Information Processing Letters 18 (1984) 59–62
North-Holland 28 February 1984

# A PARTITIONING ALGORITHM FOR MINIMUM WEIGHTED EUCLIDEAN MATCHING

M.E. DYER

*Department of Mathematics and Statistics, Teesside Polytechnic, Middlesbrough, Cleveland TS1 3BA, United Kingdom*

A.M. FRIEZE

*Department of Computer Science and Statistics, Queen Mary College, University of London, London WC1E 7HU, United Kingdom*

Communicated by H. Whitfield
Received 2 June 1983
Revised 29 September 1983

We describe an algorithm which, given n = 2m points in the unit square, finds a matching of these points. We prove that, under the assumption that the points are uniformly distributed in the square, the algorithm has a fast expected running time, and it gives a matching with value close to the optimum with probability one as n tends to infinity.

*Keywords*: Euclidean matching, heuristics, probabilistic analysis of algorithms

## 1. Introduction

The Euclidean matching problem can be described as follows: We are given $n = 2m$ points $X_1, X_2, \ldots, X_n$ in the unit square $[0, 1]^2$. The distance $d_{ij}$ between $X_i$ and $X_j$ is the usual Euclidean metric. We wish to pair the points into m pairs so that the sum of the lengths of the lines joining up the pairs is as small as possible.

This is, of course, a special case of the weighted matching problem which was so elegantly solved by Edmonds [3,4]. However, Edmonds' algorithm can run in $\Omega(n^3)$ time, which in some situations can be excessive. Therefore, there has been interest recently in heuristics which have faster running time, but which do not necessarily produce an optimal solution (see [1,2,7–12]).

We shall assume that the n points are chosen independently and uniformly at random from the unit square, and we describe a class of partitioning algorithms which have the following properties. Let w(n) be an o(n) function of n which tends to infinity with n. We describe an algorithm $A_w$ which produces a matching $M_w$ of length $L_n$ in time $T_n$ where the random variables $L_n$, $T_n$ have the following properties:

$$\Pr\Big(\lim_{n\to\infty} L_n/L_n^* = 1\Big) = 1, \tag{1.1}$$

where $L_n^*$ is the length of the minimum matching $M^*$.

$$\Pr\big(T_n > cnw^2(1+\epsilon_n)\big) = O\Big(\frac{w}{n\epsilon_n}\Big) \tag{1.2}$$

for any $\epsilon_n > 0$. Here $cn^3$ is any upper bound on the running time of an algorithm to solve the matching problem exactly.

Our algorithm is almost identical to that given in [12]. In that paper the authors concentrated on a worst-case analysis, while here we do a probabilistic analysis in the spirit of Karp [6] and, more closely, Halton and Terada [5].

## 2. The algorithm

For any function w of n, as defined above, let $k = (n/w)^{1/2}$. Note that k is a function of n such

 

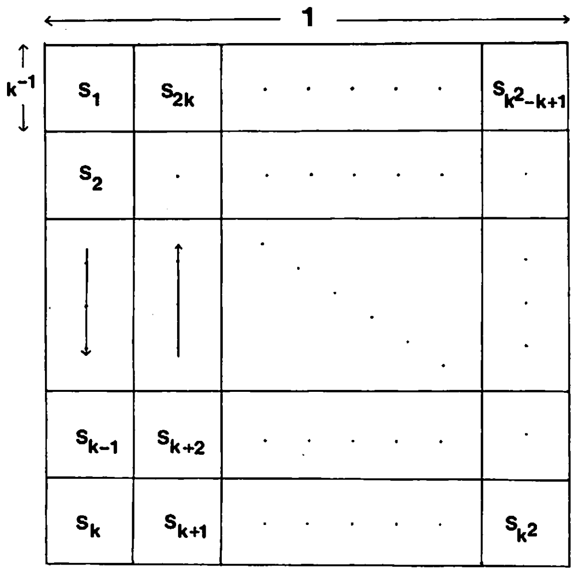


Fig. 1.

that $k = o(n^{1/2})$ and $k \to \infty$ as $n \to \infty$. For simplicity we will assume that k is integral, though this is not essential.

**Algorithm $A_w$.** Divide the unit square into $k^2$ subsquares, each of side $k^{-1}$. Number these squares $S_1, S_2, \ldots, S_{k^2}$ going alternately up and down columns (see Fig. 1).

*Step* A: Use Edmonds' algorithm to optimally match points in $S_j$, $j = 1, 2, \ldots, k^2$.

*Step* B: If there are unmatched points in any $S_j$ (i.e., $S_j$ contains an odd number of points and hence exactly one unmatched point), then use the 'strip heuristic' [12] to match the remaining points.

## 3. Error analysis of $A_w$

We shall show that

$$L_n - L_n^* < 18k^{-1}. \tag{3.1}$$

This will imply (1.1), since Papadimitriou [7] and Steele [10] have shown that there exists a constant β such that

$$\Pr\left(\lim_{n\to\infty} L_n^*/n^{1/2} = \beta\right) = 1.$$

Let $L_A$ denote the length of the matching produced in Step A. We show first that

$$L_A \leq L_n^* + (8 + 6\sqrt{2})k^{-1}. \tag{3.2}$$

To do this, we start with M* and construct from it a matching M of length L which satisfies

(a) every matching edge joins points in the same subsquare,
(b) each subsquare contains at most one unmatched point, (3.3)
(c) $L \leq L_n^* + (8 + 6\sqrt{2})k^{-1}$.

By construction, $L_A \leq L$ and (3.2) follows.

We now describe the construction of M from M*.

(i) First we eliminate 'long' edges, i.e., those that join points in non-adjacent squares in M*. Suppose some points of $S_j$ are so matched in M*. Let the *charge* for any point be half the length of its matching edge, so that the cost of the matching is the total charge. Divide these 'bad' points of $S_j$ into four groups according to which of the four subsquares $S_j^{(1)}$, $S_j^{(2)}$, $S_j^{(3)}$, $S_j^{(4)}$ they fall into, as shown in Fig. 2.

Match all points in $S_j^{(r)}$ together in an arbitrary fashion for r = 1, 2, 3, 4. There will be at most one point in each subsquare left over. Now in M* each point of this type must be charged at least $\frac{1}{2}k^{-1}$, but in the revised matching each such point is charged at most $\frac{1}{2}k^{-1}/\sqrt{2}$. Thus the revised matching has total cost less than M*. However, there will be up to four points in $S_j$ which will now become unmatched. These will be dealt with later.

(ii) We now deal with the remaining points and edges of M*. For a given j, the points in $S_j$ are

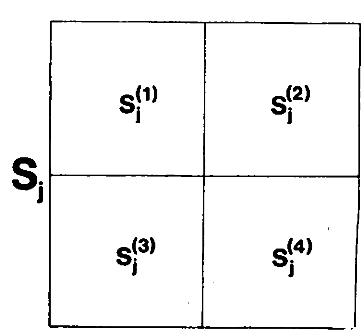


Fig. 2.

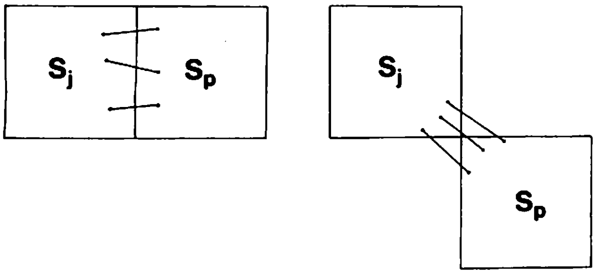


Fig. 3.

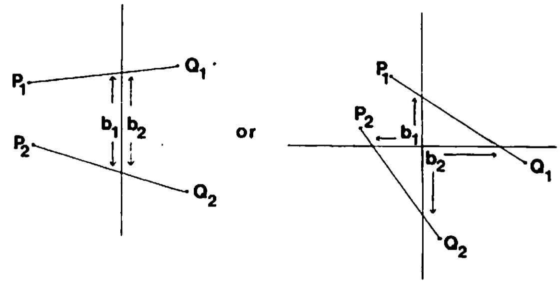


Fig. 5.

matched either with points of $S_j$ or an adjacent square. We must deal with the latter edges, which we denote by $C_j$. Consider those edges which go from $S_j$ to $S_p$. There are two possibilities as shown in Fig. 3.

Now the optimal matching edges do no cross (otherwise uncrossing any such pair would lead to a contradiction). Hence the edges of $C_j$ can be ordered around the boundary of $S_j$ such that there will be a portion of the boundary for each successive pair of edges in the ordering (see Fig. 3). Moreover, the ordering is consistent with that for $S_p$. Suppose then that $P_1, P_2, \ldots, P_t$ of $S_j$ are matched with $Q_1, Q_2, \ldots, Q_t$ of $S_p$ in this order and that $X_1, X_2, \ldots, X_t$ are where these edges cross the boundary of the two squares (see Fig. 4). We rematch $P_{2i-1}$ with $P_{2i}$, $Q_{2i-1}$ with $Q_{2i}$ for $i = 1, 2, \ldots, \lfloor \frac{1}{2}t \rfloor$. At most one point in each of $S_j$ and $S_p$ becomes unmatched in this way. By the triangle inequality, we have

$$P_1P_2 + Q_1Q_2 - P_1Q_1 - P_2Q_2 \leqslant b_1 + b_2$$

(see Fig. 5).

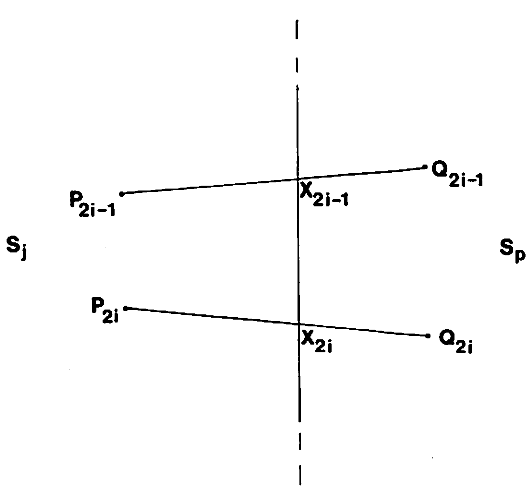


Fig. 4.

The analogous inequality will, of course, hold for all $P_i$, $Q_i$ and we deduce that the increase in the cost for rematching points of $S_j$ and $S_p$ in this way is at most half the sum of their boundaries if they meet at a corner and one quarter if they meet along an edge. Attributing one half of this increase to each square of the pair, we find that the increase in cost per square is no more than twice the perimeter of a square. Thus the total increase in length from this rematching is bounded by

$$k^2 \times 8k^{-1} = 8k.$$

We may leave at most eight points in each $S_j$ unmatched as a result of this, one for each of its adjacent subsquares.

(iii) As a result of (i) and (ii) there can be up to 12 unmatched points in $S_j$. Reduce this to at most one by matching arbitrarily in pairs. This can add at most $6k^{-1}\sqrt{2}$ to the length of the matching.

The matching M produced by (i), (ii) and (iii) clearly satisfies (3.3) and so we have proved (3.2).

Now the cost of the matching edges chosen in Step B is no more than $k^{-1}\sqrt{2}$ [12]. It follows from (3.2) that

$$L_n \leqslant L_A + k^{-1}\sqrt{2} \leqslant L_n^* + (8 + 7\sqrt{2})k^{-1},$$

which implies (3.1).

## 4. Time analysis of $A_w$

The number of points $n_j$ in subsquare $S_j$ is a binomial random variable with parameters n, p where $p = w/n$. It is a routine calculation to show that

$$\exp\{n_j^3\} < w^3 + 3w^2 + w.$$

Thus the expectation of the running time $T_A$ of Step A is $O(nw^2)$. Further calculation shows that the variance of $T_A$ is $O(nw^5)$. Inequality (1.2) then follows from the Chebychev inequality. (The running time of Step B is $O((n/w)\log(n/w))$ and therefore small by comparison.)

Note that we assume here that the 'floor' function $\lfloor \cdot \rfloor$ is available at unit cost for real numbers, so that the n points can be sorted into the $k^2$ subsquares in O(n) time. If this assumption is not made, there will be an additional $O(n \log(n/w))$ term in the running time, which becomes important if w grows very slowly with n (i.e., $w = o((\log n)^{1/2})$).

## References

[1] D. Avis, Worst-case bounds for the Euclidean matching problem, Internat. J. Comput. Math. Appl. 7 (1981) 251–257.

[2] D. Avis, A survey of heuristics for the weighted matching problem, Rept. No. SOCS-82.4, McGill University, 1982.

[3] J. Edmonds, Maximum matching and a polyhedron with 0–1 vertices, J. Res. National Bureau of Standards 69 (1965) 125–130.

[4] J. Edmonds, Paths, trees and flowers, Canad. J. Math. 17 (1965) 449–467.

[5] J.H. Halton and R. Terada, A fast algorithm for the Euclidean travelling salesman problem, optimal with probability one, SIAM J. Comput. 11 (1982) 28–46.

[6] R.M. Karp, Probabilistic analysis of partitioning algorithms for the travelling salesman problem in the plane, Math. Oper. Res. 2 (1977) 209–224.

[7] C.H. Papadimitriou, The probabilistic analysis of matching heuristics, Proc. 15th Ann. Allerton Conf. on Communication, Control and Computing (1977) pp. 368–378.

[8] E.M. Reingold and K.J. Supowit, Probabilistic analysis of divide-and-conquer heuristics for minimum weighted Euclidean matching, Networks, to appear.

[9] E.M. Reingold and R.E. Tarjan, On a greedy heuristic for complete matching, SIAM J. Comput. 10 (1981) 676–681.

[10] J.M. Steele, Subadditive Euclidean functionals and nonlinear growth in geometric probability, Ann Probab. 9 (1981) 365–376.

[11] K.J. Supowit, D.A. Plaisted and E.M. Reingold, Heuristics for weighted perfect matching, Proc. 12th Ann. ACM Symp. on Theory of Computing (1980) pp. 398–419.

[12] K.J. Supowit and E.M. Reingold, Divide-and-conquer heuristics for minimum weighted Euclidean matching, SIAM J. Comput. 12 (1983) 118–143.

[13] K.J. Supowit, E.M. Reingold and D.A. Plaisted, The travelling salesman problem and minimum matching in the unit square, Siam J. Comput. 12 (1983) 144–156.