Comp. & Maths. with Appls. Vol. 7, pp. 251–257, 1981
Printed in Great Britain.

0097-4943/81/030251-07$02.00/0
Pergamon Press Ltd

WORST CASE BOUNDS FOR THE EUCLIDEAN MATCHING PROBLEM†

DAVID AVIS
School of Computer Science, McGill University, 805 Sherbrooke Street West, Montreal, Quebec, Canada, H3A 2K6

(*Received March* 1980)

Communicated by E. Y. Rodin

Abstract—It is shown how the classical mathematical theory of sphere packing can be used to obtain bounds for a greedy heuristic for the bounded euclidean matching problem. In the case of 2 dimensions, bounds are obtained directly. For higher dimensions, an appeal is made to known bounds for the sphere packing problem that have appeared in the mathematical literature.

1. INTRODUCTION

The Euclidean matching problem may be stated as follows: given n points in Euclidean space, find a minimum weight perfect matching, where the weights are given by the Euclidean distances between the points. For convenience, we assume n is even throughout the paper. We will consider the problem where the points are constrained to lie in some bounded region, which we will assume to be the unit hypercube. This special case occurs in two dimensions as a graphics problem[1], and in both 2 and 3 dimensions as a subproblem in Christofides' heuristic for the travelling salesman problem[2]. Although there exist $0(n^3)$ implementations of Edmonds' optimal algorithm (see, e.g. [3]), we will analyze a faster GREEDY heuristic. The most obvious such heuristic is to repeatedly match the closest two unmatched vertices, resolving ties in an arbitrary fashion. This heuristic can be implemented in $0(n^2 \log n)$ time by computing all interpoint distances and by sorting these into increasing order. There are various methods of ensuring an expected running time of $0(n^2)$ under very weak assumptions (see, e.g. [4]).

Let GRE and OPT represent respectively the weight of the greedy and optimal solutions of an instance of the planar Euclidean matching problem. Reingold and Tarjan[1] have analyzed the worst case behaviour of this heuristic and have shown that, in worst case

$$\frac{\text{GRE}}{\text{OPT}} = \frac{4}{3} n^{\log_2 1.5} - 1.$$

This case occurs when all the points lie on a straight line. In the situation where the points are known to lie in a bounded region, this example has the property that, as n gets large, OPT$\to 0$ and GRE$\to$constant. This leaves open the possibility that GREEDY may not perform too poorly in bounded regions. This is the topic that is addressed in this paper.

In Section 2 we investigate the two dimensional problem. We obtain absolute worst case bounds on the behaviour of GREEDY in the unit square. These absolute bounds are of particular relevance to both of the applications mentioned above. The results obtained in this section are obtained directly, but may also be deduced from known results on packing circles into the unit square. This relationship is made precise in Section 3, where we use known sphere packing results to obtain bounds in all dimensions. The techniques illustrated appear to be useful in obtaining similar results for other geometric optimization problems.

Analysis of matching heuristics has received attention in several recent papers. In [5], variants of the GREEDY heuristic are analyzed for the non-Euclidean case. Papadamitriou[6] has given a probabilistic analysis of the so-called "strip" method for the planar matching problem. See also Steele[7]. The most comprehensive analysis of heuristics for matching has

†Research supported by N.S.E.R.C. grant No. A3013.

recently been performed by Supowit, Plaisted and Reingold [8]. Among their many results are a heuristic with a constant bound on the ratio GRE/OPT and several $0(n \log n)$ heuristics for the bounded planar case.

2. ANALYSIS OF GREEDY IN THE UNIT SQUARE

Suppose that n points are placed in the unit square. It is clear that there must be an upper bound on the distance d_n between the closest pair of points, and hence on the weight of the first edge selected by GREEDY. We begin by deducing such an upper bound.

A *planar triangulation* T of the set of points is a maximal set of edges between the points such that no two edges intersect at an interior point. T contains one unbounded face, the other faces are all triangles. We first show how the number of triangles in T is dependent on the number of extreme points on the convex hull of the set of points.

LEMMA 1

Every triangulation T with K points on the convex hull contains $2n - K - 1$ triangles.

Proof. Let V, E and F denote respectively the number of vertices, edges and faces of T. We include the unbounded face in F. Since T is a planar graph, Euler's formula holds:

$$V - E + F = 2. \tag{1}$$

We count the pairs (e, f) where e is an edge of T and f is a triangular face. Every such face has 3 edges so there are $3(F-1)$ such pairs. On the other hand, each edge appears in two triangular faces, except for the K edges on the convex hull, which appear in only one face. Hence

$$3(F-1) = 2E - K. \tag{2}$$

Combining (1) and (2) gives the required result.

We may now obtain a bound on d_n.

LEMMA 2

$$d_n \le \frac{\sqrt{2}}{\sqrt[4]{3}\sqrt{n}} + 0\left(\frac{1}{n}\right).$$

Proof. We first obtain two upper bounds on d_n that depend on K. The first bound is:

$$d_n \le \frac{4}{K} \tag{3}$$

where K is the number of points on the convex hull of the set of n points. This bound follows from the fact that the convex hull consists of K line segments, whose total length is at most 4 because the convex hull lies in the unit square. The second bound is:

$$d_n^2 \le \frac{4}{\sqrt{3}(2n-K-1)}. \tag{4}$$

This bound is obtained by taking a planar triangulation T of the points. By Lemma 1 this contains $2n - K - 1$ triangles. Each triangle has all sides of length at least d_n, and hence has area at least $d_n^2\sqrt{3}/4$. The triangles are non overlapping and fit into the unit square, hence

$$(2n - K - 1)\frac{\sqrt{3}}{4}d_n^2 \le 1,$$

and (4) follows. We equate the two r.h.s. of (3) and (4) to get a uniform bound independent of K. Setting

$$\frac{16}{K^2} = \frac{4}{\sqrt{3}(2n-K-1)},$$

and solving for K in the resulting quadratic equation gives

$$K = \frac{-4\sqrt{3} \pm \sqrt{[48 + 32\sqrt{3}(n-1)]}}{2} = 2\sqrt{(2\sqrt{3})}\sqrt{n} + 0(1).$$

The Lemma now follows upon substitution for K in (3).

Since Lemma 2 gives us a worst case bound on d_n, we may use it repeatedly to get a bound on GRE.

THEOREM 1

$$\text{GRE} \leq \frac{\sqrt{2}}{\sqrt[4]{3}}\sqrt{n} + 0(\log n), \quad \frac{\sqrt{2}}{\sqrt[4]{3}} = 1.074\ldots$$

Proof. Using Lemma 2 and setting $m = n/2$ we obtain:

$$\text{GRE} \leq \sum_{i=1}^{m} d_i \leq \sqrt{2} + \sum_{i=2}^{n}\left(\frac{\sqrt{2}}{\sqrt{(2i\sqrt{3})}} + 0\left(\frac{1}{i}\right)\right)$$

$$\leq \int_1^m \frac{dx}{\sqrt{(\sqrt{3}x)}} + 0(\log m) \leq \frac{2}{\sqrt[4]{3}}\sqrt{m} + 0(\log m),$$

thus the theorem follows.

We now turn attention to constructing a bad example for GREEDY in the unit square. Consider the set of points in the plane defined by:

$$L = \{(x, y) | (x, y) = (2i, 0) + (j, \sqrt{3}j), \text{ for all integers } i \text{ and } j\}.$$

L is a *lattice* and is known to give the densest packing of the plane by unit circles (see [9]). We consider sublattices L_n of L that are rectangular and contain 2^{2n+1} points of L. We define L_n by:

$$L_n = L\{(x, y) | 0 \leq x \leq 2^{n+1}_{-1}, \ 0 \leq y \leq (2^{n+1} - 1)\cdot\sqrt{3}\}.$$

These lattices are illustrated in Fig. 1. It may be verified that L_n does contain 2^{2n+1} points of L. It may also be verified that L is the lattice corresponding to a tesselation of the plane with equilateral triangles of side length 2. Thus each point has 6 nearest neighbours and the operation of GREEDY on this example is ambiguous. For $n \geq 2$, it is easily seen that L_n may be decomposed into 2^{2n-2} disjoint copies of the lattice L_1. We suppose that GREEDY chooses the edges $(2, 0), (3, \sqrt{3})$; $(0, 2\sqrt{3}), (1, \sqrt{3})$; and $(1, 3\sqrt{3}), (3, 3\sqrt{3})$ of L_1, as shown in Fig. 2. Further, we assume that GREEDY makes the corresponding choice in each of the disjoint copies of L_1 in L_n. The net result is that 3.2^{2n-1} points are matched with each matching edge of length 2. This leaves a lattice of 2^{2n-1} points that is isomorphic to L_{n-1} but has each interpoint distance doubled. We now recursively apply GREEDY to this lattice in exactly the way that was outlined above.

Let f_n denote the weight of the GREEDY matching on L_n. Then we have shown that:

$$f_n = 3.2^{2n-1} + 2f_{n-1}, \ n \geq 1$$

$$f_0 = 2.$$

This recurrence may readily be solved to give

$$f_n = 3.2^{2n} - 2^n.$$

Clearly, the optimal solution of the matching problem on L_n will match each of the 2^{2n+1} points to a nearest neighbour. If g_n denotes the weight of the optimal matching, then

$$g_n = 2^{2n+1}.$$

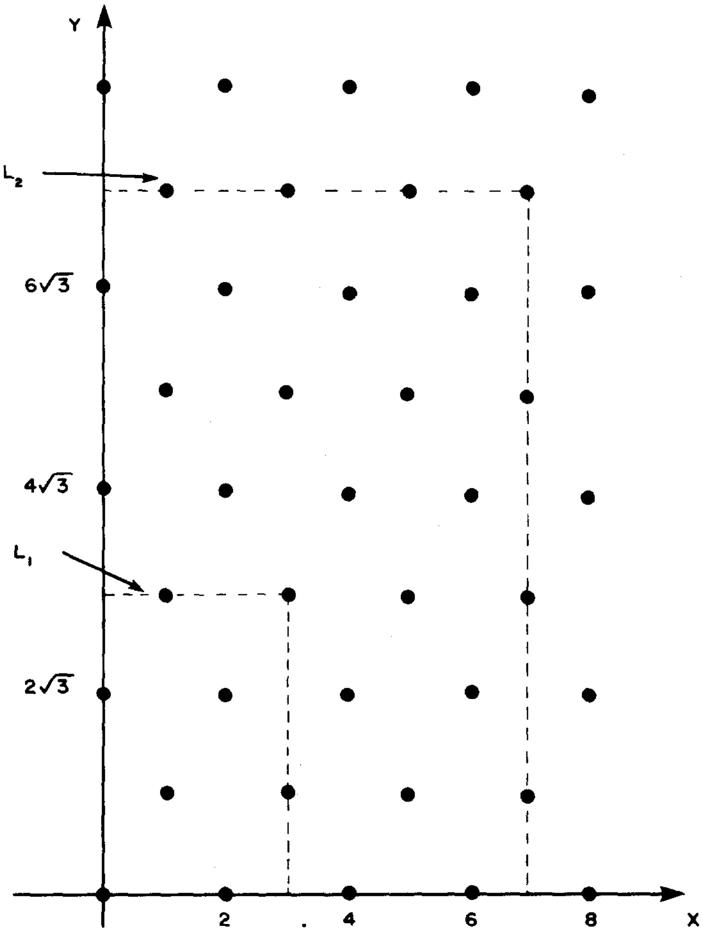

Fig. 1. Sublattices L_1 and L_2 of the lattice L.

Hence we can compute the ratio of the GREEDY solution to the optimal solution:

$$\frac{f_n}{g_n} = \frac{3}{2} - \frac{1}{2^{n+1}}.$$

Let us return to the problem of finding bad cases for GREEDY in the unit square, for sets of n points. We begin by scaling down the lattice L until n points fit inside the unit square. A straightforward argument shows that this can be done with interpoint distances of:

$$d_n = \frac{\sqrt{2}}{\sqrt[4]{3}\sqrt{n}} + 0\left(\frac{1}{n}\right).$$

Now the optimal solution will have weight

$$\mathrm{OPT} = \frac{n}{2} \cdot d_n = \frac{\sqrt{n}}{\sqrt[4]{12}} + 0(1) = 0.5372 \ldots \sqrt{n} + 0(1).$$

Thus by the preceding remarks, GREEDY will find a matching of weight:

$$\mathrm{GRE} = \frac{3}{2}\frac{\sqrt{n}}{\sqrt[4]{12}} + 0(1) = 0.8059 \ldots \sqrt{n} + 0(1).$$

A somewhat worse choice of lattice points of L can cause GREEDY to do even worse. The idea is similar to that described above, but the details are messy and are just sketched here.

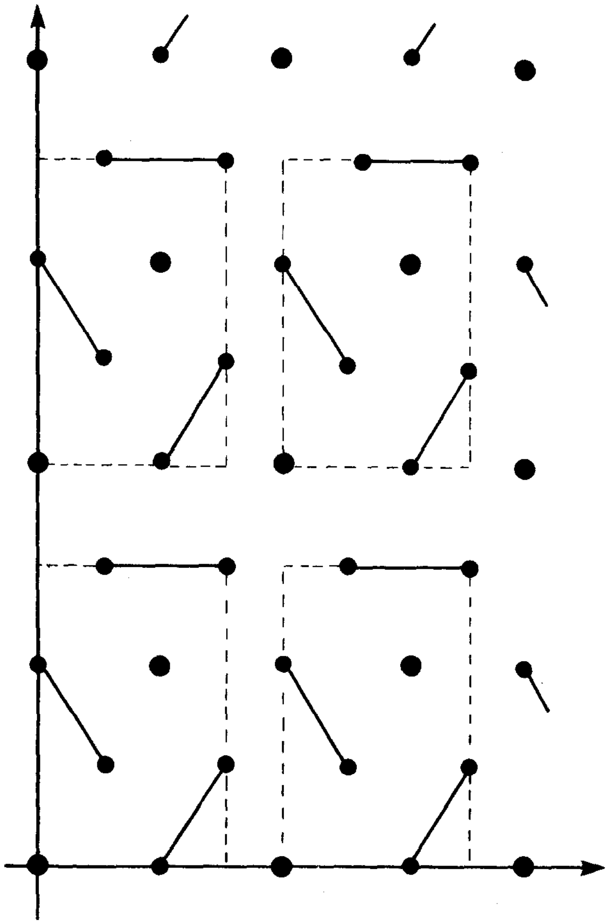

Fig. 2. GREEDY matching after first phase.

Consider the result of GREEDY matching lattice points (k, y) and $(k+2, y)$ of L for all $k \equiv 2$ or 5 (mod 6). Once these points are deleted from L, the remaining points again form a lattice isomorphic to L, but with interpoint distances increased by a factor of $\sqrt{3}$. Denote by h_n the weight of the GREEDY matching formed by making the above choices recursively on a set of 3^n points of L. Then 2/3 of the points are matched at each stage, and h_n satisfies

$$h_n = \frac{2}{3} \cdot 3^n + \sqrt{3} h_{n-1}, \qquad h_0 = 0.$$

Hence

$$h_n = \frac{2 \cdot 3^n}{\sqrt{3}(\sqrt{3}-1)} + 0(3^{n/2}).$$

In this case we find that

$$\text{GRE} = \frac{2}{\sqrt{3}(\sqrt{3}-1)} \cdot \text{OPT} + 0(1) = \frac{2}{\sqrt{3}(\sqrt{3}-1)} \cdot \frac{\sqrt{n}}{\sqrt[4]{12}} + 0(1)$$

$$= 0.8474 \ldots \sqrt{n} + 0(1).$$

This is the worst example of GREEDY in the unit square known to the author.

3. HIGHER DIMENSIONS

In this section we make use of known results for sphere packing in Euclidean space. The material on this subject that is used in this section is drawn from the excellent monograph of Rogers [9].

Let $C_{d,s}$ denote the hypercube in d dimensions of side s given by

$$C_{d,s} = \{(x_1, \dots, x_d) | 0 \le x_i \le s, \quad i = 1, 2, \dots, d\}.$$

The set $P_r = \{a_i \in C_{d,s} | 1 \le i \le n\}$ represents a *packing* of n spheres of radius r into $C_{d,s}$ if the n open spheres of radius r with centres P_r do not intersect and lie completely within $C_{d,s}$. The *packing density* of P_r in $C_{d,s}$ is denoted $\delta(P_r)$ and is given by:

$$\delta(P_r) = \frac{n\pi r^d}{s^d}.$$

The packing density is a well studied quantity, although tight upper bounds for it appear to be only known for $d = 2$. The most useful result for our purposes is the following theorem due to Blichfeldt, a proof of which is given in [9].

Blichfeldt's theorem

For any packing P_r of n spheres into $C_{d,1}$,

$$\delta(P_r) \le \frac{d+2}{2}\left(\frac{1}{\sqrt{2}}\right)^d. \tag{5}$$

We are looking for bounds on the minimum distance between n points in the unit hypercube. Let us denote this quantity by $x = x_{n,d}$. From the definitions we see that n spheres of radius $x/2$ can be packed into the cube $C_{d,1+x/2}$. Therefore we can apply Blichfeldt's theorem to show that

$$n\pi\left(\frac{x}{2}\right)^d \le \frac{d+2}{2}\left(\frac{1}{\sqrt{2}}\right)^d \left(1 + \frac{x}{2}\right)^d. \tag{6}$$

The quantity on the left of (6) is the volume of n spheres of radius $x/2$. The quantity on the right is the maximum packing density multiplied by the volume of $C_{d,1+x/2}$. Solving (6) for x we obtain

THEOREM 2

$$x_{n,d} \le \sqrt{2}\left(\frac{d+2}{2n\pi}\right)^{1/d} + 0\left(\frac{1}{n^{2/d}}\right).$$

Bounds for the total weight of the GREEDY solution in any dimension may be obtained from Theorem 2 in the same way as was demonstrated in Section 2 for $d = 2$. Note that the bound in Theorem 2 is not as sharp as that given in Lemma 2. The Ref. [9] contains a slightly sharper bound than (5) in the general case, and gives the strongest known bound for $d = 3$ as

$$\delta(P_r) \le 0.7797\dots \quad .$$

This gives the bound:

$$x_{n,3} \le 1.256 \dots \; n^{-1/3} + 0(n^{-2/3}),$$

which is somewhat sharper than Theorem 2.

Acknowledgement—The author gratefully acknowledges and thanks Edward Reingold for many helpful communications. An example similar to the first example of Section 2 was also discovered independently by him.

REFERENCES

1. E. Reingold and R. Tarjan, On a greedy heuristic for complete matching. Manuscript submitted for publication.
2. N. Christofides, *Worst case analysis of a new heuristic for the travelling salesman problem.* Tech. Report, G.S.I.A., Carnegie-Mellon, Penn. (1976).
3. E. Lawler, *Combinatorial Optimization.* Holt, Rinehart & Winston, New York (1976).
4. M. Rabin, *Probabilistic algorithms,* In *Algorithms and Complexity: New Directions and Recent Results,* (Edited by J. F. Traub). Academic Press, New York (1976).
5. D. Avis, Two greedy heuristics for the weighted matching problem. *Proc. 9th SE Conf. on Combinatorics, Graph Theory and Computing, Congressus Numerantium XXI,* pp. 65–76. Utilitas Mathematic Publishing, Winnipeg (1978).
6. C. H. Papadimitriou, The probabilistic analysis of matching heuristics. *Proc. 15th Allerton Conf. on Communication, Control and Computing,* pp. 368–378 (1977).
7. J. M. Steel, Subadditive growth functionals and non-linear growth in geometric probability Preprint. *Ann. Prob,* to appear.
8. K. Supowit, D. Plaisted and E. Reingold, Heuristics for weithted perfect matching. *12th SIGACT Conference,* pp. 398–419. ACM Publications (1980).
9. C. A. Rogers, *Packing and Covering.* Cambridge University Press (1964).