Probabilistic Analysis of Divide-and-Conquer Heuristics for Minimum Weighted Euclidean Matching*

Edward M. Reingold and Kenneth J. Supowit†
Department of Computer Science, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801

The expected costs of the matchings found by various divide-and-conquer heuristic algorithms are calculated, under the assumption that the vertices to be matched are uniformly distributed in the unit square. The expected times of the algorithms are calculated as well.

I. INTRODUCTION

Let n be an even integer, and P a set of n points in the unit square $[0, 1]^2$. A *matching* M of P is a set of $\frac{1}{2}n$ edges such that each point of P is an endpoint of exactly one edge of M. If M is a matching of P, then the *cost* of M, denoted $cost(M)$, is the sum of the lengths of the edges of M, where the length of an edge is defined by the Euclidean L_2 metric, unless otherwise stated.

The fastest known optimizing algorithm for this problem runs in $\Theta(n^3)$ time [4, 8], which is too slow to be of use for some applications [9]. Therefore, fast heuristics for Euclidean matching are of interest. Some have been analyzed in [1, 5, 12, 13] for their worst-case cost. The expected cost of several linear-time heuristics is discussed in [5], together with extensive empirical results, but no closed-form or asymptotic results are given. In [7], Papadimitriou analyzed a heuristic for its expected cost; in particular, he showed that the expected cost of the matching produced by his heuristic for n points independently and uniformly distributed on $[0, 1]^2$ is bounded above by $0.402\sqrt{n} + O(\sqrt{n})$.

In this article, we analyze the expected cost of the matchings produced by several of the divide-and-conquer matching heuristics of [13] (see also [12]), as well as their expected time. We find this analysis interesting for two reasons: (i) The rectangle heuristic can be implemented to run in $O(n)$ time. (In this paper, unless otherwise specified we use "time" to denote the worst-case time complexity using the real RAM with the floor function available at unit cost; this model of computation is described

*This research was supported in part by the National Science Foundation, grant number NSF MCS 81-17364.

†Present address: Hewlett-Packard Laboratories, Computer Research Lab, Palo Alto, CA 94304.

NETWORKS, Vol. 13 (1983) 49–66

CCC 0028-3045/83/010049-18$02.80

in [11].) It seems that $\Theta(n \log n)$ time is required for the heuristic of [7]. Therefore, although the rectangle heuristic has a slightly worse expected cost than does Papadimitriou's heuristic, its greater speed may make it more desirable. (ii) As we shall see, the expected cost analysis provides important information on how best to implement the heuristics; this information is not obtainable from the worst-case analysis [13].

In Sec. II we describe the basic heuristic and in Sec. III we prove a lemma and its corollary that allow us to analyze the expected behavior of the heuristic. Section IV derives a recurrence relation for the expected cost of the matching found by the heuristic and outlines the solution of the recurrence, but Sec. V derives the expected behavior in a different, more useful way. In Sec. VI we show how the analysis of Sec. V can be applied to related heuristics. Among the results of Sec. VI are some concerning the matching problem under the L_∞ metric, which is important when considering certain applications to plotter pen movement [9]. In Sec. VII we outline the analysis of the average time requirements of the heuristics.

II. THE RECTANGLE HEURISTIC

The rectangle heuristic as defined in [13] works as follows: n points are given in the unit square $[0, 1]^2$. Consider the rectangle $[0, \sqrt{2}] \times [0, 1]$, which contains the unit square. If $n \geqslant 2$ then this rectangle is bisected to form two congruent subrectangles, each with ratio $\sqrt{2}:1$ between the long and the short sides. The heuristic is applied recursively to each of the two subrectangles. In general, when applied to a rectangle R, the heuristic does as specified in Algorithm 1. We call this the *∞-version* of the rectangle heuristic.

Algorithm 1. The ∞-version of the rectangle heuristic.

if R contains at least two points
then
1. bisect R to form rectangles R_1 and R_2, each having the ratio $\sqrt{2}:1$ between the long and short sides
2. apply the heuristic to R_1
3. apply the heuristic to R_2
4. **if** R_1 and R_2 each contain an odd number of points
   then
   add the edge (p_1, p_2) to the matching, where p_1 is the point in R_1 not matched in step 2, and p_2 is that of R_2 not matched in step 3.

As an example, in Figure 1, $n = 4$. The first split is on the solid line and the left half is then split along the dashed line. The matching produced is shown as jagged line segments.

In order to make the rectangle heuristic run in time bounded by a function of n, a modification of the rectangle heuristic is specified in [13]: the level of recursion is not allowed to go beyond $\lceil \lg n \rceil$. More precisely, define a *rectangle* to be either the original $\sqrt{2}$ by 1 region, or one of two rectangular subregions with sides having ratio $\sqrt{2}:1$ into which a rectangle can be split. We define the *level* of a rectangle R, de-

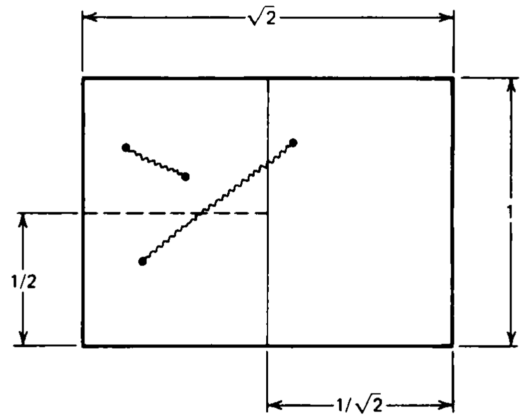

FIG. 1. Rectangle heuristic performed on four points.

noted *level*(R), as follows: *level*(R) = 0, if R is the original $\sqrt{2}$ by 1 rectangle; otherwise, *level*(R) = *level*(R') + 1, where R' is the rectangle which was bisected to form R and its mate.

The λ-version of the rectangle heuristic is given in Algorithm 2. Of special interest is the case when $\lambda = \lceil \lg n \rceil$ because the $\lceil \lg n \rceil$-version can be implemented in $\Theta(n)$ time [13]. Also, it is proved in [13] that the $\lceil \lg n \rceil$-version has the same worst-case cost as the ∞-version. More precisely, if $P \subseteq [0, \sqrt{2}] \times [0, 1]$ is a finite set of points and M is the matching produced by the ∞-version on P, then there exists a set of points $Q \subseteq [0, \sqrt{2}] \times [0, 1]$ such that $|Q| = |P|$ and $cost(M) \leqslant cost(M')$, where M' is the matching produced by the $\lceil \lg n \rceil$-version on Q.

Algorithm 2. The λ-version of the rectangle heuristic.

if *level*(R) $\leqslant \lambda$
    then
        if R contains at least two points
            then do steps 1-4 as described in Algorithm 1.
    else
        arbitrarily match the points in R until at most one is left unmatched.

Our strategy in analyzing the λ-version of the rectangle heuristic for various λ is first to analyze the expected cost on n points randomly distributed in $[0, \sqrt{2}] \times [0, 1]$, and then to use that result to analyze the expected cost on n points randomly distributed in the unit square.

III. UNIFORMITY OF THE STRANDED POINT

Let $P = \{p_1, p_2, \ldots, p_n\}$ be a set of n independent random variables, each uniformly distributed over $[0, \sqrt{2}] \times [0, 1]$. Let C_n be the expected cost of the matching produced by the ∞-version on P. In order to analyze the C_n, we need the following result.

Uniformity Lemma. Let R be a rectangle (as defined above). Let m be an *odd* integer, and let $Q = \{q_1, q_2, \ldots, q_m\}$ be a set of independent random variables each uniformly distributed over R. Let z be the random variable over R whose value represents the point stranded (i.e., not matched) by the ∞-version on Q. Then z is uniformly distributed over R.

Proof. Assume for notational convenience that $level(R) = 0$. We claim that for every $l \geqslant 0$, each level l rectangle $R_l \subseteq R$ satisfies

$$\Pr(z \in R_l) = \text{area}(R_l)/\text{area}(R). \tag{1}$$

Since for every measurable subset $S \subseteq R$ there exists a countable set of rectangles whose union is contained in S and which differs from S by a set of measure zero, it follows that (1) implies the uniform distribution of z over R. To verify (1), let $R_1, R_2, \ldots, R_{2^l}$ denote the level-l rectangles; thus $R = \cup_{1 \leqslant i \leqslant 2^l} R_i$. From the independence and uniformity of the q_j over R, it is straightforward to show that

$$\Pr(z \in R_1) = \Pr(z \in R_2) = \cdots = \Pr(z \in R_{2^l});$$

hence $\Pr(z \in R_i) = 2^{-l} = \text{area}(R_i)/\text{area}(R)$ for all i, $1 \leqslant i \leqslant 2^l$. ∎

Corollary. Let λ be a positive integer, and let R be a rectangle such that $level(R) \leqslant \lambda$. Let m and Q be defined as in the Uniformity Lemma. Let z be the random variable over R whose value represents the point stranded by the λ-version on Q. Then z is uniformly distributed over R.

Proof. Let R' be a rectangle such that $R' \subseteq R$. As in the proof of the lemma, it suffices to show that

$$\Pr(z \in R') = \text{area}(R')/\text{area}(R).$$

If $level(R') \leqslant \lambda$ then the same argument as in the proof of the lemma applies, so assume $level(R') > \lambda$. Let R'' denote the level-λ rectangle such that $R' \subseteq R'' \subseteq R$. As in the lemma, we have

$$\Pr(z \in R'') = \text{area}(R'')/\text{area}(R).$$

The point stranded by a random matching of the points in R'' is uniformly distributed over R'', hence

$$\Pr(z \in R' \mid z \in R'') = \text{area}(R')/\text{area}(R'').$$

Thus

$$\Pr(z \in R') = \Pr(z \in R' \mid z \in R'') \times \Pr(z \in R'') = \text{area}(R')/\text{area}(R). \quad ∎$$

IV. ANALYSIS OF THE ∞-VERSION

Let $n \geqslant 0$ be an integer (not necessarily even), and let $P = \{p_1, p_2, \ldots, p_n\}$ a set of independent random variables each uniformly distributed over the rectangle $R = [0, \sqrt{2}] \times [0, 1]$. Let C_n be the expected cost of the matching produced by the ∞-version of the rectangle heuristic on R. (We need the C_n for odd n in order to analyze the C_n for even n, which are our chief concern.) We have

$$C_0 = 0,$$

$$C_n = \sum_{k=0}^{n} \binom{n}{k} \left(\frac{1}{2}\right)^k \left(\frac{1}{2}\right)^{n-k} \left(\frac{1}{\sqrt{2}} C_k + \frac{1}{\sqrt{2}} C_{n-k}\right) + D \, \Pr(\text{odd-odd split}),$$

where D is the expected distance between two randomly chosen points one in each of the two halves of the rectangle R, and Pr(odd-odd split) is the probability that the two halves of R each contain an odd number of points of P. This recurrence is derived from the facts that $\binom{n}{k}(\frac{1}{2})^k(\frac{1}{2})^{n-k}$ is the probability of a split of k points in the left half of R and $n - k$ in the right half of R, $C_k/\sqrt{2}$ and $C_{n-k}/\sqrt{2}$ are the expected costs of the matchings in the left and right halves, respectively, and D Pr(odd-odd split) is the expected cost of the edge-matching the stranded points, if any. The factor D is derived from the Uniformity Lemma; that is, the two stranded points are each uniformly distributed in their respective halves of R. If n is odd, Pr(odd-odd split) = 0. If n is even, that probability is

$$\sum_{k \text{ even}} \binom{n}{k} \left(\frac{1}{2}\right)^k \left(\frac{1}{2}\right)^{n-k} = \frac{1}{2}.$$

Thus

$$C_0 = 0,$$

$$C_n = \frac{\sqrt{2}}{2^n} \sum_{k=0}^{n} \binom{n}{k} C_k + \begin{cases} 0 & (n \text{ odd}) \\ \frac{1}{2} D & (n \text{ even}) \end{cases} \qquad n \geqslant 1.$$

For the $\sqrt{2} \times 1$ rectangle, D can be calculated by (tedious) elementary calculus to be

$$D = 2 \int_{-\sqrt{2}/2}^{0} \int_{0}^{\sqrt{2}/2} \int_{0}^{1} \int_{0}^{1} [(x_1 - x_2)^2 + (y_1 - y_2)^2]^{1/2} \, dy_1 \, dy_2 \, dx_1 \, dx_2$$

$$= \frac{4 - \sqrt{2}}{2} \left[\frac{1 + 4\sqrt{2} + \sqrt{3}}{30} + \frac{1}{3} \ln\left(\frac{\sqrt{2} + \sqrt{6}}{2}\right) + \frac{\sqrt{2}}{12} \ln(\sqrt{2} + \sqrt{3})\right]$$

$$\approx 0.819960615.$$

Rewriting the recurrence relation for C_n more generally as

$$x_0 = 0,$$
$$x_n = a_n + \frac{\alpha}{2^n} \sum_{k=0}^{n} \binom{n}{k} x_k, \qquad n \geq 1, \tag{2}$$

we have $C_n = Dx_n$ when $\alpha = \sqrt{2}$ and

$$a_n = \begin{cases} 0 & (n \text{ odd or } n = 0) \\ \frac{1}{2} & (n \text{ even}, \quad n \geq 2). \end{cases}$$

Although a closed-form solution is not known for the recurrence (2), it is possible to derive asymptotic estimates. Knuth [6] presents a general technique for handling such recurrences in the course of analyzing the average behavior of the radix exchange sort (when the inputs are uniformly distributed random real numbers). Saxe [10] used an ad hoc technique for the case $\alpha = 1, a_n = 1$ in order to analyze a certain coin tossing problem (n coins are given, all initially heads up; at each stage, all heads-up coins are flipped—what is the expected number of *stages* until all of the coins are tails up?).

The general technique of Knuth is to use the *binomial transform* $\langle \hat{a}_n \rangle$ of the sequence $\langle a_n \rangle$,

$$\hat{a}_n = \sum_{k=0}^{n} \binom{n}{k} (-1)^k a_k;$$

it can be shown that

$$x_n = a_n + \sum_{k=2}^{n} \binom{n}{k} (-1)^k \frac{\alpha \hat{a}_k}{2^k - \alpha}.$$

Following a suggestion of de Bruijn, Knuth observes that

$$\frac{\alpha}{2^k - \alpha} = \sum_{j=1}^{\infty} \left(\frac{\alpha}{2^k} \right)^j$$

and interchanging the order of summation gives

$$x_n = a_n + \sum_{j=1}^{\infty} \alpha^j \sum_{k=2}^{n} \binom{n}{k} \hat{a}_k \left(\frac{-1}{2^j} \right)^k.$$

In the various cases of interest, $\langle \hat{a}_k \rangle$ turns out to be simple enough to allow the inner summation to be evaluated in closed form, and the infinite sum that results can be approximated by gamma-function integrals.

Applying this technique to C_n we get

$$C_n = \tfrac{1}{4}D[2(2-\sqrt{2})\sqrt{\pi}/\ln 2 - f(n)]\sqrt{n} + O(1)$$
$$\approx 0.614118\sqrt{n},$$

where $f(n)$ is negligible and is defined below. (We do not go into the details of this derivation of C_n because we get the same result by a more useful level-by-level approach in the next section.)

Recall that we are interested in the rectangle heuristic applied to points in the unit square. The argument of [13] can be applied here to show that the expected cost of the matching on n points independently and uniformly distributed on $[0, 1]^2$ is

$$\frac{1}{\sqrt{\sqrt{2}}} C_n \approx 0.516410\sqrt{n},$$

a bit larger than the upper bound of $0.402\sqrt{n}$ found by Papadimitriou for the minimum matching of n random points in the unit square.*

V. ANALYSIS OF THE λ-VERSION

Fix some $\lambda \in \{1, 2, 3, \ldots\} \cup \{\infty\}$. To analyze the λ-version of the algorithm we proceed as follows. Let a_{il} be the expected number of level-l rectangles containing exactly i points. The probability that a level-l rectangle contains exactly i points is

$$\binom{n}{i}\left(1 - \frac{1}{L}\right)^{n-i}\left(\frac{1}{L}\right)^i,$$

where $L = 2^l$ is the number of level-l rectangles. The expected number of level-l rectangles with exactly i points is this probability times the number of level-l rectangles; thus

$$\begin{aligned} a_{il} &= \binom{n}{i}(1 - 2^{-l})^{n-i}(2^{-l})^i 2^l \\ &= \binom{n}{i} 2^{-l(n-1)}(2^l - 1)^{n-i} \end{aligned} \tag{3}$$

(with 0^0 taken as 1).

*More precisely, the argument of [13] says that

$$(1-\epsilon)\frac{1}{\sqrt{\sqrt{2}}}C_n + \delta_1(\epsilon) \leqslant \text{expected cost} \leqslant (1+\epsilon)\frac{1}{\sqrt{\sqrt{2}}}C_n + \delta_2(\epsilon),$$

for all $\epsilon > 0$, where $\delta_1(\epsilon)$ and $\delta_2(\epsilon)$ do not depend on n.

Let e_l be the expected number of *nonempty* level-l rectangles with an *even* number of points:

$$e_l = \sum_{\substack{i>0 \\ i \text{ even}}} a_{il}$$

$$= \sum_{k=1}^{n/2} \binom{n}{2k} 2^{-l(n-1)} (2^l - 1)^{n-2k}.$$

This can be evaluated by elementary techniques to yield

$$e_l = 2^{l-1} [1 + (1 - 2^{-l+1})^n - 2(1 - 2^{-l})^n], \qquad l \geqslant 0.$$

By the corollary to the Uniformity Lemma, each nonempty level-l rectangle containing an even number of points contributes $D/\sqrt{2^l}$ to the cost of the matching, *if* it splits odd-odd (this happens with probability $\frac{1}{2}$, as calculated above), and 0 if it splits even-even. Thus the contribution of the lth level to the expected cost of the matching is $\frac{1}{2}De_l/\sqrt{2^l}$. Define

$$S(\lambda) = 2 \sum_{l=1}^{\lambda} \frac{e_l}{\sqrt{2^l}}$$

$$= \sum_{l=1}^{\lambda} \sqrt{2^l}[1 + (1 - 2^{-l+1})^n - 2(1 - 2^{-l})^n];$$

the contribution of levels $1, 2, \ldots, \lambda$ is then $\frac{1}{4}DS(\lambda)$. Our cost C_n from above is

$$C_n = \tfrac{1}{2}D + \tfrac{1}{4}DS(\infty),$$

and we are interested in the expected cost of the first λ levels,

$$C_n(\lambda) = \tfrac{1}{2}D + \tfrac{1}{4}DS(\lambda).$$

We now use the de Bruijn-Knuth gamma-function integral method to develop an asymptotic estimate of $S(\lambda)$ in order to estimate $C_n(\lambda)$. Define

$$g(x) = (1 - e^{-x})^2/\sqrt{x}.$$

Knuth proves that

$$\sum_{j=1}^{\infty} 2^j [(1 - 2^{-j})^n - e^{-n/2^j}] = O(1)$$

(see Sec. 5.2.2, pp. 131, 132 and exercises 46 and 47 of [6]). From this it follows directly that

$$S(\lambda) = \sqrt{n} \sum_{j=1}^{\lambda} g\left(\frac{n}{2^j}\right) + O(1).$$

Furthermore, since

$$e^{-x} - 1 + x = \frac{1}{2\pi i} \int_{-3/2-i\infty}^{-3/2+i\infty} \Gamma(z) x^{-z} \, dz$$

(see, e.g., p. 132 of [6]),

$$g(x) = \frac{\sqrt{2}}{2\pi i} \int_{-3/2-i\infty}^{-3/2+i\infty} \Gamma(z) \, [(2x)^{-z-1/2} - \sqrt{2} \, x^{-z-1/2}] \, dz$$

and we can write

$$S(\lambda) = \frac{\sqrt{n}}{2\pi i} \sum_{j=1}^{\lambda} \int_{-3/2-i\infty}^{-3/2+i\infty} \Gamma(z) \left[\sqrt{2} \left(\frac{n}{2^{j-1}}\right)^{-z-1/2} - 2\left(\frac{n}{2^j}\right)^{-z-1/2}\right] dz + O(1).$$

Following Knuth, we observe that the convergence is uniform and so we can bring the summation under the integral:

$$= \frac{\sqrt{n}}{2\pi i} \int_{-3/2-i\infty}^{-3/2+i\infty} \Gamma(z) \left[\sqrt{2} \sum_{j=1}^{\lambda} \left(\frac{n}{2^{j-1}}\right)^{-z-1/2} - 2 \sum_{j=1}^{\lambda} \left(\frac{n}{2^j}\right)^{-z-1/2}\right] dz + O(1).$$

For $\mathrm{Re}(w) > 0$ we have

$$\sum_{j=1}^{\lambda} \left(\frac{n}{2^{j-1}}\right)^{w} = (2n)^{w} \, \frac{1 - (1/2^{\lambda})^{w}}{2^{w} - 1},$$

and hence also

$$\sum_{j=1}^{\lambda} \left(\frac{n}{2^j}\right)^{w} = n^{w} \, \frac{1 - (1/2^{\lambda})^{w}}{2^{w} - 1}.$$

Thus

$$S(\lambda) = \frac{\sqrt{n}}{2\pi i} \int_{-3/2-i\infty}^{-3/2+i\infty} \Gamma(z) \, n^{-z-1/2} \left[1 - \left(\frac{1}{2^{\lambda}}\right)^{-z-1/2}\right] \frac{2^{-z} - 2}{2^{-z-1/2} - 1} \, dz + O(1).$$

the two cases of interest are $\lambda = \lceil \lg n \rceil$ and $\lambda = \infty$, we consider them separately, first, $\lambda = \infty$.

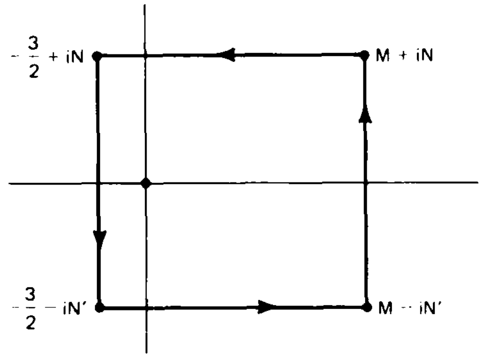

FIG. 2. Contour of integration C.

We need to evaluate

$$\frac{\sqrt{n}}{2\pi i}\int_{-3/2-i\infty}^{-3/2+i\infty} \Gamma(z)\, n^{-z-1/2}\, \frac{2^{-z}-2}{2^{-z-1/2}-1}\, dz.$$

Let $F(z)$ be the integrand. Integrating on the contour C as shown in Figure 2 we have

$$\frac{1}{2\pi i}\int_C F(z)\,dz = \frac{1}{2\pi i}\int_{-3/2+iN}^{-3/2-iN'} F(z)\,dz + \frac{1}{2\pi i}\int_{M+iN}^{-3/2+iN} F(z)\,dz$$

$$+ \frac{1}{2\pi i}\int_{M-iN'}^{M+iN} F(z)\,dz + \frac{1}{2\pi i}\int_{-3/2-iN'}^{M-iN'} F(z)\,dz$$

$$= \Sigma \text{ residues of } F(z) \text{ in } C$$

by the Cauchy residue theorem. As in Knuth, the last three integrals (corresponding to the top, right, and bottom lines of C) are negligible as $N, N', M \to \infty$, so that

$$S(\infty) = -n^{1/2}\ [\Sigma \text{ residues of } F(z) \text{ in } C] + O(1).$$

The poles of $F(z)$ in C as $N, N', M \to \infty$ are $z = 0$ and $z = -1$ [for $\Gamma(z)$] and $z = -\frac{1}{2} + (2\pi i/\ln 2)k$, $k = 0, 1, 2, \ldots$. The residue of $\Gamma(z)$ at the pole $z = -j$ is $(-1)^j/j!$. The residues of $F(z)$ in C are thus

$z = 0$:

$$\text{residue} = [(-1)^0/0!]\,n^{-0-1/2}(2^{-0}-2)/(2^{-0-1/2}-1) = (2-\sqrt{2})/\sqrt{n},$$

$z = -1$:

$$\text{residue} = [(-1)^1/1!]\,n^{1-1/2}(2^1-2)/(2^{1-1/2}-1) = 0,$$

$z = -\frac{1}{2}$:

$$\text{residue} = \lim_{z \to -1/2} (z + \tfrac{1}{2}) F(z)$$

$$= -\Gamma(-\tfrac{1}{2})(\sqrt{2} - 2)/\ln 2 \qquad \text{(by L'Hospital's rule)}$$

$$= -2(2 - \sqrt{2})\sqrt{\pi}/\ln 2,$$

$z = -\frac{1}{2} + (2\pi i/\ln 2)\, k, k \geqslant 1$:

$$\text{residue} = \lim_{z \to -1/2 + (2\pi i/\ln 2) k} \left(z + \frac{1}{2} - \frac{2\pi i}{\ln 2} k\right) F(z)$$

$$= \frac{2 - \sqrt{2}}{\ln 2} \Gamma\left(-\frac{1}{2} + \frac{2\pi i}{\ln 2} k\right) e^{2\pi i \{k \lg n\}} \qquad \text{(by L'Hospital's rule),}$$

where $\{x\}$ is the fractional part of x.

Thus

$$S(\infty) = [2(2 - \sqrt{2})\sqrt{\pi}/\ln 2 - f(n)]\sqrt{n} + O(1),$$

where

$$f(n) = \frac{2 - \sqrt{2}}{\ln 2} \sum_{k=1}^{\infty} \operatorname{Re}[\Gamma(-\tfrac{1}{2} + 2\pi i/\ln 2)\, e^{2\pi i \{k \lg n\}}].$$

To bound the effect of $f(n)$, note that $|e^{2\pi i \{k \lg n\}}| \leqslant 1$ for all n and k, so

$$f(n) \leqslant \frac{2 - \sqrt{2}}{\ln 2} \sum_{k=1}^{\infty} \left| \Gamma\left(-\frac{1}{2} + \frac{2\pi k}{\ln 2} i\right) \right|.$$

Since $\Gamma(z + 1) = z\Gamma(z)$ and $|\Gamma(\frac{1}{2} + iy)| = (\pi/\cosh \pi y)^{1/2}$,

$$|\Gamma(-\tfrac{1}{2} + iy)| = [8\pi/(1 + 4y^2)(e^{\pi y} + e^{-\pi y})]^{1/2}.$$

Thus

$$|\Gamma(-\tfrac{1}{2} + 2\pi i/\ln 2)| < 1.81 \times 10^{-7},$$

and

$$|\Gamma(-\tfrac{1}{2} + 2\pi(k + 1)i/\ln 2)| < |\Gamma(-\tfrac{1}{2} + 2\pi ki/\ln 2)| \times 6.6 \times 10^{-7}.$$

It follows easily that

$$\sum_{k=1}^{\infty} \left| \Gamma\left(-\frac{1}{2} + \frac{2\pi k}{\ln 2} i\right) \right| < 1.82 \times 10^{-7},$$

so that

$$f(n) < 1.54 \times 10^{-7}.$$

The behavior of the ∞-version is thus determined:

$$\begin{aligned} C_n &= \tfrac{1}{4}D\,[2(2-\sqrt{2})\sqrt{\pi}/\ln 2 - f(n)]\sqrt{n} + O(1) \\ &\approx 0.614118\sqrt{n}. \end{aligned}$$

For the $\lceil \lg n \rceil$-version we need to analyze $S(\lceil \lg n \rceil)$. Unfortunately, the gamma-function integral technique *cannot* be used in this case since the values of the integrals for the top, right, and bottom boundaries are *not* negligible unless $n/2^\lambda \to \infty$ as $n \to \infty$.

Let $\theta = \lceil \lg n \rceil - \lg n, 0 \leqslant \theta < 1$. Then

$$\begin{aligned} S(\lceil \lg n \rceil) &= S(\theta + \lg n) \\ &= S(\infty) - \sqrt{n} \sum_{j=1+\theta+\lg n}^{\infty} g\left(\frac{n}{2^j}\right) + O(1) \\ &= S(\infty) - \sqrt{n} \sum_{j=1}^{\infty} g(2^{-j-\theta}) + O(1). \end{aligned}$$

Define

$$h(\theta) = \sum_{j=1}^{\infty} g(2^{-j-\theta}).$$

We have

$$S(\lceil \lg n \rceil) = [2(2-\sqrt{2})\sqrt{\pi}/\ln 2 - h(\theta) - f(n)]\sqrt{n} + O(1),$$

or

$$C_n(\lceil \lg n \rceil) = \tfrac{1}{4}D[2(2-\sqrt{2})\sqrt{\pi}/\ln 2 - h(\theta) - f(n)]\sqrt{n} + O(1).$$

Finally, we must take into consideration the cost of the edges chosen randomly at level $\lambda = 1 + \lceil \lg n \rceil$. In a level-$\lambda$ rectangle with i points, $\frac{1}{2}i$ random edges are picked if i is even, and $(\frac{1}{2}i - 1)$ if i is odd. Thus the randomly chosen edges contribute

$$\sum_{i\text{ even}} \sqrt{2}^{-\lambda} \frac{D}{2} i a_{i\lambda} + \sum_{i\text{ odd}} \sqrt{2}^{-\lambda} \frac{D}{2}(i-1) a_{i\lambda} = \sqrt{2}^{-\lambda} \frac{D}{2}\left(\sum_{i=0}^{n} i a_{i\lambda} - \sum_{i\text{ odd}} a_{i\lambda}\right),$$

which is evaluated without difficulty to be

$$\tfrac{1}{4}D\,[2n\sqrt{2}^{-\lambda} - \sqrt{2}^{\lambda} + \sqrt{2}^{\lambda}(1 - 2^{-\lambda} + 1)^{n}].$$

For $\lambda = 1 + \lceil \lg n \rceil = 1 + \theta + \lg n$ this gives

$$\tfrac{1}{4}D(\sqrt{2}^{1-\theta} - \sqrt{2}^{1+\theta} + \sqrt{2}^{1+\theta}e^{-2^{-\theta}})\sqrt{n} + O(1).$$

Let

$$r(\theta) = \sqrt{2}^{1-\theta} - \sqrt{2}^{1+\theta} + \sqrt{2}^{1+\theta}\,e^{-2^{-\theta}};$$

the expected total cost of the matching produced by the $\lceil \lg n \rceil$-version of the algorithm on a set of n random points in the $\sqrt{2} \times 1$ rectangle is thus

$$\tfrac{1}{4}D\,[2(2 - \sqrt{2})\sqrt{\pi}/\ln 2 - f(n) + r(\theta) - h(\theta)]\sqrt{n} + O(1)$$

and is

$$M(n) = \frac{1}{\sqrt{\sqrt{2}}}\frac{D}{4}\left(\frac{2(2 - \sqrt{2})\sqrt{\pi}}{\ln 2} - f(n) + r(\theta) - h(\theta)\right)\sqrt{n} + O(1)$$

for all $\epsilon > 0$, if the points all lie in the unit square, by the technique of [13].* The function $r(\theta) - h(\theta)$ is strictly decreasing on $[0, 1]$, with $r(0) - h(0) \approx 0.141375671$ and $r(1) - h(1) \approx 0.053122782$. Thus, for example, when n is a power of 2 we have $\theta = 0$ and

$$M(n) \approx 0.540779\sqrt{n}.$$

When $n = 2^k + 2$, $\theta \to 1$ as $n \to \infty$ and we have

$$M(n) \approx 0.525567\sqrt{n}.$$

The ∞-version would give $C_n/\sqrt{\sqrt{2}} \approx 0.516410\sqrt{n}$, so the truncation at level $\lceil \lg n \rceil$ increases the expected cost of the matching by between 2.2 and 4.7%, depending on the value of θ. That the truncation does not increase the cost of the matching in the worst case was demonstrated in [13].

VI. RELATED ANALYSES

All of the above analysis can be applied with little effort to the L_∞ distance norm and/or to the triangle heuristic of [13]—all that changes is the constant D. In the case of the L_∞ norm, we observe that distances in a level-$(l + 1)$ rectangle are just $1/\sqrt{2}$

*See the footnote to p. 55.

times those in a level-l rectangle, exactly as with the L_2 norm, D, however, changes to

$$D = 2\int_{-\sqrt{2}/2}^{0}\int_{0}^{\sqrt{2}/2}\int_{0}^{1}\int_{0}^{1} \max(|x_1 - x_2|, |y_1 - y_2|)\,dy_1\,dy_2\,dx_1\,dx_2$$
$$= \frac{25\sqrt{2} - 1}{60}$$
$$\approx 0.572588984.$$

To apply the analysis to the triangle heuristic we begin with a brief summary of that heuristic. The triangle heuristic does a divide-and-conquer approach identical with that of the rectangle heuristic *except* that it works directly on the unit square: The first bisection divides the square diagonally into two 45°-45°-90° triangles and later bisections divide the triangles into similar 45°-45°-90° triangles. An example parallel to that of Figure 1 is shown in Figure 3.

Let T_n be the expected cost of the matching produced by the ∞-version of the triangle heuristic when n uniformly distributed points are chosen in a 45°-45°-90° triangle with hypotenuse $\sqrt{2}$. Except for the particular value of D, T_n satisfies precisely the same recurrence relation as C_n does and we have

$$T_n = \tfrac{1}{4}D_T[2(2 - \sqrt{2})\sqrt{\pi}/\ln 2 - f(n)]\sqrt{n} + O(1),$$

where

$$D_T = 8\int_{-\sqrt{2}/2}^{0}\int_{0}^{\sqrt{2}/2}\int_{0}^{1+\sqrt{2}x_2}\int_{0}^{1-\sqrt{2}x_1} d[(x_1,y_1),(x_2,y_2)]\,dy_1\,dy_2\,dx_1\,dx_2$$

and $d[(x_1, y_1), (x_2, y_2)]$ is either the L_2 distance or the L_∞ distance between the points (x_1,y_1) and (x_2,y_2).

The determination of S_n, the expected cost of the matching on the unit square

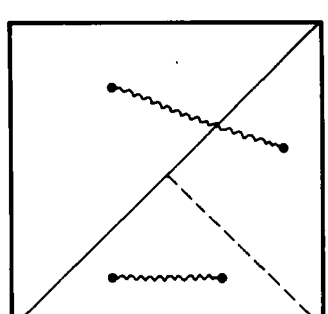

FIG. 3. Triangle heuristic performed on four points.

found by the ∞-version of the triangle heuristic, proceeds by observing that

$$S_n = \sum_{k=0}^{n} \binom{n}{k} \left(\frac{1}{2}\right)^k \left(\frac{1}{2}\right)^{n-k} (T_k + T_{n-k}) + D_S \operatorname{Pr}(\text{odd-odd split})$$
$$= \frac{2}{2^n} \sum_{k=0}^{n} \binom{n}{k} T_k + O(1),$$

where D_S is the expected distance between two randomly chosen points, one in each of the two triangles formed by bisecting the square diagonally. We know from the results of Sec. V that $T_n = K\sqrt{n} + O(1)$, so that

$$S_n = \frac{2K}{2^n} \sum_{k=0}^{n} \binom{n}{k} \sqrt{k} + O(1).$$

Let $\beta_k = \binom{n}{\nu+k}/2^n$, $\nu = \frac{1}{2}n$. We have the well-known approximations

$$\beta_k \approx \beta_0 e^{-k^2/\nu}$$
$$\beta_0 \approx 1/\sqrt{\pi\nu}$$

(see, e.g., [3]). Thus

$$\frac{1}{2^n} \sum_{k=0}^{n} \binom{n}{k} \sqrt{k} \approx \frac{1}{\sqrt{\pi\nu}} \sum_{k=-\nu}^{\nu} e^{-k^2/\nu} \sqrt{\nu + k}.$$

The straightforward application of Euler's summation formula (see, e.g., [2]) gives

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

so that

$$S_n = \sqrt{2}\, T_n + O(1)$$
$$= (D_T/2\sqrt{2})[2(2 - \sqrt{2})\sqrt{\pi}/\ln 2 - f(n)]\sqrt{n} + O(1).$$

For the L_2 norm, $D_T \approx 0.54$ so that $S_n \approx 0.57\sqrt{n}$ and the triangle heuristic is inferior to the rectangle heuristic on the average as well as in the worst case (see [13]).

The analysis of the $\lceil \lg n \rceil$-version of the triangle heuristic is the same as that for the rectangle heuristic, except (again) for the constant D and the factoring down from the $\sqrt{2} \times 1$ rectangle to the unit square: The expected cost of the matching is $M(n)$ with $D_T/(2\sqrt{2})$ in place of $D/(4\sqrt{\sqrt{2}})$.

VII. AVERAGE TIME OF THE ∞-VERSION

We have already observed above that the $\lceil \lg n \rceil$-version (of either the rectangle heuristic or the triangle heuristic) can be implemented in $\Theta(n)$ worst-case time using the real RAM model with the floor function available at unit cost. For the ∞-version, the worst-case time is not bounded by any function of n because the points can be clustered arbitrarily close together. However, the ∞-version can be implemented in $\Theta(n)$ expected time, and even the (naive) implementation of the ∞-version in Algorithm 1 requires only $\Theta(n \log n)$ expected time.

The expected time of the naive ∞-version (Algorithm 1) on a set of n points uniformly distributed over the region is proportional to z_n satisfying

$$z_0 = z_1 = 1,$$

$$z_n = O(n) + \sum_{k=0}^{n} \binom{n}{k} \left(\frac{1}{2}\right)^k \left(\frac{1}{2}\right)^{n-k} (z_k + z_{n-k}),$$

where the $O(n)$ term represents the time to partition the points into the two subregions. This can be rewritten in the form of Eq. (2) and estimated by the techniques described earlier. The result is $z_n = \Theta(n \log n)$.

A more subtle implementation has $\Theta(n)$ expected time: Partition the region into the $2^{\lceil \lg n \rceil}$ subregion of level $\lceil \lg n \rceil$ and apply *this* recursively to any subregion with three or more points. The expected time of this implementation satisfies

$$Z_n = \sum_{i=0}^{n} Z_i a_{i, \lceil \lg n \rceil}$$

where a_{il} is given by Eq. (3). Elementary manipulations and mathematical induction show that $Z_n = \Theta(n)$.

VIII. SUMMARY

The results of Secs. V and VI for the L_2 and the L_∞ metrics are summarized in Table I.

Table II contains the dominant term of the expected cost of the matching produced by the λ-version of the rectangle heuristic when $\lambda = \lceil \lg n \rceil - 7, \lceil \lg n \rceil - 6, \cdots, \lceil \lg n \rceil + 7$ on n points in the unit square, for the L_2 metric. We assume n is a power of 2; recall from Section V that the costs are maximized when n is a power of 2, hence the

TABLE I. Expected costs of matching for various versions of rectangle heuristic for L_2 and L_∞ metrics

	L_2	L_∞
∞-version	$0.516410\sqrt{n}$	$0.428846\sqrt{n}$
$\lceil \lg n \rceil$-version, $n = 2^k$	$0.540779\sqrt{n}$	$0.449084\sqrt{n}$
$\lceil \lg n \rceil$-version, $n = 2^k + 2$	$0.525567\sqrt{n}$	$0.436451\sqrt{n}$

TABLE II. Expected cost of matching produced by various versions of rectangle heuristic for n a power of 2 points uniformly distributed in the unit square.

λ	Expected cost of matching
$\lceil \lg n \rceil - 7$	$2.7885\sqrt{n}$
$\lceil \lg n \rceil - 6$	$1.9933\sqrt{n}$
$\lceil \lg n \rceil - 5$	$1.4399\sqrt{n}$
$\lceil \lg n \rceil - 4$	$1.0613\sqrt{n}$
$\lceil \lg n \rceil - 3$	$0.8114\sqrt{n}$
$\lceil \lg n \rceil - 2$	$0.6590\sqrt{n}$
$\lceil \lg n \rceil - 1$	$0.5779\sqrt{n}$
$\lceil \lg n \rceil$	$0.5408\sqrt{n}$
$\lceil \lg n \rceil + 1$	$0.5256\sqrt{n}$
$\lceil \lg n \rceil + 2$	$0.5198\sqrt{n}$
$\lceil \lg n \rceil + 3$	$0.5176\sqrt{n}$
$\lceil \lg n \rceil + 4$	$0.5168\sqrt{n}$
$\lceil \lg n \rceil + 5$	$0.5166\sqrt{n}$
$\lceil \lg n \rceil + 6$	$0.5165\sqrt{n}$
$\lceil \lg n \rceil + 7$	$0.5164\sqrt{n}$

expressions in Table II are upper bounds for all values of n. These expressions were calculated from the formulas of Section V.

Recall that for finite λ, the λ-version can be implemented in $O(2^\lambda + n)$ time [13]. Hence, for fixed k, the $(\lceil \lg n \rceil + k)$-version can be implemented in $O(n)$ time. Therefore, for some purposes it may be best to use the $(\lceil \lg n \rceil + 1)$- or the $(\lceil \lg n \rceil + 2)$-version, since their expected costs are only 1.77 and 0.6% worse, respectively, than that of the ∞-version, whereas for the $\lceil \lg n \rceil$-version the expected cost is 4.77% worse than that of the ∞-version. It is interesting to note that from the worst-case analysis [13] alone, one could not justify the use of the λ-version for $\lambda > \lceil \lg n \rceil$, since the worst-case cost of the $\lceil \lg n \rceil$-version equals that of the ∞-version.

The authors are grateful to L. Monier and D. S. Watanabe for their helpful remarks and to J. Purtilo for wrestling with MACSYMA in the evaluation of the quadruple integrals.

References

[1] D. Avis, Two greedy heuristics for the weighted perfect matching problem. In *Proc. Ninth Southeastern Conf. on Combinatorics, Graph Theory, and Computing*, (1978) 65–76.

[2] N. G. De Bruijn, *Asymptotic Methods in Analysis*. North-Holland, Amsterdam (1961). Republished by Dover, New York (1981).

[3] W. Feller, *An Introduction to Probability Theory and Its Applications*. Wiley, New York (1968), Vol. I.

[4] H. Gabow, An efficient implementation of Edmond's algorithm for maximum matching on graphs. *J. Assoc. Comput. Mach.* 23 (1976) 221–234.

[5] M. Iri, M. Murota, and S. Matsui, Linear-time heuristics for the minimum-weight perfect matching on a plane with an application to the plotter algorithm. Unpublished (1980).
[6] D. E. Knuth, *The Art of Computer Programming, Volume 3: Searching and Sorting*. Addison-Wesley, Reading, MA (1973).
[7] C. H. Papadimitriou, The probabilistic analysis of matching heuristics. In *Proc. Fifteenth Annual Allerton Conf. on Communication, Control and Computing* (1977) 368–378.
[8] C. H. Papadimitriou and K. Steiglitz, *Combinatorial Optimization: Algorithms and Complexity*, Prentice-Hall, Englewood Cliffs, NJ (1982).
[9] E. M. Reingold and R. E. Tarjan, On a greedy heuristic for complete matching. *SIAM J. Comput.* 10 (1981) 676–681.
[10] J. Saxe, personal communication (1981).
[11] M. I. Shamos, *Computational Geometry*, Ph.D. Thesis, Yale University (1978).
[12] K. J. Supowit, D. A. Plaisted, and E. M. Reingold, Heuristics for weighted perfect matching. In *Proc. Twelfth Annual ACM Symp. on Theory of Computing* (1980) 398–419.
[13] K. J. Supowit and E. M. Reingold, Divide-and-Conquer heuristics for minimum weighted Euclidean matching. *SIAM J. Comput.* To appear.

Received May 1, 1981
Accepted May 11, 1982