B\left(\frac{d}{2} + \frac{1}{2}, \frac{1}{2}\right)}.$$

Finally, we have

$$E(|\Omega p|^\alpha, p \in \mathcal{B}) = 2^{d+\alpha}\left(\frac{2d+\alpha}{2d+2\alpha}\right)\frac{B\left(\frac{d}{2} + \frac{1}{2}, \frac{d}{2} + \frac{1}{2} + \frac{\alpha}{2}\right)}{B\left(\frac{d}{2} + \frac{1}{2}, \frac{1}{2}\right)}.$$

$\square$

## 4   Some Ratios

We may ask now what is the value of the ratio $\rho(d, \alpha)$ between $E(|\Omega p|^\alpha, p \in \mathcal{B})$ and $E(|Op|^\alpha, p \in \mathcal{B})$. It is an easy exercise to verify that $\rho(1, \alpha) = 2^\alpha$. In Corollary 12, we compute $\lim_{d \to \infty} \rho(d, \alpha)$.

**Corollary 12.** *The ratio $\rho(d, \alpha) = E(|\Omega p|^\alpha, p \in \mathcal{B})/E(|Op|^\alpha, p \in \mathcal{B})$ when $d \to \infty$ is given by $2^{\alpha/2}$.*

*Proof.* Computing $\rho(d, \alpha)$ with Theorems 8 and 9 gives:

$$\rho(d, \alpha) = 2^{d+\alpha}\left(\frac{2d+\alpha}{2d}\right)\frac{B\left(\frac{d}{2} + \frac{1}{2}, \frac{d}{2} + \frac{1}{2} + \frac{\alpha}{2}\right)}{B\left(\frac{d}{2} + \frac{1}{2}, \frac{1}{2}\right)} \tag{8}$$

Using the Stirling's identities:

$$B(a,b) \sim \sqrt{2\pi} \frac{a^{a-\frac{1}{2}} b^{b-\frac{1}{2}}}{(a+b)^{a+b-\frac{1}{2}}}, a, b \gg 0, \tag{9}$$

$$B(a,b) \sim \Gamma(b) a^{-b}, a \gg b > 0, \tag{10}$$

we have:

$$
\begin{aligned}
\lim_{d\to\infty} \rho(d,\alpha) &= \lim_{d\to\infty} \left\{ 2^{d+\alpha} \left( \frac{2d+\alpha}{2d} \right) \frac{B\left( \frac{d}{2} + \frac{1}{2} + \frac{\alpha}{2}, \frac{d}{2} + \frac{1}{2} \right)}{B\left( \frac{d+1}{2}, \frac{1}{2} \right)} \right\} \\
&= \lim_{d\to\infty} \left\{ \frac{2^{d+\alpha} \sqrt{2\pi} \left( \frac{d}{2} + \frac{1}{2} \right)^{\frac{d}{2}} \left( \frac{d}{2} + \frac{1}{2} + \frac{\alpha}{2} \right)^{\frac{d}{2}+\frac{\alpha}{2}}}{\sqrt{\pi} \left( d + 1 + \frac{\alpha}{2} \right)^{d+\frac{1}{2}+\frac{\alpha}{2}} \left( \frac{d}{2} + \frac{1}{2} \right)^{-\frac{1}{2}}} \right\} \\
&= 2^{\alpha/2} \cdot \lim_{d\to\infty} \left\{ \frac{(d+1)^{\frac{d+1}{2}} (d+1+\alpha)^{\frac{d+\alpha}{2}}}{\left( d+1+\frac{\alpha}{2} \right)^{\frac{d+1}{2}} \left( d+1+\frac{\alpha}{2} \right)^{\frac{d+\alpha}{2}}} \right\} \\
&= 2^{\alpha/2} \cdot e^{-\alpha/4} \cdot e^{\alpha/4} = 2^{\alpha/2}
\end{aligned}
$$

$\square$

If we consider a tree which is a random path in the unit ball, then the average size of its edge is given by the expected weighted-length of a random segment determined by two evenly distributed points in the unit ball. This is given by:

$$2^{d+\alpha} \left( \frac{d}{d+\alpha} \right) \frac{B\left( \frac{d}{2} + \frac{1}{2}, \frac{d}{2} + \frac{1}{2} + \frac{\alpha}{2} \right)}{B\left( \frac{d}{2} + \frac{1}{2}, \frac{1}{2} \right)} \tag{11}$$

The reader may refer to Tu and Fischbach [7] for a proof.

From Theorem 8, Corollary 12 and Eq.(11) we obtain the following corollary:

**Corollary 13.** *The ratio between the weighted-length of a random path in the unit ball and the weighed-length of a star centered at the center of the unit ball is $\frac{2^\alpha}{1+\alpha/2}$ when $d = 1$ and $2^{\alpha/2}$ when $d \to \infty$.*

From Theorem 9 and Eq.(11) we obtain the following corollary:

**Corollary 14.** *The ratio between the weighted-length of a star centered at the boundary of the unit ball and the weighted-length of a random path in the unit ball is given by $\frac{2d+\alpha}{2d}$.*

# 5   Weighted Euclidean Minimum $k$-Insertion Tree for Random Points

Now, we will compute a bound on the expected weighted-length of an edge of $|EMIT_k|_\alpha$ for points evenly distributed in the unit ball.

**Theorem 15.** *When points are uniformly i.i.d in a ball, the expected length $E(length)$ of an edge of $|EMIT_k|_\alpha$, with positive $\alpha$, verifies:*

$$\left( \frac{\alpha}{d} \right) B\left( k+1, \frac{\alpha}{d} \right) \le E(length) \le 2^\alpha \left( \frac{\alpha}{d} \right) B\left( k+1, \frac{\alpha}{d} \right), \tag{12}$$

*where $B(x,y) = \int_0^1 \lambda^{x-1} (1-\lambda)^{y-1} d\lambda$ is the Beta function.*

First, we will evaluate the weighted-distance between the origin and the closest amongst $k$ points $\{p_1, p_2, \ldots, p_k\}$ evenly distributed in the unit ball. This provides the lower-bound. Then, we will find an upper-bound on the weighted-distance between any point inside the ball and the closest amongst $k$ points $\{p_1, p_2, \ldots, p_k\}$ evenly distributed in the unit ball.

**Lemma 16.** *Let $c$ be a point inside the unit ball, $Prob\,(|cp| \leq l) = P_c(l)$ be the probability that the distance between a point $p \in \mathcal{B}$ and $c$ is less or equal to $l$, and $P_{c,k}(l) = Prob\,(\min(|cp_j|)_{1 \leq j \leq k} \leq l)$ be the cumulative distribution function of the minimum distance among $k$ points following a uniformly i.i.d inside the unit ball and $c$, then*

$$P_{c,k}(l) = 1 - (1 - P_c(l))^k.$$

*Proof.*

$$
\begin{aligned}
P_{c,k}(l) &= Prob\,(\min(|cp_j|)_{1 \leq j \leq k} \leq l) \\
&= 1 - Prob\,(|cp_j| > l\,, 1 \leq j \leq k) \\
&= 1 - Prob\,(|cp_1| > l)^k \\
&= 1 - (1 - P_c(l))^k.
\end{aligned}
$$

$\square$

A direct consequence of Lemma 16 is the following corollary.

**Corollary 17.** *Let $P_{\mathcal{B},k}(l) = Prob\,(\min(|Op_j|)_{1 \leq j \leq k} \leq l)$ be the cumulative distribution function of the minimum distance among $k$ points following a uniformly i.i.d inside the unit ball, and the center of the unit ball, then*

$$P_{\mathcal{B},k}(l) = 1 - (1 - l^d)^k.$$

**Lemma 18.** *The expected value $E\,(\min(|Op_j|^\alpha)_{1 \leq j \leq k})$ of the minimum weighted-distance among $k$ points following a uniformly i.i.d inside the unit ball and the center of the unit ball, with positive $\alpha$, is given by*

$$E\,(\min(|Op_j|^\alpha)_{1 \leq j \leq k}) = \left(\frac{\alpha}{d}\right) B\left(k + 1, \frac{\alpha}{d}\right).$$

*Proof.* Using Corollary 17, we have:

$$
\begin{aligned}
E\,(\min(|Op_j|^\alpha)_{1 \leq j \leq k}) &= \int_0^1 l^\alpha P'_{\mathcal{B},k}(l)\mathrm{d}l = kd \int_0^1 l^{d-1+\alpha}\left(1 - l^d\right)^{k-1}\mathrm{d}l \\
&= k \int_0^1 \lambda^{\alpha/d}\left(1 - \lambda\right)^{k-1}\mathrm{d}\lambda = kB\left(k, 1 + \frac{\alpha}{d}\right) \\
&= \left(\frac{\alpha}{d}\right) B\left(k + 1, \frac{\alpha}{d}\right).
\end{aligned}
$$

$\square$

Now, we shall obtain the upper-bound, which is more involved. First we will obtain a general expression for the expected value of $\min(|cp_j|^\alpha)_{1 \leq j \leq k}$. Assume $\delta(c) = 1 + |Oc|$ in what follows.

**Lemma 19.** *The expected value $E\left(\min(|cp_j|^\alpha)_{1\le j\le k}\right)$ of the minimum weighted-distance among $k$ points following a uniformly i.i.d inside the unit ball and $c$, with positive $\alpha$, is given by*

$$E\left(\min(|cp_j|^\alpha)_{1\le j\le k}\right) = \int_0^{\delta(c)} \alpha l^{\alpha-1}(1 - P_c(l))^k \mathrm{d}l.$$

*Proof.* As $P_c(0) = 0$ and $P_c(\delta(c)) = 1$, integration by parts gives us the following identity:

$$\int_0^{\delta(c)} l^\alpha P_c'(l) P_c^{i-1}(l) = \frac{\delta(c)^\alpha - \int_0^{\delta(c)} \alpha l^{\alpha-1} P_c^i(l)}{i}, i > 0. \tag{13}$$

From Lemma 16, we also have the following expression for $E\left(\min(|cp_j|^\alpha)_{1\le j\le k}\right)$:

$$
\begin{aligned}
E\left(\min(|cp_j|^\alpha)_{1\le j\le k}\right) &= \int_0^{\delta(c)} k l^\alpha P_c'(l)(1 - P_c(l))^{k-1} \\
&= \sum_{i=0}^{k-1} (-1)^i \binom{k-1}{i} \int_0^{\delta(c)} k l^\alpha P_c'(l) P_c^i(l) \mathrm{d}l. \tag{14}
\end{aligned}
$$

Replacing Eq.(13) in Eq.(14) leads to:

$$
\begin{aligned}
\sum_{i=0}^{k-1} (-1)^i \binom{k-1}{i} \int_0^{\delta(c)} k l^\alpha P_c'(l) P_c^i(l) \mathrm{d}l &= \sum_{i=0}^{k} (-1)^i \binom{k}{i} \int_0^{\delta(c)} \alpha l^{\alpha-1} P_c^i(l) \mathrm{d}l \\
&= \int_0^{\delta(c)} \alpha l^{\alpha-1} \sum_{i=0}^{k} (-1)^i \binom{k}{i} P_c^i(l) \mathrm{d}l \\
&= \int_0^{\delta(c)} \alpha l^{\alpha-1}(1 - P_c(l))^k \mathrm{d}l.
\end{aligned}
$$

$\square$

*Proof of Theorem 15.* Lemma 18 gives us the lower-bound in Theorem 15. Now, if we take a function $\Psi(l)$ such that $\Psi(l) \le P_c(l)$ for $0 \le l \le \delta(c)$, it upper-bounds the integral in Lemma 19. Take $\Psi(l) = (l/\delta(c))^d$, then we have:

$$
\begin{aligned}
E\left(\min(|cp_j|^\alpha)_{1\le j\le k}\right) &= \int_0^{\delta(c)} \alpha l^{\alpha-1}(1 - P_c(l))^k \mathrm{d}l \\
&\le \int_0^{\delta(c)} \alpha l^{\alpha-1}(1 - \Psi(l))^k \mathrm{d}l \\
&= \int_0^{\delta(c)} \alpha l^{\alpha-1} \left(1 - \frac{l^d}{\delta(c)^d}\right)^k \mathrm{d}l \\
&= \int_0^1 \alpha \delta(c)^\alpha \lambda^{\alpha-1}(1 - \lambda^d)^k \mathrm{d}\lambda \\
&= \delta(c)^\alpha \left(\frac{\alpha}{d}\right) \int_0^1 y^{(\alpha/d)-1}(1 - y)^k \mathrm{d}y \\
&= \delta(c)^\alpha \left(\frac{\alpha}{d}\right) B\left(k + 1, \frac{\alpha}{d}\right).
\end{aligned}
$$

And thus, $\delta(c) = 2$ ($c$ on the boundary) maximizes the value above. This completes the proof. $\square$

**Remark:** The results obtained here for evenly distributed points in the unit ball (expected values, ratios, bounds) so far are valid for any positive $\alpha$. Then, we have that the expected weighted-length of the $i$-th step $s_i$ of $|EMIT|_d$ in the unit ball for evenly distributed point is

$$\frac{1}{i+1} = B(i+1,1) \le s_i \le 2^d B(i+1,1) = \frac{2^d}{i+1}.$$

Evaluating for $n$ steps leads to

$$\Omega(\log n) = \sum_{i=2}^{n+1} \frac{1}{i} \le |EMIT|_d \le 2^d \sum_{i=2}^{n+1} \frac{1}{i} = O(\log n).$$

Unlike $|EMST|_d$ of points evenly distributed inside the unit cube, which converges to a constant as $n \to \infty$ [1], the expected value of $|EMIT|_d$ of points evenly distributed inside the unit ball is $\Theta(\log n)$. With the same argument, for $k > 0$ we have that the expected $|EMIT_k|_d = \Theta\left(\log k + \frac{n}{k}\right)$. ∎

# References

[1] D. Aldous and J. M. Steele. Asymptotics for euclidean minimal spanning trees on random points. *Probab. Theory Related Fields*, 92:247–258, 1992.

[2] Pedro Machado Manhães de Castro and Olivier Devillers. Self-Adapting Point Location. Technical report.

[3] Luis A. Santaló. *Integral Geometry and Geometric Probability*. Addison-Wesley, 1976.

[4] J. M. Steele. Growth rates of euclidean minimal spanning trees with power weighted edges. *AnnProb*, 16:1767–1787, 1988.

[5] J. M. Steele. Cost of sequential connection for points in space. *Operations Research Letters*, 8:137–142, 1989.

[6] J. M. Steele and T. L. Snyder. Worst-case growth rates of some classical problems of combinatorial optimization. *SIAM J. Comput.*, 18:278–287, 1989.

[7] Shu-Ju Tu and Ephraim Fischbach. A new geometric probability technique for an n-dimensional sphere and its applications to physics. *arXiv.org Mathematical Physics*, 2000.