PARTITIONING HEURISTICS FOR TWO GEOMETRIC MAXIMIZATION PROBLEMS

M.E. DYER

Department of Mathematics and Statistics, Teesside Polytechnic, Middlesbrough, Cleveland TS1 3BA, UK

A.M. FRIEZE

Graduate School of Industrial Administration, Carnegie-Mellon University, Pittsburgh, PA 15213, USA (on leave from Queen Mary College London)

C.J.H. McDIARMID

Institute of Economics and Statistics, Oxford University, Oxford OX1 3UL, UK

Received June 1984
Revised August 1984

Given n points randomly selected from a uniform distribution on the unit square, we describe linear-time partitioning heuristics which will construct a matching or a tour of these points. We show that the heuristics closely approximate the optimum values as $n \to \infty$. Hence we show that the asymptotic values of the maximum matching and tour are about $0 \cdot 3826n$ and twice this value respectively.

heuristics * partitioning

1. Introduction

Let $N = \{1, 2, \dots, n\}$, and E^2 be the plane $\mathbb{R}^2$ equipped with the usual Euclidean metric $d(\cdot, \cdot)$. Then, given a set $X = \{x_i \in E^2: i \in N\}$ with $n = 2m$ even, the geometric maximum matching problem is to divide N into a set of m disjoint pairs M (i.e. a matching) such that the sum of pairwise distances

$$L(M) = \sum_{(i,j) \in M} d(x_i, x_j) \tag{1.1}$$

is maximized. Similarly, the geometric maximum tour (or Hamiltonian circuit) problem is to determine a permutation T of N (a tour) such that the sum of successive distances

$$L(T) = \sum_{i=1}^{n-1} d(x_{T(i)}, x_{T(i+1)}) + d(x_{T(n)}, x_{T(1)}) \tag{1.2}$$

is maximized. (Here we do not assume n is even.)

The best (known) exact solution algorithms for these problems run in the potentially large times $O(n^3)$ [3] and $O(n^2 2^n)$ [6] respectively. It has been shown [9] that the geometric *minimum* tour problem (the Travelling Salesman Problem) is NP-hard, but no similar result appears to exist for the maximum tour problem although this is probably also NP-hard.

Here we examine faster heuristic methods when X is randomly sampled from a uniform distribution on the unit square $[0, 1]^2$. (Actually, our results have somewhat wider applicability. This is discussed in Section 5.) The minimization versions of these two problems have been well studied in this setting [2,5,8,10,11,12]. In particular it is known that there exist heuristics which run in *almost* linear time and produce solutions arbitrarily close to the optimum with probability close to 1 as $n \to \infty$. The maximization versions do not

0167-6377/84/$3.00 © 1984, Elsevier Science Publishers B.V. (North-Holland)

appear to have received as much attention. This is quite natural, since the minimizing versions are probably of more practical importance. However, it is interesting to investigate what results are available in the maximizing case. We will show that more satisfactory answers can be given, from the theoretical point of view, than in the minimizing case. We present *linear*-time partitioning heuristics which closely approximate the optimum with probability close to 1 as $n \to \infty$. We show that the asymptotic values of a maximum matching and tour are respectively about $0 \cdot 3826n$ and twice this value. In the minimizing case the values are only known to be $a\sqrt{n}$ and $b\sqrt{n}$, for some constants a, b and it is an open question whether $b = 2a$. The partitioning methods we present are not entirely new, see for example [2,5,8]. However, their application to *maximization* versions of geometric problems does not appear to have been considered previously.

2. Maximum matching

We will assume we have a set of $n = 2m$ points $x_1, x_2, \ldots, x_n$ in the unit square $[0, 1)^2$. Observe that if points are selected randomly from any probability density on $[0, 1]^2$, the event that any point has a co-ordinate equal to 1 has probability 0. Let $k(n)$ be any even-integer valued function of n such that $k \to \infty$ as $n \to \infty$ and $k = O(n^\alpha)$ for some $\alpha < \frac{1}{4}$. Let $K = \{1, 2, \ldots, k\}$. For notational convenience, we will let i' denote $(k + 1 - i)$ for any $i \in K$.

Matching Heuristic

Step A: Divide $[0, 1)^2$ into k^2 equal-area subsquares $S(i, j)$ for $i, j \in K$, each of side k^{-1}. Thus

$$S(i, j) = \{(x, y): \lfloor kx \rfloor + 1 = i, \lfloor ky \rfloor + 1 = j\}.$$

Sort the n points into these subsquares. Let $X(i, j)$ be the set of points in $S(i, j)$. Note that $S(i', j')$ is the subsquare obtained by reflecting $S(i, j)$ in the point $(\frac{1}{2}, \frac{1}{2})$, the centre of $[0, 1)^2$. Let $m \leftarrow \emptyset$.

Step B: **for** $i = 1, 2, \ldots, \frac{1}{2}k$ **do**
  for $j = 1, 2, \ldots, k$ **do**
    while $X(i, j)$ and $X(i', j')$ are nonempty **do**
      Choose any $x_p \in (i, j)$, $x_q \in X(i', j')$.
      Add (p, q) to M.
      Remove x_p from $X(i, j)$, x_q from $X(i', j)$.

Step C: Let $X' = \bigcup_{i, j \in K} X(i, j)$ be the remaining points. Find a matching M' of X' using any linear-time heuristic (e.g. arbitrary matching). Add M' to M.

This heuristic clearly runs in $O(n)$ time, assuming we have the floor function $\lfloor \cdot \rfloor$ available to do the sorting in Step A. In Step C we could, of course, improve on arbitrary matching by, for example using steps A and B a further constant number of times, halving the value of k at each repetition, before resorting to arbitrary matching.

3. Analysis of the matching heuristic

Let $c_{ij} = k^{-1}(i - \frac{1}{2}, j - \frac{1}{2})$ be the centre of $S(i, j)$. We show first that any matching may be reasonably approximated by assuming that all points of $X(i, j)$ are at c_{ij}. Now given a point x in $[0, 1)^2$ let x' be the centre c_{ij} of the square containing x. Note that $d(x, x') \leqslant k^{-1}/\sqrt{2}$. Also, given a matching M containing m pairs, let M' be the corresponding matching of the set of points obtained by moving each point x to x'. Then, by the triangle inequality, we have

$$|L(M') - L(M)| \leqslant \sqrt{2}\, mk^{-1} = nk^{-1}/\sqrt{2}. \qquad (3.1)$$

We now examine the behavior of an optimal matching when there is exactly one point at each c_{ij}. Let this have value $k^2 w(k)$. We note that the points c_{ij} are centrally symmetric with respect to the point $(\frac{1}{2}, \frac{1}{2})$, since we may verify $c_{i'j'} - (\frac{1}{2}, \frac{1}{2}) = (\frac{1}{2}, \frac{1}{2}) - c_{ij}$. We need the following

Lemma 3.1. *Let C be any finite set which is centrally symmetric with respect to the origin (i.e. if $c \in C$ then $-c \in C$). Then a maximum matching of C is given by matching c with $-c$ for all $c \in C$.*

Proof. A sufficient condition for a matching to be maximal (see [3]) is that there exists a set of real numbers $\{u(c): c \in C\}$ such that

(i) for all $c_1, c_2 \in C$, $u(c_1) + u(c_2) \geqslant d(c_1, c_2)$ and

(ii) if (c_1, c_2) is a matching edge, then $u(c_1) + u(c_2) = d(c_1, c_2)$. Let $u(c) = d(0, c)$, then (i) is simply the triangle inequality $d(c_1, 0) + d(0, c_2) \geqslant d(c_1, c_2)$, and (ii) is the equality $d(c, 0) + d(0, -c) = d(c, -c)$.

Now Lemma 3.1 and inequality (3.1) justify the approach used in the matching heuristic. Before proceeding with the analysis, we consider briefly the asymptotic behaviour of $w(k)$. For any point $x \in [0, 1]^2$, let $r(x)$ denote its distance form the point $(\frac{1}{2}, \frac{1}{2})$. Then from Lemma 3.1 we have

$$w(k) = k^{-2} \sum_{i,j} r(c_{ij}). \tag{3.2}$$

Letting $k \to \infty$ we see that

$$w(k) \to w^* = \int_0^1 \int_0^1 r(x, y) \mathrm{d}x \mathrm{d}y. \tag{3.3}$$

A straightforward computation yields

$$w^* = (\sqrt{2} + \log_e(1 + \sqrt{2}))/6 \simeq 0.3826.$$

Further fairly straightforward analysis also gives the bounds

$$0 < w^* - w(k) < k^{-1}/\sqrt{2}. \tag{3.4}$$

We now show that the optimal and heuristic matchings have the same asymptotic behaviour with high probability.

Let $N_{ij} = |X(i, j|$. Then the random variable N_{ij} has the binomial distribution with parameters n and k^{-2}. Let A_n be the event that

$$N_{ij} \geqslant \lceil (1 - k^{-1}) nk^{-2} \rceil = p \quad \text{for all } i, j.$$

The the Chernoff bounds (see, for example, [1]) yield

$$1 - \Pr(A_n) < k^2 \exp(-\tfrac{1}{2} nk^{-4}) = o(n^{-2}) \tag{3.5}$$

for any constant s.

Now let M_0 be an optimal matching, and M_H be the matching produced by our heuristic. Assume that the event A_n occurs. Then, by (3.1),

$$L(M_H) \geqslant pk^2 w(k) - \tfrac{1}{2} pk^2 (\sqrt{2} k^{-1})$$

and it follows using (3.4) that

$$L(M_H) > nw^* - 2nk^{-1}.$$

To deal with M_0, let Y be a subset of X with $|Y \cap S(i, j| = p$ for all i, j. Let M_Y be the set of pairs in M_0 which are contained in Y. Then, by (3.1),

$$L(M_Y) \leqslant pk^2 w(k) + \tfrac{1}{2} pk^2 (\sqrt{2} k^{-1})$$

and

$$L(M_0 \setminus M_Y) \leqslant (n - pk^2)\sqrt{2}.$$

It follows, using (3.4), that

$$L(M_0) < nw^* + 2nk^{-1}.$$

Thus, we have shown that if A_n occurs then

$$nw^* - 2nk^{-1} < L(M_H) \leqslant L(M_0) < nw^* + 2nk^{-1}. \tag{3.6}$$

The inequalities (3.5) and (3.6) tie down the behaviour of $L(M_H)$ and $L(M_0)$ rather tightly. In particular, if we were to pick the random points $x_1, x_2, \ldots$ sequentially in the natural way, then the Borel–Cantelli lemmas yield

Theorem 3.2. *As $n \to \infty$, almost surely $L(M_H)/L(M_0) \to 1$ and $L(M_0)/n \to w^* \simeq 0.3826$.*

Note that we could convert our heuristic into a linear *expected* time heuristic with *guaranteed* performance bounds as in (3.6) if, whenever the event A_n fails to occur, we call a polynomial optimizing algorithm to determine the matching.

By way of comparison, consider matching the points of X randomly in pairs to give a random matching M_R. We can compute the expected value of $L(M_R)$ to be

$$\begin{aligned} \mu_n = n\mu &= n(2 + \sqrt{2} + 5 \log_e(1 + \sqrt{2}))/30 \\ &\simeq 0.2607n. \end{aligned} \tag{3.7}$$

It follows by Hoeffding's extension of the Chernoff bounds (Theorem 1 of [7]) that as $n \to \infty$, almost surely

$$L(M_R)/n \to \mu \simeq 0.2607$$

and

$$L(M_R)/L(M_0) \to \mu/w^* \simeq 0.6814.$$

Thus, asymptotically, a random matching will have value around 68% of the optimal.

4. Maximum tour

We can construct a partition heuristic for the maximum tour problem which closely parallels the matching heuristic of Section 2. We will merely outline the method. Step A is identical with the matching heuristic. In Step B, when $x_{T(r)} \in X(i, j)$ we set, whenever possible, $x_{T(r+1)} \in X(i', j')$ (i.e. we 'move' successively between $S(i, j)$ and $S(i', j')$ until either $X(i, j)$ or $X(i', j')$ is exhausted). In Step C the partial tours formed in Step B are joined together and the remaining points 'patched in', in arbitrary fashion, to form a complete tour.

Again the heuristic can be implemented in $O(n)$ time. Let T_H denote our heuristic tour and T_0 an optimal tour, and note that $L(T_0) \leqslant 2L(M_0)$. Now, much as in the last section, if the event A_n defined there occurs then

$$2nw^* - 4nk^{-1} < L(T_H) \leqslant L(T_0) < 2nw^* + 4nk^{-1}. \tag{4.1}$$

Thus we have

Theorem 4.1. *As* $n \to \infty$, *almost surely* $L(T_H)/L(T_0) \to 1$ *and* $L(T_0)/n \to 2w^* \simeq 0.7652$.

We may also see by arguing as in the last section that a random tour will have about 68% of the length of a maximum tour.

5. Conclusions and comments

We have shown that partitioning heuristics are capable of closely approximating the solutions of the maximum matching and tour problems, assuming the points are uniformly distributed in the unit square. Moreover the approach yields the asymptotic behaviour of the optimal values.

Our results have many easy generalizations, for example to problems in higher numbers of dimensions. They readily generalize to any distribution centrally symmetric about some point, provided it tends rapidly enough to zero at large distances. (A uniform distribution on the unit square is clearly a special case.) They also generalize to any distance metric resulting from a norm. Also, we believe these, or similar, methods are applicable to other related problems.

Finally, observe that our results indicate that, in a certain theoretical sense, the maximization problems seem to be 'easier' than the corresponding minimization problems in this situation. This is not the first time that this type of behaviour has been observed. For example, in [4], Fisher, Nemhauser and Wolsey show that a similar phenomenon occurs in the worst-case analysis of heuristics for the maximum and minimum tour problems in (complete) weighted graphs without triangle inequality. It is not entirely clear why this difference exists, but it arises at least in part from using a ratio measure to assess the performance of heuristics.

Acknowledgements

We are grateful to B. King for useful discussions, and to the referee for pointing out the analogy with the results presented in [4].

References

[1] D. Angluin and L.G. Valiant, "Fast probabilistic algorithms for Hamiltonian circuits and matchings", *J. Comput. Syst. Sci.* **19**, 155–193 (1979).

[2] M.E. Dyer and A.M. Frieze, "A partitioning heuristic for minimum weighted Euclidean matching", *Inform. Proc. Letters* **18**, 59–62 (1984).

[3] J. Edmonds, "Maximum matching and a polyhedron with 0–1 vertices", *J. Res. Nat. Bur. Standards Sect. B* **69**, 125–130 (1965).

[4] M.L. Fisher, G.L. Nemhauser and L.A. Wolsey, "An analysis of approximation for finding a maximum weight Hamiltonian circuit", *Operations Research* **27**, 799–809 (1979).

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

[6] M. Held and R.M. Karp, "A dynamic programming approach to sequencing problems", *J. SIAM* **10**, 196–210 (1962).

[7] W. Hoeffding, "Probability inequalities for sums of bounded random variables", *J. Amer. Statist. Assoc.* **58**, 13–30 (1963).

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

[9] C.H. Papadimitriou, "The Euclidean travelling salesman problem is NP-complete", *Theoret. Comput. Sci.* **4**, 237–244 (1977).

[10] C.H. Papadimitriou, "The probabilistic analysis of matching heuristics", *Proc. of 15th Allerton Conference on Communication, Control and Computing* (1977) 368–378.

[11] J.M. Steele, "Subadditive Euclidean functionals and nonlinear growth in geometric probability", *Annals of Probability* **9**, 365–376 (1981).

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