LINDA HALL LIBRARY
5109 CHERRY STREET
KANSAS CITY, MISSOURI 64110-2498
PHONE (816) 363-4600
FAX: (816) 926-8785

10/27/11DocServ #: 654321
11:23

SHIP TO:

University of Michigan/ILL
106 Hatcher Graduate Library
913 South University Ave.
Ann Arbor, MI 48109-1205

EYM
Patron: Pettie, Seth

Odyssey: 141.211.175.136
Phone:
Ariel: 141.211.175.21

Regular

Odyssey

LHL

CCG

Max Cost: $26.00IFM

Notes:

Shelved as: X TK3226 .A472 1977

Location: LHL Closed Stacks

Title: Proceedings ; Fifteenth annual Allerton Conference on Communication, Control, and Computing ; conference held September 28-30, 1977, Allerton House, M

Volume:

Issue:

Date: 1977

Article/Paper: Papadimitriou, C.; The probabilistic analysis of matching heuristics

Pages: 368-378

OCLC #: 30630139
ISSN:

Lender string: CNS,*LHL,MUQ,MUQ,UMV

IL: 83379021

DOCSERV / OCLC / PULL SLIP

11

THE PROBABILISTIC ANALYSIS OF MATCHING HEURISTICS*

CHRISTOS H. PAPADIMITRIOU
Center for Research on Computing Technology
Harvard University
Cambridge, Massachusetts 02138

ABSTRACT

It can be shown that the total length of the optimum matching of a set of n points uniformly distributed on the unit square--or any other Lebesgue area of measure 1 --is, when divided by $\sqrt{n}$, almost certainly equal to a constant $\mu \geq .25$. We analyze probabilistically several heuristics for the matching problem and obtain upper bounds on the value of μ. We show that $\mu \leq .40106$, and conjecture that $\mu \simeq .35$.

1. INTRODUCTION

Many combinatorial optimization problems call for the construction of the shortest possible network of some kind, given the matrix of distances among a set of n points. These problems include the traveling salesman problem (TSP), the minimum spanning tree problem, Steiner's problem, the matching problem, etc. (See [18] for definitions.) When the given points are actually realized as points on the plane--and the distance matrix is thus induced by the two-dimensional Euclidean metric--we obtain the Euclidean case of such a problem. Different problems, however, behave differently under this restriction. The minimum spanning tree problem becomes easier [12], but the TSP and Steiner's problem remain hard [5, 9]. There is no obvious advantage to the matching problem [11], and the complexity of another problem is reduced from probably exponential to $O(n \log n)$ [11].

For these Euclidean cases of "network design" problems a very strong probabilistic result is available [2]. In the case of the traveling salesman problem this result essentially asserts that the value of the optimal tour of n points drawn from a uniform distribution in the unit square is almost certainly equal to a constant β times $\sqrt{n}$. (See Theorem 1 of Section 2 for a precise statement.) This result is surprisingly stable when the assumptions are relaxed significantly. In particular, it holds for arbitrary measurable regions with arbitrary probability distributions, and can be generalized to many dimensions. Furthermore, it is also valid for the minimum spanning tree problem, Steiner's problem, and, in fact, any network design-type of problem as long as it satisfies four conditions stated in our Section 2. In particular, it holds for the matching problem [3, 8] (Corollary 2).

Recently, the result of [2] received some well-deserved attention from computer scientists and was used as the main argument in the probabilistic analysis of heuristics for the TSP [7] and other hard Euclidean problems [10]. However, heuristics such as the ones in [7, 10] are susceptible to evaluation of the solution yielded only relative to the optimum solution, and not in absolute terms (e.g., calculating the actual length of the derived solution). Hence, no information about the magnitude of the constant β in the theorem of [2] can be deduced by analyzing them. We also notice here that in contrast to the TSP, the matching problem (Euclidean or otherwise) can be solved exactly in $O(n^3)$ time by the algorithm of Edmonds [3] as implemented by Gabow and Lawler [4, 8]. The conceptual complexity of this algorithm, however, renders it beyond any detailed probabilistic analysis.

*The research reported here was supported by NSF Grant MCS 77-01193.

Determining the cons
In [2] the bounds $.61 \leq \beta \leq$
even an improvement on th
departure from the existin
estimate β to be .75; bas
estimates it to .765.

The value of the corr
is unknown as well. Since
shortest tour, we notice in
that $\mu \geq .25$. The last pa
upper bound using techniqu
TSP. For this purpose we
heuristics, most notably a
The constant corresponding
than .40106.

2. THE SHORTE

We will state the resu
and uniform distribution.
Introduction.

Let $P = (p_1, p_2, \ldots)$
sequences of points indepen
square. Let $P^n = \{p_1, \ldots$
of all the points in P^n.

THEOREM 1 [2]. W
exists and is equal to a cons

The proof of Theorem
TSP is shown to satisfy four
finite set of points in an are
$Q_1, \ldots, Q_m$ is the corresp
conditions:

1. $T(Q)$

where δ_{ij} is the dian

2. $T(Q)$

where $|\partial A_j|$ is the l

3. There exists a constan

4. There exists a constan

The second part of the
uses only properties (1-4) of
the argument appears in [11]

Let P be as in Theore
segments (p_{ik}, p_{jk}), $k = 1,$
Let $M(P^n)$ be the shortest

Determining the constant β is an interesting problem in its own right. In [2] the bounds $.61 \le \beta \le .92$ are proved. Explicit calculation of β --or even an improvement on the above bounds--appears to require a significant departure from the existing methodology. Monte Carlo experiments in [2] estimate β to be .75; based on more elaborate experiments, D. Stein [13] estimates it to .765.

The value of the corresponding constant for the matching problem, μ, is unknown as well. Since the optimum matching is at most a half of the shortest tour, we notice immediately that $\mu \le .5\beta \le .46$. We also show that $\mu \ge .25$. The last part of this paper concentrates on improving the .46 upper bound using techniques that are applicable to matching and not to the TSP. For this purpose we analyze probabilistically certain matching heuristics, most notably a three-phase extension of the "monotonic" heuristic. The constant corresponding to the three-phase heuristic is shown to be less than .40106.

2. THE SHORTEST PAIRING OF MANY POINTS

We will state the result of [2] for two dimensions, the unit square and uniform distribution. The possible extensions were sketched in the Introduction.

Let $P = (p_1, p_2, \dots)$ be a random variable with range infinite sequences of points independently and uniformly distributed in the unit square. Let $P^n = \{p_1, \dots, p_n\}$, and let $T(P^n)$ denote the shortest tour of all the points in P^n.

THEOREM 1 [2]. With probability 1, the limit $\lim_{n\to\infty} T(P^n)\cdot n^{-1/2}$ exists and is equal to a constant β. ■

The proof of Theorem 1 essentially consists of two parts. First, the TSP is shown to satisfy four simple combinatorial conditions. Let Q be a finite set of points in an area A, and let $A_1, \dots, A_m$ be a partition of A. $Q_1, \dots, Q_m$ is the corresponding partition of Q. T satisfies the following conditions:

1. $$T(Q) \le \sum_{j=1}^{m} (T(Q_j) + \delta_{j,j+1}) \quad ,$$

   where δ_{ij} is the diameter of $A_i \cup A_j$ and $\delta_{m,m+1} = \delta_{m1}$.

2. $$T(Q) \ge \sum_{j=1}^{m} (T(Q_j) - 2\cdot|\partial A_j|) \quad ,$$

   where $|\partial A_j|$ is the length of the boundary of A_j.

3. There exists a constant α such that $T(P^n) \le \alpha\sqrt{n}$.

4. There exists a constant γ such that $\mathscr{E}(T(P^n)) \ge \gamma\sqrt{n}$.

The second part of the proof involves a deep analytic argument which uses only properties (1-4) of the TSP. An elementary exposition of some of the argument appears in [11].

Let P be as in Theorem 1. A <u>matching</u> of P^n is a set of line segments (p_{i_k}, p_{j_k}), $k = 1, \dots, \lfloor n/2 \rfloor$ such that all i_k, j_m are distinct. Let $M(P^n)$ be the shortest such matching.

COROLLARY 2. With probability 1, the limit $\lim_{n\to\infty} M(P^n)$ exists and is equal to a constant μ.

Proof. We will prove properties (1-4) with T replaced by M. The corollary will then follow from the proof of Theorem 1 given in [2].

1. Let $k_1, \ldots, k_{q'}$ be the indices j for which $|Q_j|$ is odd. If q' is itself odd, consider $q = q'-1$; otherwise, $q = q'$. The optimal matchings of Q_j together with the segments $(p_{k_1}, p_{k_2}), \ldots, (p_{k_{q-1}}, p_{k_q})$ is a matching of Q, hence

$$M(Q) \le \sum_{j=1}^{m} M(Q_j) + \sum_{j=1}^{q/2} \delta_{k_{2j-1}, k_{2j}} \le \sum_{j=1}^{m} (M(Q_j) + \delta_{j,j+1}) .$$

2-3. Precisely as in Lemmas 1 and 4 of [2]. Instead of 2, we can show the stronger

$$M(Q) \ge \sum_{j=1}^{m} \left(M(Q_j) - \frac{1}{2} |\partial A_j| \right) .$$

4. Consider a point $p \in P^n$, and let p' be a point in P^n different from but closest to p. Let d(p,p') be the Euclidean distance between p and p'.

To calculate $s(p) = \mathscr{E}(d(p,p'))$, we notice that $\mathrm{prob}(d(p,p') > r) = (1 - A_p(r))^{n-1}$, where $A_p(r)$ is the area of the intersection of the unit square and the disc of center p and radius r. Thus,

$$s(p) = \int_0^{\infty} (1 - A_p(r))^{n-1} \, dr \ge \int_0^{1} (1 - \min(1, \pi r^2))^{n-1} \, dr$$

$$= \frac{1}{\sqrt{\pi}} \int_0^1 (1 - x^2)^{n-1} \, dx$$

$$= \frac{\Gamma(n)}{2\Gamma\left(n + \frac{1}{2}\right)} \ge \frac{.25}{\sqrt{n}} .$$

Since a matching of P^n at best will pair each $p \in P^n$ with its closest neighbor, we have that

$$\mathscr{E}(M(P^n)) \ge \frac{1}{2} \mathscr{E}\left(\sum_{p \in P^n} s(p) \right) = \frac{1}{2} \sum_{p \in P^n} \mathscr{E}(s(p)) \ge .25\sqrt{n} \quad . \blacksquare$$

Notice that from the last inequality it follows that $\mu \ge .25$. In Section 5 we show that $\mu \le .40106$.

3. THE

The probabilistic anal
the following framework. L
points on the plane, produce
applied to sets of points on a
process. Let $2\mu_{\mathscr{A}}$ be the
as $\xi \to \infty$. Then it follows
argument in [2].)

Our first algorithm is
heuristic (Fig. 1). In this a
divide the square into vertic
width w. In each strip we s
points bottom-up, matching
point that we meet with the o
next higher y-coordinate. T
is possibly left unmatched in
does not influence our calcul
its contribution is asymptoti
ible (recall that $\xi \to \infty$).

This heuristic is esser
same as the one analyzed in
10; following their analysis,
determine $\mu_M(w)$, the $\mu_{\mathscr{A}}$
tonic algorithm with width w
value of $\mu(w)$ is .46, attair
$w^2 = 3$. Notice that the .46
same as that derived in the I
using an independent argume

A version of this monot
whose average performance i
[2] resorts to numerical quad
will see in the next section th
relaxed combinatorially, will
effective heuristics, and an i

4. THE THR

The algorithm whose ar
elaborate implementation of t
clarity we first present the ba
the modifications necessary i
analytic tractability.

As in the monotonic heu
width w. In the monotonic he
possibly one--are matched to
are very close to each other b
Fig. 1) cannot be matched tog
algorithm processes the point
process the points in the colur
two adjacent strips (see Fig.
paired with the one having imr
points marked 1 in Fig. 2). F
vertical distance L away (as
we leave this point unmatched
L are parameters.

3. THE MONOTONIC HEURISTIC

The probabilistic analyses of this and the next section will fall into the following framework. Let $\mathscr{A}$ be an algorithm which, given a set of points on the plane, produces a feasible matching. Consider the algorithm applied to sets of points on a $\xi \times \xi$ square drawn by a unit density Poisson process. Let $2\mu_{\mathscr{A}}$ be the limit of the average length of a matched segment, as $\xi \to \infty$. Then it follows that $\mu_{\mathscr{A}} \geq \mu$. (This is a major part of the argument in [2].)

Our first algorithm is the <u>monotonic heuristic</u> (Fig. 1). In this algorithm we divide the square into vertical strips of width w. In each strip we scan the points bottom-up, matching each free point that we meet with the one with the next higher y-coordinate. The point that is possibly left unmatched in each strip does not influence our calculations, since its contribution is asymptotically negligible (recall that $\xi \to \infty$).

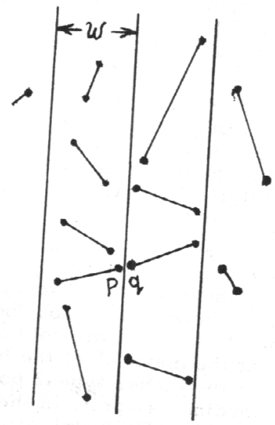

Figure 1

This heuristic is essentially the same as the one analyzed in [2], Lemma 10; following their analysis, we may determine $\mu_M(w)$, the $\mu_{\mathscr{A}}$ for the monotonic algorithm with width w. The least value of $\mu(w)$ is .46, attained for $w^2 = 3$. Notice that the .46 bound is the same as that derived in the Introduction using an independent argument.

A version of this monotonic algorithm is the best known TSP heuristic, whose average performance is analytically tractable (still, its analysis in [2] resorts to numerical quadrature and special function calculations). We will see in the next section that the matching problem, by being more relaxed combinatorially, will allow the probabilistic analysis of more effective heuristics, and an improvement on the .46 upper bound for μ.

4. THE THREE-PHASE ALGORITHM

The algorithm whose analysis will yield our upper bound for μ is an elaborate implementation of the monotonic heuristic. For the purposes of clarity we first present the basic ideas in the algorithm and then describe the modifications necessary in order to improve its effectiveness and analytic tractability.

As in the monotonic heuristic, we divide the square into strips of width w. In the monotonic heuristic all points in a strip--except for possibly one--are matched to points in the <u>same</u> strip. Thus, points that are very close to each other but in different strips (the points p,q in Fig. 1) cannot be matched together. To avoid such anomalies, our new algorithm processes the points in three phases. In the first phase, we process the points in the <u>column</u> of width 2d around the dividing line of two adjacent strips (see <u>Fig. 2</u>). Each free point in these columns is paired with the one having immediately higher y-coordinate (these are the points marked 1 in Fig. 2). However, if this next point is more than a vertical distance L away (as is the case of the points marked 2 in Fig. 2) we leave this point unmatched and proceed to the next point. w, d and L are parameters.

In the second phase, we match the points in the columns that were left unmatched in the first phase. We match them with the point in the main strip (in fact, in the corresponding left and right halves of the two neighboring main strips) that has the next greater y-coordinate. (These are the points marked 2' in Fig. 2.) Finally, in the third phase we match the remaining points in the main strips (marked 3 in Fig. 2) in the monotonic fashion.

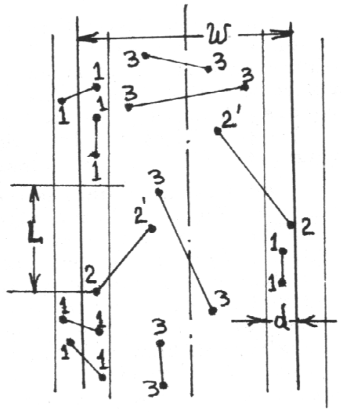

Figure 2

There are certain difficulties in the analysis of the algorithm as stated. We briefly discuss these, together with the modifications that result in an analytically tractable and effective algorithm.

First, we would like the y distance between the points of type 2 and their mates 2' to be exponentially distributed. It is not, however. The reason is that two successive type-2 points in the same column may interact in a non-trivial way. It may be that the mate 2' of the lowest is above the higher type-2 point. This means that the higher type-2 point will not be matched with the next--in terms of y coordinates--point in the main strip, but to the second next. The distribution is not Poisson.

We can avoid this difficulty by an easy modification of the algorithm. Whenever the mate 2' of a type-2 point is further--with respect to y coordinates--than the next point of the column, then we match the type-2 point at hand not to the 2' point, but to the next point in the column. It is easy to see that this does not increase the expected value of the total matching. This has the effect of "decoupling" successive type-2 points and rendering the distribution of the y-lengths of matched segments of phase 2 exponential. Note that in order to implement this modification, phases 1 and 2 must be interleaved.

In phase 3 the algorithm matches monotonically the points that were left unmatched in phases 1 and 2 (type 3 points). These points are, roughly speaking, the sum of two Poisson processes: the original one of unit density, and another process corresponding to the type-2 points. The latter process is almost Poisson--with a "dead" period of length L following each event. The effect of this process is that each even in it cancels the event of the original process that comes after it. An alternative formulation would be via a simple marked process [14]. Unfortunately, the resulting process of the surviving points is not Poisson. Quite to the contrary, the interarrival interval distribution function Laplace-transforms to a complicated irrational function, and it is not at all clear even how it could be approximated by a Poisson process.

To overcome this difficulty, we modify phase 3 of our algorithm. For each type-3 point--i.e., a point left unmatched by phases 1 and 2 --we scan the entire strip by increasing y's. If a type-3 point is met before a type-2 point (Fig. 3a), then the two points are matched (dotted segment). However, if a type-2 point is met first (Fig. 3b), the matching is changed as shown in Fig. 3c. We call such points type-4, to distinguish from the type 3 points of case (a). The mate of the type-2 point is now considered as an ordinary type-3 point (Fig. 3c). The result of this modification is that the

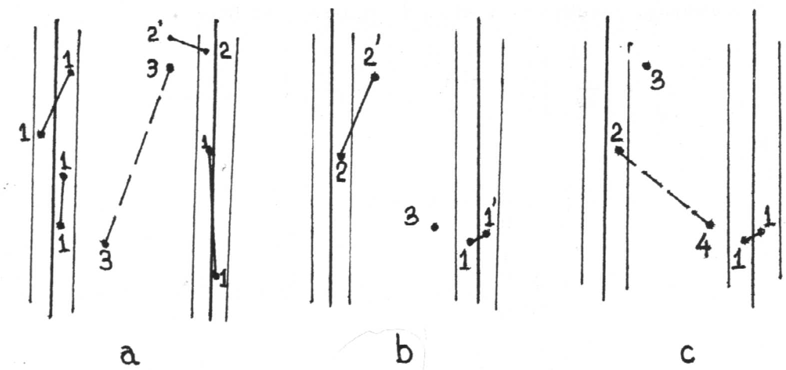

Figure 3

distribution of the y distances between two matched type-3 points is now exponential. Notice that this modification, unlike the previous one, has the effect of worsening the average overall length of the matching produced; its main purpose is to ease the analysis that follows.

5. ANALYSIS OF THE THREE-PHASE ALGORITHM

Let $2 \cdot \mu(w,d,L)$ be the average value of a matched segment for the three-phase algorithm of the previous section. In this section we will bound $\mu(w,d,L)$ from above. The derivation is complicated and lengthy, and due to space constraints, we will present here a sketch of the basic ideas involved. The methodology is, generally, analytical using approximations and special functions, and the trade-off is between accuracy and analytical tractability.

In order to estimate $\mu(w,d,L)$, we shall first calculate the densities r_1, r_2, r_3, r_4 of the points of the different types, and the average matched half-segments between points of the different types μ_1, μ_2, μ_3.

5.1 Phase 1

The process of scanning a column in phase 1 and marking its points 1 or 2 can be envisioned as a Markov process with the state diagram shown in Fig. 4. State 2 corresponds to a type-2 point; state 1 corresponds to a type-1 point that is the lower end of the matched segment, and state 1' corresponds to the rest of type-1 points. q' is the probability that, given a point in the column, the next point in this column will be the next point in the corresponding half of the strip. q' can be easily calculated as $q' = q \cdot (1 - 4\exp(-wL/2)\, d/w)$, $q = \exp(-2dL)$. The analysis of the state diagram of Fig. 4 gives:

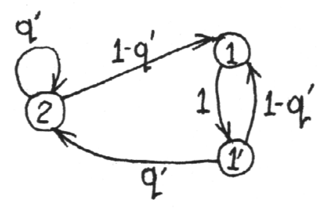

Figure 4

$$r_1 = \frac{4d(1-q')}{w(2-q')}, \quad r_2 = \frac{4dq'}{w(2-q')} .$$

For the average length of a matched segment, we have

$$\mu_1 = \frac{1}{2a}\left[\left(1 - \frac{q'}{q}\right) J_1 + \frac{q' J_2}{q(1-q)}\right] \quad ,$$

where

$$J_1 = a^4 \int_0^1 \int_0^1 \int_0^\infty \exp(-a^2 y)[(x-x')^2 + y^2]^{1/2} \, dy\, dx'\, dx$$

$$J_2 = a^4 \int_0^1 \int_0^1 \int_0^k \exp(-a^2 y)[(x-x')^2 + y^2]^{1/2} \, dy\, dx'\, dx \quad .$$

Here we let $a = 2d$ and $k = L/a$. We calculate J_2 first.

$$J_2 = 2a^2 \int_0^1 (1-u) \int_0^k \exp(-a^2 y)[u^2+y^2]^{1/2} \, du\, dy$$

$$= 2a^2\left[\frac{1}{6} - q\left(\frac{\sqrt{1+k^2}}{2} + \frac{k^2}{2}\,\ln\left(1+\sqrt{1+k^2}\right) - \frac{(1+k^2)^{3/2}}{3} - \frac{k^2 \ln k}{2} + \frac{k^3}{3}\right)\right] + I_1 + I_2$$

$$I_1 = 2a^2 \int_0^1 u \int_0^{k/u} \exp(-a^2 uz) z (1+z^2)^{1/2} \, dz\, du$$

$$= 2 \int_0^1 u \int_0^{k/u} \frac{\partial}{\partial u}\left[\exp(-a^2 uz)(1+z^2)^{-1/2}\right] dz\, du$$

$$= 2kq \int_0^1 \frac{du}{(u^2+k^2)^{1/2}} - 2\int_0^k \exp(-a^2 z)(1+z^2)^{1/2} \, dz - 2\int_0^1 \int_0^{k/u} \frac{\exp(-a^2 uz)}{(1+z^2)^{1/2}} \, dz\, du$$

$$\leq -2kq\left[\ln\left(1+\sqrt{1+k^2}\right) - \ln k\right] - I_3 + \frac{\pi}{a^2}\int_0^a (H_0(t) - Y_0(t))\, dt - 2E_1(a^2 k)$$

$I_2 = -2a^2 \int_0^k$ ex

$= -2 \int_0^k \exp($

$\leq \frac{2\pi}{a^2} (H_1(a^2$

Here H_0, H_1 are Struve fu
E_1 is the exponential integr

$J_2 \leq 2kq(\ln(1+$

$+ \frac{4 - 2qk^2 a}{a^4}$

By $\phi(x)$ we denote the func
values of ϕ calculated from

x	2	
$\phi(x)$	-.206689	-.

J_1 can now be calculated f

$J_1 = \frac{\pi\phi(a}{a^2}$

5.2 Phase 2

Let $\mu_{21}(\mu_{22})$ be the
that the two mates are in th
Fig. 6b). They can be estir

$$I_2 = -2a^2 \int_0^k \exp(-a^2 y) y \int_0^1 u(u^2+y^2)^{-1/2}\, du\, dy$$

$$= -2 \int_0^k \exp(-a^2 y)(2(1+y^2)^{1/2} - (1+y^2)^{-1/2})\, dy$$

$$- 2a^2 \int_0^k y \exp(-a^2 y)\, dy$$

$$\le \frac{2\pi}{a^2}(H_1(a^2)+Y_1(a^2)) + I_3 + \frac{4 - 2qk^2a^4}{a^4}$$

Here H_0, H_1 are Struve functions, Y_0, Y_1 are Bessel-2 functions and E_1 is the exponential integral. Then

$$J_2 \le 2kq\left(\ell n\left(1+\sqrt{1+k^2}\right) - \ell n\, k\right) + \frac{\pi\phi(a^2)}{a^2} - 2E_1(a^2 k)$$

$$+ \frac{4 - 2qk^2a^4}{a^4} + 2a^2\left[\frac{1}{6} - q\left[\frac{\sqrt{1+k^2}}{2} + \frac{k^2}{2}\left(\ell n\left(1+\sqrt{1+k^2}\right) - \ell n\, k\right)\right.\right.$$

$$\left.\left. - \frac{1}{3}((1+k^2)^{3/2} - k^3)\right]\right]$$

By $\phi(x)$ we denote the function $\int_0^x (H_0(t)-Y_0(t))\, dt - 2(H_1(x)-Y_1(x))$; values of ϕ calculated from [1] are tabulated below

x	2	2.5	3	3.5	4
$\phi(x)$	-.206689	-.006253	.144263	.264366	.364285

J_1 can now be calculated from J_2 with $k \to \infty$, $q = \exp(-a^2 k)$.

$$J_1 = \frac{\pi\phi(a^2)}{a^2} + \frac{4}{a^4} + \frac{a^2}{3} .$$

5.2 Phase 2

Let $\mu_{21}(\mu_{22})$ be the expected length of a phase-2 segment, given that the two mates are in the same strip (Fig. 5a) (resp. different strips, Fig. 6b). They can be estimated by

$$\mu_{21}^2 \le \frac{1}{4cd}\int_0^c\int_0^d (x+x')^2\,dx'\,dx + \frac{c}{2}\int_0^\infty \exp(-2cy)y^2\,dy$$

$$= \frac{(c+d)^4 - c^4 - d^4}{48dc} + \frac{1}{8c^2}$$

$$\mu_{22}^2 \le \frac{1}{4cd}\int_0^c\int_0^d (x+x'+d)\,dx'\,dx + \frac{c}{2}\int_0^\infty \exp(-2y)y^2\,dy$$

$$= \frac{(2d+c)^4 - (2d)^4 - (d+c)^4 + d^4}{48dc} + \frac{1}{8c^2},$$

where $c = (w/2) - d$. Finally, $\mu_2 = \frac{1}{2}(\mu_{21} + \mu_{22})$.

5.3 <u>Phase 3</u>

An unmatched point during phase-3 a type-3 or type-4 point depending on whether the next (in increasing y's) unmatched point comes before a type-2 point. Since both events have exponentially distributed times of first occurrence, it follows that

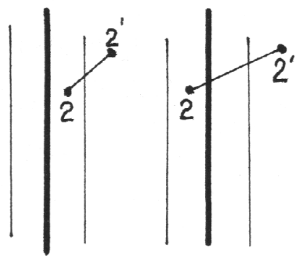

Figure 5

$$r_3 = 1 - r_1 - r_2, \quad r_4 = \frac{2c}{w}\cdot\frac{q''}{2c+q''},$$

where

$$q'' = \frac{dq'}{2-q'}.$$

$$\mu_3 = \frac{(2c)^3}{2}\int_0^1\int_0^1\int_0^\infty \exp(-4c^2y)[(x-x')^2+y^2]^{1/2}\,dx'\,dx\,dy \quad ;$$

this is identical to the integral J_1 of 5.1. Hence

$$\mu_3 = \frac{\pi\cdot\Phi((2c)^2)}{2(2c)^3} + \frac{2}{(2c)^5} + \frac{(2c)^2}{6}.$$

Finally, each type-4 point increases the estimate of μ_2 of 5.2 from $\mu_2 = \frac{1}{2}(\mu_{21}+\mu_{22})$ to $\mu_2' = \frac{1}{2}(\mu_{22}+\mu_{23})$, where

$$\mu_{23}^2 \le \frac{1}{4cd}\int_0^c\int_0^d (x+x'+d+c)^2\,dx'\,dx + \frac{c}{2}\int_0^\infty \exp(-2cy)y^2\,dy$$

$$= \frac{(2d+2c)^4 - (2d+c)^4 - (2c+d)^4 + (c+d)^4}{48cd} + \frac{1}{8c^2}.$$

5.4 <u>Computation</u>

For values of the Str
exponential integral we use

$E_1(x) \simeq \frac{1}{x\ \text{ex}}$

For $w = 2.6$, $d = .55$ a
.4010552.

<u>Monte Carlo experim</u>
ly distributed on the unit s
shown below.

n	20	
$M(P^n)/\sqrt{n}$	.38955	.3

<u>A greedy algorithm</u>
follows: process the point
point with the closest free
ered w or less. (w is a
much harder. There is a
ings, and it appears that t
greedy algorithm is by ass
bound thus achieved is .6
tional probabilities on the
regions--a bound of .49 i
were able to prove about t
for other algorithms, this
applied to random instance
<u>Improved upper bou</u>
left open here. For the T
.92 upper bound appears t
to notice that the gap betw
upper bound appears to be
problem. For the spannin
[6]; for matching--certai
tree problem--we pointed
difficult TSP, the obscure
certainly due to the algori
are the only ones availabl
<u>All known lower bou</u>
Euclidean network design
tive as the one in the proo
require totally different te

5.4 Computation

For values of the Struve and Bessel-2 functions we used [1]. For the exponential integral we used

$$E_1(x) \simeq \frac{1}{x\exp(x)} \frac{x^2+2.3347x+.2506}{x^2+3.3306x+1.6815}.$$

For $w = 2.6$, $d = .55$ and $L = 2.31$, the value of $\mu(w,d,L)$ is found .4010552.

6. REMARKS

Monte Carlo experiments on problems of size up to 200 points uniformly distributed on the unit square suggest that $\mu \simeq .35$. The results are shown below.

n	20	40	40	60	80	100	118	194
$M(P^n)/\sqrt{n}$	.38955	.319232	.3407	.3808	.3878	.3702	.37882	.36177

Table 2

A greedy algorithm can be devised to solve the matching problem as follows: process the points in order of increasing y's. Match each free point with the closest free point that has an x-distance from the one considered w or less. (w is a parameter.) The analysis of this algorithm is much harder. There is a non-trivial interaction between subsequent matchings, and it appears that the only way to estimate the performace of the greedy algorithm is by assuming the worst possible such interaction. The bound thus achieved is .675. By a more elaborate analysis--using conditional probabilities on the existence of previously matched points in certain regions--a bound of .49 is achieved. Despite the fact that the bounds we were able to prove about the greedy algorithm are inferior to the ones proved for other algorithms, this heuristic appears to perform much better when applied to random instances.

Improved upper bounds for matching is a problem that is certainly left open here. For the TSP, improving on the monotonic heuristic and the .92 upper bound appears to be extremely hard. Incidently, one cannot fail to notice that the gap between the conjectured constant and the best known upper bound appears to be increasing with the algorithmic complexity of the problem. For the spanning tree problem, this gap is just from .68 to .707 [6]; for matching--certainly a problem harder than the minimum spanning tree problem--we pointed out a .35 to .401 gap; finally, for the notoriously difficult TSP, the obscure area spans .765 to .92. Naturally, this is certainly due to the algorithmic techniques for proving upper bounds that are the only ones available to us now.

All known lower bounds for the constant of the TSP and other Euclidean network design problems are all derived by arguments as primitive as the one in the proof of Corollary 2. Improved lower bounds may require totally different techniques.

References

1. M. Abramowitz, I.A. Stegun, Handbook for Mathematical Functions, Dover, 1965.

2. J. Beardwood, J.H. Halton, J.N. Hammersley, "The Shortest Path Through Many Points", Proc. Camb. Phil. Soc. 55, 4 (1959), pp. 299-326.

3. J. Edmonds, "Paths, Trees, and Flowers", Canad. Math. J. 17 (1965), pp. 449-465.

4. H. Gabow, E.L. Lawler, "An Efficient Implementation of Edmonds' Algorithm for Maximum Matching on Graphs", TR No. 075, Dept. of Computer Science, University of Colorado, 1975.

5. M.R. Garey, R.L. Graham, D.S. Johnson, "Some NP-Complete Geometric Problems", Proc. 8th ACM Symp. on the Theory of Computing, Hershey, Pa., May 1976.

6. E.N. Gilbert, "Random Minimal Trees", J. SIAM 13, 2 (1965), pp. 376-387.

7. R.M. Karp, "The Probabilistic Analysis of Some Combinatorial Search Algorithms", in J.F. Traub (ed.), Algorithms and Complexity, Academic Press, 1976.

8. E.L. Lawler, Combinational Optimization: Networks and Matroids, Holt, Rinehart and Winston, 1976.

9. C.H. Papadimitriou, "The Euclidean TSP is NP-Complete", J. Th. Comp. Sci., to appear, 1977.

10. __________, "The Complexity of the Capacitated Tree Problem", Networks, to appear, 1978.

11. C.H. Papadimitriou and K. Steiglitz, Combinatorial Optimization Algorithms, manuscript, 1977.

12. M.I. Shamos, "Geometric Complexity", Proc. 7th ACM Symp. on the Theory of Computing, pp. 224-233, 1975.

13. D.M. Stein, Scheduling Dial-a-Ride Transportation Systems: An Asymptotic Approach, Ph.D. thesis, Harvard University, 1977.

14. J. Snyder, Random Point Processes, Wiley, 1975.

APPROXIMATE SOLUTIONS FOR THE T

CHUL E. KIM
Department of Computer Sci
College Park, Maryland 20

Abstract SIMPLE-SPLIT and LPT
are analyzed in terms of their
relative errors are shown to be
the problem size. The results
Also an ε-approximation algorit
$O(\varepsilon^{-2})$ is presented.

Consider the two-partition
A set $S = \{s_1, s_2, \ldots, s$
Partition S into P and Q
This problem is NP-hard [1,7] a
polynomial time algorithm to so
solve the problem efficiently b
Let A be an approximation
solution and the solution by A,
measured in terms of either the
$E(C_A/C^*)-1$, the expected relati
First, two simple heuristi
lyzed. It can be shown that th
Graham showed that for the LPT
In this paper the expected rela
$n^{-3/2}$ and n^{-2}, respectively, un
distribution of the sample spac
given, and seem to confirm the
is presented. It is shown that
the time and space complexity O
fying $C_\varepsilon/C^* < 1+\varepsilon$.

SIMPL

A heuristic algorithm SIMP
1. Sort S in nonincreasi
where $n = 4m+k$ and $0 \le k$
2. Partition S into P_S an
$P_S = \{s_1, s_3, \ldots, s_{2m-}$
$Q_S = \{s_2, s_4, \ldots, s_{2m},$

This research was supported in p