TR-88-012

AN UPPER BOUND FOR THE AVERAGE LENGTH OF THE EUCLIDEAN MINIMUM SPANNING TREE

TR-88-012

An Upper Bound for the Average Length of the Euclidean Minimum Spanning Tree

Ning-Yang Baby Wang* and Ruei-Chuan Chang**

* Ning-Yang Baby Wang is with the Institute of Information Science, Academia Sinica, Nankang, Taipei, Taiwan, Republic of China.

** Ruei-Chuan Chang is with the Institute of Computer Engineering, National Chiao Tung University, Hsinchu, Taiwan and Institute of Information Science, Academia Sinica, Nankang, Taipei, Taiwan, Republic of China.

Correspondence Address:

Professor Ruei-Chuan Chang
Institute of Information Science
Academia Sinica
Nankang, Taipei, Taiwan
Republic of China 11529

Abstract:

In this paper we estimate the average length of an Euclidean minimum spanning tree. If the vertices of an EMST are uniformly distributed on a unit disk in R^2, we can derive the behavior of the edge lengths of the EMST constructed from these vertices by using the ordered statistics, and obtain an upper bound for the expected length of an EMST to be $\sqrt{\pi \cdot n/2} + o(\sqrt{n})$, where n is the number of vertices of the EMST. This estimation is much closer to the actual values than the previous results. Experiment values are also given.

CR Categories: F.2.2

KEY WORDS: EMST, ETSP, Average case analysis, ordered statistics, binomial distribution

1. Introduction

The Euclidean minimum spanning tree (EMST) problems can be defined as follows. Given n points in the plane, construct a spanning tree of minimum total length whose vertices are the given points. This paper considers the problem of estimating the expected length of an EMST. This is hard to do because the edge lengths are highly dependent on each other, but some lower and upper bound were derived in Marks[2] and Beardwood[1].

Let X be a random point which is uniformly distributed on a finite set $S \subset R^2$. Let $X_1, \cdots, X_n$ be a random sample from X, and T_n be the length of the EMST connecting $X_1, \cdots, X_n$, and let $c_n = E(T_n)$, the expected value of T_n. Marks[2] derived a lower bound for the average length of the Euclidean Travelling Salesperson problem (ETSP) (as well as for that of EMST) as

$$c_n \geq \frac{1}{2}\sqrt{A}\,\frac{n-1}{\sqrt{n}},$$

where A is the Lebesgue measure of S when $S \subset R^2$; whereas Beardwood[1] found that an upper bound for the average length of ETSP (also for that of EMST) is

$$c_n = 0.92\sqrt{n \cdot A} + o(\sqrt{n})$$

when $S \subset R^2$. In Supowit[4] some average and worst cases of the ETSP problem were considered when S is a unit square in R^2.

It seems that there is a huge gap between the upper bound and the lower bound for the EMST. We hereby get an upper bound for c_n as

$$c_n = 0.707\sqrt{n \cdot A} + o(\sqrt{n}) ,$$

where S is confined to be a unit disk in R^2.

This paper is organized as follows. In section 2 we give a definite model of our problem and the statement of the main theorem we have reached. The proof of the main theorem is in section 3. The concluding remarks are given in section 4.

2. Model of the Problem

For convenience of computation, we confine S, the set which the vertices of the EMST are from, to be the unit disk $D = \{ p \in R^2 : |p| \leq 1 \}$. Assume that X is a random point which is uniformly distributed on D, and $(X_1, \cdots, X_n)$ is an independent random sample from X; then we can derive a closer upper bound of c_n as stated in the following theorem.

[Theorem 1]:

$$c_n = 0.707\sqrt{n \cdot A} + o(\sqrt{n}),$$

where S is a unit disk in R^2.

The proof of theorem 1 will be given in the next section.

3. Proof of the main theorem

In this section we give the proof of theorem 1. This requires some preparatory work. First of all, we discuss the behavior of the length of an edge that connects two points with one inside a circle and the other, outside.

[Lemma 2]:

Let X be a random point uniformly distributed on a disk of radius $a \in R$ satisfying $a > 0$, and p be a point with distance b from the center of the disk and $b \geq a$ as shown in Fig. 1. Let random variable Z be the distance between X and p; then the cumulative distribution function $F(a,b;\cdot)$ and the probability density function $f(a,b;\cdot)$ of Z are

$$F(a,b;z)=\begin{cases} 0 & z<b-a \\ \frac{1}{\pi}\left[\cos^{-1}\left(\frac{a^2+b^2-z^2}{2ab}\right)+\frac{z^2}{a^2}\cos^{-1}\left(\frac{b^2+z^2-a^2}{2bz}\right)-\frac{b}{a}\sqrt{1-\left(\frac{a^2+b^2-z^2}{2ab}\right)^2}\right], & b-a\leq z<b+a \\ 1 & b+a\leq z \end{cases}$$

$$\cdots\cdots(1)$$

and

$$f(a,b;z)=\begin{cases}\frac{2z}{\pi a^2}\cos^{-1}\left(\frac{z^2+b^2-a^2}{2bz}\right), & \text{when } b\text{-}a\le z\le b+a, \\ 0, & \text{otherwise}\end{cases} \quad\cdots\cdots(2)$$

respectively.

<proof>:

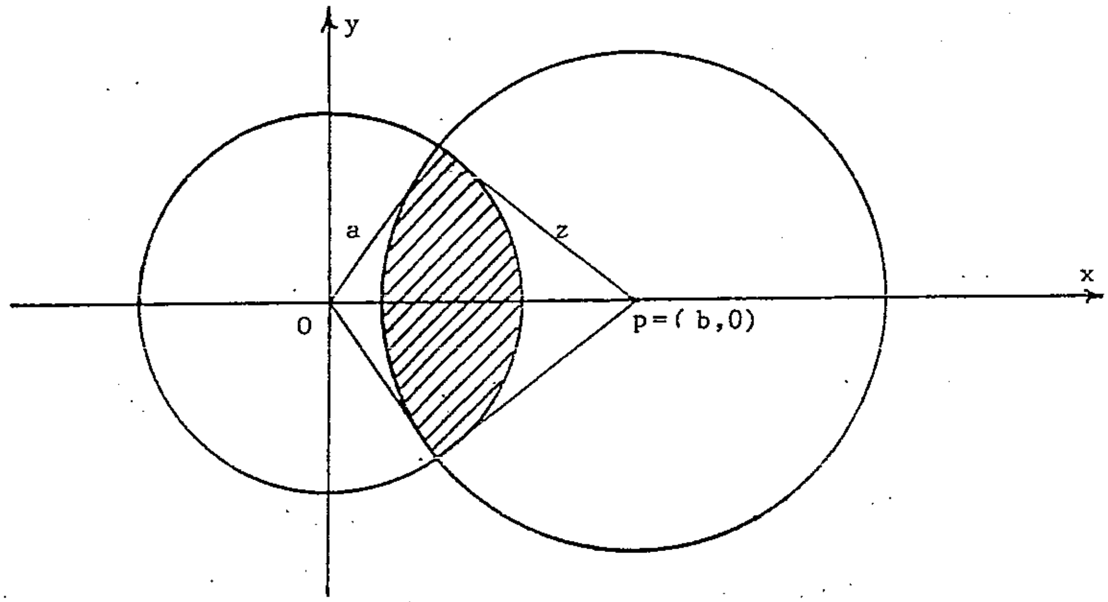

Fig. 1

The probability that the event $Z\le z$ occurs, $Pr(Z\le z)$, is the ratio of the area of the shaded region in Fig. 1 to the area of the original disk πa^2 because X is uniformly distributed in the disk; therefore we have F(a,b;z) as (1) and $f(a,b;z) = \frac{d}{dz}F(a,b;z)$ as (2) (see appendix 1).

Q.E.D.

The expected value of the edge length in lemma 2 could be bounded as in the following lemma. We hereby introduce a symbol u_n to replace $\sqrt{(n-1)/n}$ for brevity in the remaining part of this paper.

[Lemma 3]:

Let n be any integer greater than 1. For every integer m satisfying $1 \leq m < n$, define $X_1, \cdots, X_m$ the m i.i.d. random points which are uniformly distributed in B, and Y a random point uniformly distributed in A as in Fig. 2. Define random variables $Z_i = |X_i - Y|$, $i = 1, \cdots, m$, where $|\cdot|$ is the Euclidean norm, and $Z_1^* \leq Z_2^* \leq \cdots \leq Z_m^*$ the ordered statistics of $Z_1, \cdots, Z_m$. Then the expected value of Z_1^* has the following property

$$E(Z_1^*) < \int_{1-u_n}^{1+u_n} [1 - F(u_n, 1; t)]^m dt + \frac{1}{n}, \qquad \cdots\cdots(3)$$

where $F(u_n, 1; t)$ is as derived in lemma 2.

<Proof>:

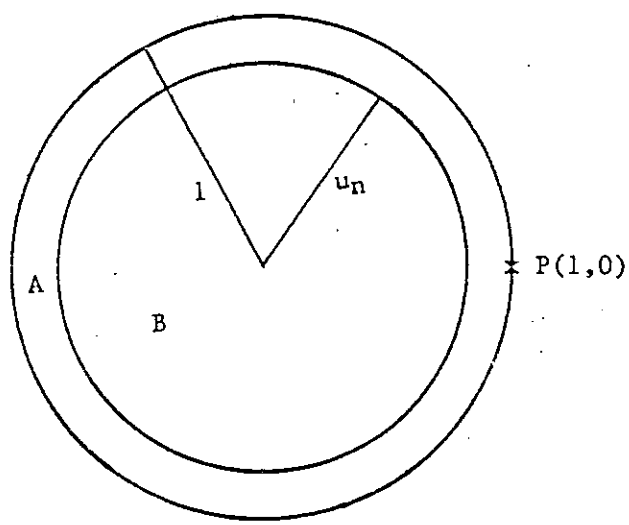

Fig. 2

Let P be the point with Cartesian coordinates (1,0); define random variable $Z_{i,p} = |X_i - P|$, $i = 1, \cdots, m$, where $|\cdot|$ is the Euclidean norm, and $Z_{1,p}^* \leq Z_{2,p}^* \leq \cdots \leq Z_{m,p}^*$ the ordered statistics of $Z_{1,p}, \cdots, Z_{m,p}$; then

$$E(Z_1^*) < E(Z_{1,p}^*).$$

By lemma 2, we have

$$E(Z^*_{1,p}) = \int_{1-u_n}^{1+u_n} t\cdot m\cdot[1-F(u_n,1;t)]^{m-1}\cdot f(u_n,1;t)dt$$

$$= 1 - u_n + \int_{1-u_n}^{1+u_n}[1-F(u_n,1;t)]^m dt\ ;$$

therefore

$$E(Z^*_1) < \int_{1-u_n}^{1+u_n}[1-F(u_n,1;t)]^m dt + \frac{1}{n}.$$

Q.E.D.

The integral in (3) is not simple, since there are unknown parameters in it. The following lemma will give a terser experssion which will not affect the result.

[**Lemma 4**]:

Define $F:(0\ ,2)\longrightarrow R$,

$$F(x) = \frac{1}{\pi}\cdot\left[\cos^{-1}(1-\frac{x^2}{2}) + x^2\cos^{-1}(\frac{x}{2}) - \sqrt{x^2-\frac{x^4}{4}}\right]$$

which is limiting function of $F(u_n,1;x)$ as n approaches to infinity; then

$$\int_{1-u_n}^{1+u_n}[1-F(u_n,1;t)]^m dt < \int_0^2[1-F(x)]^m dx + \frac{6}{n},\cdots\cdots\cdots\cdots\cdots(4)$$

for all integer $n \geq 2$ and all m satisfying $1 \leq m < n$.

<Proof>:

For all integer $n\geq 2$, define $g:[n\ ,\infty)\longrightarrow R$,

$$g(r) = \int_{1-u_r}^{1+u_r}[1-F(u_r,1;x)]^m dx\ .$$

Applying the Leibniz's rule we have

$$\frac{d}{dr}g(r) = \left[\frac{d}{dr}(1+u_r)\right]\cdot[1-F(u_r,1;1+u_r)] - \left[\frac{d}{dr}(1-u_r)\right]\cdot[1-F(u_r,1;1-u_r)]$$

$$- m\cdot\int_{1-u_r}^{1+u_r}[1-F(u_r,1;x)]^{m-1}\cdot\left[\frac{\partial}{\partial r}F(u_r,1;x)\right]dx$$

$$= \frac{1}{2r\sqrt{r(r-1)}} - m\cdot\int_{1-u_r}^{1+u_r}[1-F(u_r,1;x)]^{m-1}\cdot\left[\frac{\partial}{\partial r}F(u_r,1;x)\right]dx\ ,$$

where

$$\frac{\partial}{\partial r}F(u_r,1;x) = \frac{1}{\pi(r-1)^2}\left[\frac{1}{2}\cdot\sqrt{-x^4+(4-\frac{2}{r})x^2-\frac{1}{r^2}} - x^2\cos^{-1}(\frac{x}{2}+\frac{1}{2xr})\right]$$

$$= \frac{1}{2r(r-1)}\left[\sqrt{-x^4+(4-\frac{2}{r})x^2-\frac{1}{r^2}}\Big/\left[2x\cdot\cos^{-1}(\frac{x}{2}+\frac{1}{2xr})\right]-x\right]\cdot f(u_r,1;x)\ ,$$

so

$$\left|\frac{d}{dr}g(r)\right| < \frac{1}{2(r-1)^2}\left[1+\int_{1-u_r}^{1+u_r}\left|x-\sqrt{-x^4+(4-\frac{2}{r})x^2-\frac{1}{r^2}}\Big/\left[2x\cos^{-1}(\frac{x}{2}+\frac{1}{2xr})\right]\right|\right.$$

$$\left.\cdot[1-F(u_r,1;x)]^{m-1}\cdot m\cdot f(u_r,1;x)dx\right]\ ,$$

since $\forall\ r > 1$, $\forall\ x \in (1-u_r\ ,\ 1+u_r)$, it can be shown that

(see appendix 2)

$$\left|x-\sqrt{-x^4+(4-\frac{2}{r})x^2-\frac{1}{r^2}}\Big/\left[2x\cdot\cos^{-1}(\frac{x}{2}+\frac{1}{2xr})\right]\right| < 2\ ;$$

thus

$$\left|\frac{d}{dr}g(r)\right| < \frac{1}{2(r-1)^2}\left[1+2\int_{1-u_r}^{1+u_r}[1-F(u_r,1;x)]^{m-1}\cdot m\cdot f(u_r,1;x)dx\right]$$

$$= \frac{1}{2(r-1)^2}(1+2) \le \frac{6}{r^2}\ .$$

This provides that $\lim_{r\to\infty} g(r)$ exists, and

$$\left|g(r) - \lim_{r\to\infty} g(r)\right| \le \int_r^\infty \left|\frac{d}{dr'}g(r')\right|dr'$$

$$< \int_r^\infty \frac{6}{(r')^2}\,dr' = \frac{6}{r}\ ,$$

hence

$$g(r) < \lim_{r\to\infty} g(r) + \frac{6}{r}\ ,$$

and

$$\int_{1-u_n}^{1+u_n}[1-F(u_n,1;x)]^m dx < \int_0^2 [1-F(x)]^m dx + \frac{6}{n}\ .$$

Q.E.D.

The right–hand side of (4) is simple enough to evaluate. In the next lemma we will give the limiting value of the right–hand side of (4) as m approaches to infinity.

[Lemma 5]:

$$\lim_{m\to\infty} \sqrt{m}\cdot\int_0^2 [1-F(x)]^m dx = \sqrt{\frac{\pi}{2}},$$

where F is as in lemma 4.

<Proof>:

Let $k_m = \int_0^2 [1-F(x)]^m dx$ and $\theta = 2\cos^{-1}(\frac{x}{2})$; then

$$k_m = \int_0^\pi \sin(\tfrac{\theta}{2})\left[\frac{\sin\theta - \theta\cdot\cos\theta}{\pi}\right]^m d\theta .$$

For all $\theta \in (\frac{\pi}{2}, \pi)$, we have

(see appendix 3)

$$-\pi\cdot\cos\theta < \sin\theta - \theta\cdot\cos\theta ;$$

thus

$$k_m = \int_0^{\pi/2} \sin(\tfrac{\theta}{2})\left[\frac{\sin\theta - \theta\cdot\cos\theta}{\pi}\right] d\theta + \int_{\pi/2}^{\pi} \sin(\tfrac{\theta}{2})(-\cos\theta)^m d\theta$$

$$> \int_{\pi/2}^{\pi} \sin(\tfrac{\theta}{2})(-\cos\theta)^m d\theta ;$$

replace θ by $2\phi_1+\frac{\pi}{2}$; we have

$$k_m > 2^{m+0.5}\cdot\int_0^{\pi/4} (\sin\phi_1 + \cos\phi_1)\cdot\sin^m\phi_1\cdot\cos^m\phi_1\cdot d\phi_1$$

$$= 2^{m+0.5}\cdot\int_0^{\pi/2} \sin^{m+1}\phi_2\cdot\cos^m\phi_2\cdot d\phi_2$$

$$= 2^{m-0.5}\cdot\int_0^{\pi/2} 2\cdot\sin^{m+1}\phi_2\cdot\cos^m\phi_2\cdot d\phi_2$$

The integral of the last expression is the beta function of parameters $\frac{m}{2}+1$ and $\frac{m+1}{2}$, so we have

$$k_m \geq 2^{m-0.5}\cdot\frac{\Gamma(\frac{m}{2}+1)\cdot\Gamma(\frac{m+1}{2})}{\Gamma(m+\frac{3}{2})},$$

by the duplication formula[3] and Stirling's formula,

$$k_m \geq \frac{1}{\sqrt{m}}\left(\sqrt{\frac{\pi}{2}} - \frac{c_1}{m}\right) \qquad \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots (5)$$

where c_1 is a positive constant.

On the other hand, $\forall$ w $\in (0, 1)$, since $\forall\ \varphi_1 \in (0, \frac{\pi}{2})$, we have (see appendix 4)

$$\cot\varphi_1 - \frac{\pi}{2} + \varphi_1 > 0$$

and

(see appendix 5)

$$\lim_{\varphi_1 \to \frac{\pi}{2}^-} \frac{\cot\varphi_1 - \pi/2 + \varphi_1}{\cos^2\varphi_1} = 0 ,$$

for $\frac{\pi(1-w)}{2} > 0$, there exists $\theta_w \in (\frac{\pi}{2}, \pi)$ such that $\forall\ \varphi_1 \in (\theta_w - \frac{\pi}{2}, \frac{\pi}{2})$,

$$0 < \frac{\cot\varphi_1 - \pi/2 + \varphi_1}{\cos^2\varphi_1} < \frac{\pi(1-w)}{2},$$

and

$$\cot\varphi_1 + \varphi_1 + \frac{\pi}{2} < \pi\cdot\left[1 + \left(\frac{1-w}{2}\right)\cdot\cos^2\varphi_1\right]$$

(see appendix 6)

$$< \pi\cdot(1 - \cos^2\varphi_1)^{\frac{w-1}{2}}$$

$$< \pi\cdot\sin^{w-1}\varphi_1$$

so

$$\cos\varphi_1 + (\varphi_1 + \frac{\pi}{2})\sin\varphi_1 < \pi\cdot\sin^w\varphi_1 ;$$

let $\varphi_2 = \varphi_1 + \frac{\pi}{2}$, then $\forall\ \varphi_2 \in (\theta_w, \pi)$,

$$\sin\varphi_2 - \varphi_2\cdot\cos\varphi_2 < \pi\cdot(-\cos\varphi_2)^w ;$$

thus $\forall$ m ,

$$k_m \leq \int_0^{\theta_w} \sin(\frac{\theta}{2})\left[\frac{\sin\theta - \theta\cos\theta}{\pi}\right]^m d\theta + \int_{\theta_w}^{\pi} \sin(\frac{\theta}{2})\cdot(-\cos\theta)^{mw} d\theta$$

$$< c_w^m + \sqrt{\frac{\pi}{2mw}} ,$$

where $c_w \in (0,1)$ is independent on m ; therefore $\forall\ \epsilon > 0$,$\exists$ M, $\forall$ m $\geq$ M ,

(see appendix 7)

$$k_m < \frac{1}{\sqrt{m}} \cdot \left[\sqrt{\frac{\pi}{2}} + \epsilon \right] \qquad \cdots\cdots\cdots\cdots\cdots\cdots (6)$$

Combine (5) and (6), we have

$$\lim_{m \to \infty} \sqrt{m} \cdot k_m = \sqrt{\frac{\pi}{2}} .$$

Q.E.D.

Now we begin the proof of theorem 1.

<Proof of theorem 1>:

Assume that the average length of the EMST in the unit disk has an upper bound $c\sqrt{n}$, where n is the number of vertices of the spanning tree and c is a constant independent on n.

As in Fig.2, the area of A is $\frac{\pi}{n}$. Since $X_1, \cdots, X_n$ are uniformly distributed on this disk, the probability of a point fallen in A is the ratio of the area of A to the area of the disk. Let random variable Z_n be the number of the X_i's fallen in A; then Z_n has the binomial distribution $B(n, \frac{1}{n})$. Now we divide the problem to the following cases:

[case 1: $Z_n = 0$]:

The probability that this case occurs is $P_r(Z_n = 0) = (\frac{n-1}{n})^n$, and all X_i's are in B. Therefore the length of the EMST has an upper bound $c\sqrt{n} \cdot \sqrt{\frac{n-1}{n}}$ because the edge length is proportional to the radius of the disk.

[case 2: $Z_n = n$]:

The Probability that this case occurs is $P_r(Z_n = n) = (\frac{1}{n})^n$, and all X_i's are in A. Thus the length will never exceed

(see appendix 8)

$$2\pi + n \cdot (1 - \sqrt{\frac{n-1}{n}}) < 2\pi + 1 .$$

[case 3: $1 \le Z_n < n$]:

The probability that this case occurs is $P_r(Z_n = k) = \binom{n}{k}\frac{(n-1)^{n-k}}{n^n}$, and there are (n–k) X_i's in B and k X_i's in A. From the previous result, the length has an upper bound

(see appendix 9)

$$c\cdot\sqrt{(n-k)}\sqrt{\frac{n-1}{n}} + k\cdot\left[\sqrt{\frac{\pi}{2(n-k)}} + \frac{h(n-k)}{\sqrt{n-k}}\right],$$

where $h(n) = o(n)$.

Summing up the above three cases, we have

$$c\sqrt{n} \le c\cdot\left(\frac{n-1}{n}\right)^n\cdot\sqrt{n-1} + \frac{2\pi+1}{n^n} +$$
$$\sum_{k=1}^{n-1}\binom{n}{k}\frac{(n-1)^{n-k}}{n^n}\cdot\left[c\cdot\sqrt{\frac{(n-1)(n-k)}{n}} + k\left[\sqrt{\frac{\pi}{2(n-k)}} + \frac{h(n-k)}{\sqrt{n-k}}\right]\right],$$

and

$$c \le \frac{p(n)}{q(n)},$$

where

$$p(n) = \sqrt{\frac{\pi}{2}}\cdot\sum_{k=1}^{n-1}\binom{n}{k}\frac{(n-1)^{n-k}}{n^n}\cdot\frac{k}{\sqrt{n-k}} + \frac{2\pi+1}{n^n} + \sum_{k=1}^{n-1}\binom{n}{k}\frac{(n-1)^{n-k}}{n^n}\cdot k\cdot\frac{h(n-k)}{\sqrt{n-k}}$$

and

$$q(n) = \sqrt{n} - \sum_{k=0}^{n-1}\binom{n}{k}\frac{(n-1)^{n-k}}{n^n}\sqrt{\frac{(n-1)(n-k)}{n}}.$$

We can rewrite q(n) as

$$q(n) = \frac{1}{\sqrt{n}}\cdot\left[n - (n-1)\cdot\sum_{k=0}^{n}\binom{n}{k}\frac{(n-1)^{n-k}}{n^n}\sqrt{\frac{n-k}{n-1}}\right],$$

because

(see appendix 10)

$$1 - \frac{k}{n-1} < \sqrt{\frac{n-k}{n-1}} < 1 + \frac{1}{2(n-1)}, \text{ for } k = 0,1,\cdots,n$$

we have

$$\sum_{k=0}^{n}\binom{n}{k}\frac{(n-1)^{n-k}}{n^n}\left(1-\frac{k}{n-1}\right) < \sum_{k=0}^{n}\binom{n}{k}\frac{(n-1)^{n-k}}{n^n}\sqrt{\frac{n-k}{n-1}} < \sum_{k=0}^{n}\binom{n}{k}\frac{(n-1)^{n-k}}{n^n}\left(1+\frac{1}{2(n-1)}\right)$$

and

$$1-\frac{1}{n-1}<\sum_{k=0}^{n}\binom{n}{k}\frac{(n-1)^{n-k}}{n^n}\sqrt{\frac{n-k}{n-1}}<1+\frac{1}{2(n-1)},$$

so we have

$$\lim_{n\to\infty}\sqrt{n}\cdot q(n)=1\,. \qquad \cdots\cdots(7)$$

On the other hand, because

(see appendix 11)

$$1+\frac{k}{2n}<\sqrt{\frac{n}{n-k}}<1+\frac{k^2}{n}\text{, for } k=1,2,\cdots,n\,,$$

we also have

$$\sum_{k=1}^{n-1}\binom{n}{k}\frac{(n-1)^{n-k}}{n^n}\cdot k\cdot(1+\frac{k}{2n})<\sum_{k=1}^{n-1}\binom{n}{k}\frac{(n-1)^{n-k}}{n^n}\cdot k\cdot\sqrt{\frac{n}{n-k}}<\sum_{k=1}^{n-1}\binom{n}{k}\frac{(n-1)^{n-k}}{n^n}\cdot k\cdot(1+\frac{k^2}{n}),$$

so

$$\lim_{n\to\infty}\sqrt{n}\cdot\sum_{k=1}^{n-1}\binom{n}{k}\frac{(n-1)^{n-k}}{n^n}\cdot\frac{k}{\sqrt{n-k}}=1\,,$$

which implies that the last two terms in p(n) are negiligible; therefore

$$\lim_{n\to\infty}\sqrt{n}\cdot p(n)=\sqrt{\frac{\pi}{2}}\,. \qquad \cdots\cdots(8)$$

Combine (7) and (8), we reach

$$\lim_{n\to\infty}\frac{p(n)}{q(n)}=\sqrt{\frac{\pi}{2}},$$

hence

$$c\le\frac{p(n)}{q(n)}=\sqrt{\frac{\pi}{2}}+o(1)\,,$$

and

$$c\sqrt{n}=\sqrt{\frac{\pi}{2}\cdot n}+o(\sqrt{n}).$$

Q.E.D.

4. Concluding remarks

In order to compare the upper bound we derived with the actual length of the EMST, we generated 300 samples for each $n = 2^2, 3^2, 4^2, \cdots, 20^2$ and computed its average length. The results are shown in Table 1.

The relative error is about –6.437% eventually; this confirms that the upper bound we derived is reasonably good. On the other hand, this upper bound might not be the actual expected length because, as in the proof of theorem 1, any point (if exists) in set A might connect with more than one point in B; this means that the tree we construct is not "minimum", and this upper bound could be reduced by further work.

Table 1. Comparison between theoretic and experiment values:

Number of points n	Theoretic value $\sqrt{\pi \cdot n/2}$	Experiment value θ_n	Coefficient $\theta_n/\sqrt{\pi \cdot n}$	Relative error $(\theta_n-\sqrt{\pi \cdot n/2})/\sqrt{\pi \cdot n/2}$
4	2.5066	1.9232	0.54252	−23.276%
9	3.7599	3.4608	0.65085	− 7.957%
16	5.0133	4.6758	0.65951	− 6.732%
25	6.2666	6.0126	0.67845	− 4.052%
36	7.5199	7.1721	0.67440	− 4.625%
49	8.7732	8.3940	0.67654	− 4.323%
64	10.0265	9.5114	0.67078	− 5.137%
81	11.2798	10.7081	0.67126	− 5.069%
100	12.5331	11.8994	0.67135	− 5.056%
121	13.7865	13.0467	0.66916	− 5.366%
144	15.0398	14.1862	0.66698	− 5.675%
169	16.2931	15.3320	0.66540	− 5.899%
196	17.5464	16.5315	0.66621	− 5.784%
225	18.7997	17.6946	0.66554	− 5.878%
256	20.0530	18.8232	0.66374	− 6.133%
289	21.3063	19.9954	0.66360	− 6.153%
324	22.5597	21.1560	0.66311	− 6.222%
361	23.8130	22.3074	0.66240	− 6.323%
400	25.0663	23.4527	0.66159	− 6.437%

Reference:

[1] J.H.H Beardwood, J.H. Halton and J.M. Hammersley, The shortest path through many points, Proc. Cambridge Phil. Soc.,Vol. 55(1959), pp. 299–327.

[2] E.S. Marks, A lower bound for the expected travel among m random points. Ann. Math. Statistics. Vol. 19(1948), pp. 419–422.

[3] Murray R. Spiegel, Mathematical handbook of formulas and tables, McGraw–Hill Inc., 1968.

[4] K. Supowit, E.M. Reingold and D.A. Plasted, The traveling salesman problem and minimum matching in the unit square, SIAM J. Comput. Vol. 12, No.1(1983), pp. 144–156.

List of figure captions:

[Figure 1]:

We have a disk with center the origin and radius a and a point p having distance b from the origin with $b \geq a$. No loss the generality, we can assume that the coordinate of p is (b,0). If a point X in the original disk has distance z from p, then the set of all such X is the shaded region on this figure.

[Figure 2]:

Given n points, we divide the unit disk into two regions by drawing a circle with center the same as the unit disk and radius $u_n = \sqrt{(n-1)/n}$ in order to obtain the behavior of the length of an edge that connects two points with one inside the circle and the other, outside.

Appendix

1.

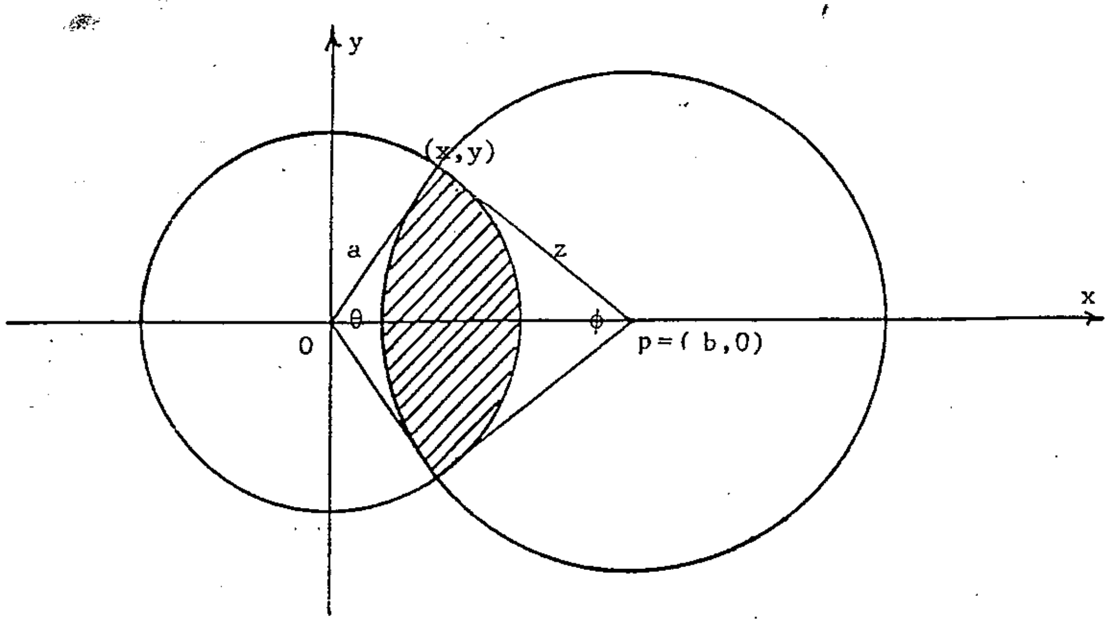

Fig. 1′

The two angles θ and ϕ can be derived as

$$\theta = 2\cdot\cos^{-1}(\frac{a^2+b^2-z^2}{2ab})$$

and

$$\phi = 2\cdot\cos^{-1}(\frac{b^2+z^2-a^2}{2bz})$$

and the y–coordinate of the point (x,y) is

$$y = \sqrt{a^2-(\frac{a^2+b^2-z^2}{2b})^2},$$

so the area of the shaded region is

$$\pi a^2\cdot\frac{\theta}{2\pi} + \pi b^2\cdot\frac{\phi}{2\pi} - b\cdot y$$

$$= a^2\cdot\cos^{-1}(\frac{a^2+b^2-z^2}{2ab}) + z^2\cdot\cos^{-1}(\frac{b^2+z^2-a^2}{2bz}) - ab\cdot\sqrt{1-(\frac{a^2+b^2-z^2}{2ab})^2}.$$

2. Because $\sqrt{-x^4 + (4-\frac{2}{r})x^2 - \frac{1}{r^2}} \Big/ \left[2x\cdot\cos^{-1}(\frac{x}{2}+\frac{1}{2xr})\right]$

$$= \sqrt{1-(\frac{x}{2}+\frac{1}{2xr})^2} \Big/ \cos^{-1}(\frac{x}{2}+\frac{1}{2xr});$$

let

$$\frac{d}{dx}\left(\frac{x}{2}+\frac{1}{2xr}\right)=\frac{1}{2}-\frac{1}{2x^2r}=0\ ;$$

it is evident that $(\frac{x}{2}+\frac{1}{2xr})$ has minimum value $\sqrt{\frac{1}{r}}$ when $x=\sqrt{\frac{1}{r}}$ and has maximum value 1 when $x = 1-u_r$ or $1+u_r$; therefore we can define $w = \cos^{-1}(\frac{x}{2}+\frac{1}{2xr})$ and

$$\sqrt{1-\left(\frac{x}{2}+\frac{1}{2xr}\right)^2}\Big/\cos^{-1}\left(\frac{x}{2}+\frac{1}{2xr}\right)=\frac{\sqrt{1-\cos^2 w}}{w}=\frac{\sin w}{w}$$

which is in the interval (0,1) and since $x \in (1-u_r,1+u_r) \subset (0,2)$, we have

$$\left|x-\sqrt{-x^4+(4-\frac{2}{r})x^2-\frac{1}{r^2}}\Big/\left[2x\cdot\cos^{-1}\left(\frac{x}{2}+\frac{1}{2xr}\right)\right]\right|<2\ .$$

3. It is well–known that $\forall\ \alpha \in (0,\frac{\pi}{2})$, $\alpha < \tan\alpha$. Replace α by $\pi-\theta$; then $\forall\ \theta \in (\frac{\pi}{2},\pi)$,

$$\pi-\theta<\tan(\pi-\theta)=-\frac{\sin\theta}{\cos\theta}$$

and

$$(\pi-\theta)(-\cos\theta)<\sin\theta$$

because $(-\cos\theta)>0\ \forall\theta \in (\frac{\pi}{2},\pi)$, hence

$$-\pi\cdot\cos\theta<\sin\theta-\theta\cdot\cos\theta\ .$$

4. As in appendix 3, $\forall\ \varphi_1 \in (0,\frac{\pi}{2})$, we have

$$\varphi_1<\tan\varphi_1\ ,$$

so

$$\frac{\pi}{2}-\varphi_1<\tan(\frac{\pi}{2}-\varphi_1)=\cot\varphi_1\ ;$$

hence

$$\cot\varphi_1-\frac{\pi}{2}+\varphi_1>0\ .$$

5. We want to prove that $\lim\limits_{\varphi_1\to\frac{\pi}{2}^-}\frac{\cot\varphi_1-\pi/2+\varphi_1}{\cos^2\varphi_1}=0$. Because the numerator and the denominator both approach to 0 as φ_1 approaches to $\frac{\pi}{2}^-$, we could apply L'Hospital $\frac{0}{0}$ rule and find that

$$\lim_{\varphi_1 \to \frac{\pi}{2}^-} \frac{\cot\varphi_1 - \pi/2 + \varphi_1}{\cos^2\varphi_1}$$

$$= \lim_{\varphi_1 \to \frac{\pi}{2}^-} \frac{-1-\cot^2\varphi_1+1}{-2\cdot\sin\varphi_1\cdot\cos\varphi_1}$$

$$= \lim_{\varphi_1 \to \frac{\pi}{2}^-} \frac{\cos\varphi_1}{2\cdot\sin^3\varphi_1} = 0\ .$$

6. From the formula of the binomial expansion, we have

$$(1-\cos^2\varphi_1)^{\frac{w-1}{2}}$$

$$= 1 + \frac{w-1}{2}(-\cos^2\varphi_1) + \frac{w-1}{2}\cdot\frac{w-3}{2}(-\cos^2\varphi_1)^2 + \cdots$$

$$> 1 + \frac{1-w}{2}\cos^2\varphi_1$$

because $w \in (0,1)$.

7. $\forall\, w \in (0,1)$, $\forall\, m$, we have

$$k_m < c_w^m + \sqrt{\frac{\pi}{2mw}}\,,$$

where $c_w \in (0\ ,\ 1)$ is independent on m ; therefore $\forall\ \epsilon > 0$, choose $w \in (0,1)$ such that

$$\sqrt{\frac{\pi}{2}}\left(\sqrt{\frac{1}{w}} - 1\right) < \frac{\epsilon}{2}$$

and for the ϵ and the associated c_w , the inequality

$$\sqrt{m}\cdot c_w^m < \frac{\epsilon}{2}$$

holds for sufficiently large m because $c_w \in (0,1)$. Thus

$$\sqrt{m}\left(k_m - \sqrt{\frac{\pi}{2}}\right) < \sqrt{m}\cdot c_w^m + \sqrt{\frac{\pi}{2}}\left(\sqrt{\frac{1}{w}} - 1\right) < \epsilon$$

and hence

$$k_m < \frac{1}{\sqrt{m}}\cdot\left[\sqrt{\frac{\pi}{2}} + \epsilon\right].$$

8. The first term of $\left[2\pi + n\cdot(1-\sqrt{\frac{n-1}{n}})\right]$ is the perimeter of the unit disk, and the

expression in the parentheses of the second term is the width of the region A in figure 2. This means that we can use the sum of the preimeter of the unit disk and n times of the width of A to bound the total length of the EMST whose vertices are uniformly distributed on A.

9. The first term of $\left[c\cdot\sqrt{(n-k)}\sqrt{\frac{n-1}{n}} + k\cdot\left[\sqrt{\frac{\pi}{2(n-k)}} + \frac{h(n-k)}{\sqrt{n-k}}\right]\right]$ is the upper bound of the average length of the EMST whose (n–k) vertices are uniformly distributed on the region B in the figure 2, and the expression in the parentheses of the second term is the upper bound of the expected value of the smallest edge length that connects one of the (n–k) points uniformly distributed in B and a point uniformly distributed in A, as derived in the previous lemmas. Therefore we can bound the total length of this kind of EMST by the above expression.

10. It can be derived by the binomial expansion.

11. It can be derived by the binomial expansion.