A Linear Time Algorithm for a Matching Problem on the Circle

Carlos A. Cabrelli   Ursula M. Molter

Departamento de Matemática
Facultad de Ciencias Exactas y Naturales
Universidad de Buenos Aires
Ciudad Universitaria, Pabellón I
1428 Capital Federal
ARGENTINA*

KEYWORDS: Algorithms, Computational Complexity, Minimal Matching.

*Partially supported by Grants UBACyT EX048 and CONICET, PIA 6246/96

1 Introduction

Given two ordered sets of n points on the line $A = \{a_1, \ldots, a_n\}, B = \{b_1, \ldots, b_n\}$, a matching between them is a bijection from A onto B. Each matching is characterized by a permutation σ of $\{1, \ldots, n\}$ (i.e., to each a_i there is a corresponding $b_{\sigma(i)}$). The cost of the matching pair $(a_i, b_{\sigma(i)})$ is the distance between the points of the pair. The cost of the matching, is the sum of the cost of each pair in the matching.

The minimal matching problem is to find a matching which minimizes the cost.

It is known, that when the sets are on the line, a minimal matching is given by the identity permutation, that is $(a_1, b_1), (a_2, b_2), \ldots, (a_n, b_n)$ [4], [11]. Actually, there could be many minimal matchings, for example, if the point sets are separated.

Let us now consider the same problem in the circle. The points here are ordered by its angle, and the distance between two points is the angle between them. This problem was studied in [4], [11], [7] and [2]. In [10] it is shown that when the sets are not necessarily ordered, to find the minimal cost matching requires at least $\Omega(n \log n)$ operations. When the sets are ordered, they describe an $\mathcal{O}(n \log n)$ algorithm to find the minimal cost matching.

In this paper we describe a simple linear time algorithm to find a minimal matching for two ordered sets in the circle. This is done by showing that the problem can be reduced to a weighted median problem which has been proven to be of linear complexity.

The minimal matching problem can be related to the computation of the Kantorovich-Hutchinson distance between measures ([3]).

2 The matching problem on the unit circle

Following the notation in [10], let $A = \{a_1, \ldots, a_n\}$ and $B = \{b_1, \ldots, b_n\}$ be two ordered sets of points on the unit circle, $0 \leq a_i, b_i < 2\pi$, $C = \{c_1, \ldots, c_{2n}\}$ be the sorted union of A and B and $\ell(a, b)$ the distance between the points a and b.

Define for $\alpha \in [0, 2\pi)$

$$\begin{aligned} F_A(\alpha) &= \#\{i : a_i < \alpha\} \\ F_B(\alpha) &= \#\{i : b_i < \alpha\} \\ \text{and} \qquad F &= F_A - F_B. \end{aligned}$$

In [10] it was shown that the cost of a minimal matching is given by:

$$\min_{0 \le s \le 2\pi} \int_0^{2\pi} |F(x) - F(s)| \, dx. \tag{2.1}$$

Since F takes only a finite number of values $y_1 = F(c_1), \ldots, y_{2n} = F(c_{2n})$ then (2.1) becomes

$$\min_{1 \le j \le 2n} \sum_i |y_i - y_j| d_i, \tag{2.2}$$

where $d_1 = \ell(c_1, c_{2n})$ and $d_i = \ell(c_i, c_{i-1})$ $i = 2, \ldots, 2n$.

3 The weighted median problem

Given $x_1, \ldots, x_m$, m unsorted real numbers with associated weights $w_1, \ldots, w_m$, the weighted median problem is to find

$$\min_{x \in \mathcal{R}} \sum_{i=1}^{m} w_i |x - x_i|.$$

In the case that all the weights are equal to one, the problem consists in finding the median of m numbers.

The solution will always be given by one of the points, therefore the problem becomes to solve

$$\min_{1 \le j \le m} \sum_{i=1}^{m} w_i |x_j - x_i|.$$

The weighted median problem is known to be a $\mathcal{O}(n)$ problem.

Linear time algorithms for the median problem are found in [1], [6], [8], [9] and [5] provides a comparative study of these algorithms.

4 A linear time algorithm for the matching problem on the circle

From formula (2.2), it is clear that the minimal matching turns out to be a weighted median problem, where the points x_j are the values $F(c_j)$ and the weights w_j the distance between two consecutive points on the circle.

As mentioned before, in the case that the sorted points are on the line, the minimal matching is found by pairing the numbers in order i.e. (a_i, b_i). When looking at the problem in the circle, in [10] it is shown that one can transform the problem into a problem on the line, by cutting the circle at some particular point. Any solution to (2.2). is a suitable point for cutting. The reason for this is rather technical, and we refer the reader to [10]. In that paper, to find that point, they sort the values of $F(c_j)$.

However, the cutting point can be found using the weighted median algorithm, which avoids sorting the values of $F(c_j)$ and therefore downsizes the complexity to linear time. The algorithm can be described by:

Given the ordered points c_j,

1. Calculate the values $y_j = F(c_j) = \#\{s : a_s < c_j\} - \#\{s : b_s < c_j\}$.
2. Calculate $d_j = \ell(c_{j-1}, c_j)$.
3. Use $\{y_j\}$ and $\{d_j\}$ as input for any of the weighted median algorithms. The output will be a pair $(j_0, c_{\min})$ where y_{j_0} is the point that produces the minimum cost and $c_{\min} = \sum w_i |y_i - y_{j_0}|$ will be the minimum cost.
4. Cut the circle at c_{j_0} and solve the problem on the line, i.e. match the elements in the sets sequentially starting at c_{j_0} and moving clockwise.

5 Example

Let us consider the same example as in [10]. Let $A = \{0, 3, 5, 5\}$ and $B = \{1, 1, 2, 4\}$ be two sets of points in $[0, 2\pi)$.

Then $C = \{0, 1, 1, 2, 3, 4, 5, 5\}$ and $F = \{0, 1, 1, -1, -2, -1, -2, -2\}$.

Figure I

The j^{th} element of the set F corresponds to $F(c_j)$ and finally the set $D = \{2\pi - 5, 1, 0, 1, 1, 1, 1, 0\}$ is the set of all distances, that is it's j^{th} element is $d_j = \ell(c_{j-1}, c_j)$. We then have the table:

Table I

The entries of the last row of the table correspond to $\sum_j d_j|F(c_j) - \gamma|$ for all possible values of $\gamma \in F$. Then the minimum value, as shown in the table, is $2\pi - 1$ and corresponds to $s \in [0, 2\pi)$ where $F(s) = -1$. We see that this occurrs at c_4 and c_6. That means that we can cut the circle at any point in the intervals $(1, 2]$ and $(3, 4]$ and then the minimal matching will be given by $< 3, 2 >, < 5, 4 >, < 5, 1 >, < 0, 1 >$.

6 Acknowledgments

We want to thank Ricardo Baeza-Yates for pointing us out to the reference [5] and for suggesting to write up this result.

References

[1] C. Bleich and M. L. Overton. A linear-time algorithm for the weighted median problem. Technical Report 75, Department of Computer Science, New York University, 1983.

[2] S. R. Buss and P. N. Yianilos. Linear and $O(n \log n)$ Time Minimum-Cost Matching Algorithms for Quasi-convex Tours. In *SODA94*, 65–76, 1994.

[3] C. A. Cabrelli and U. M. Molter. The Kantorovich metric between probability measures on the circle. *Journal of Computational and Applied Mathematics*, 57:345–361, 1995.

[4] H. C. Chen and A. K. C. Wong. Generalized texture representation and metric. *Comput. Vision Graphics Image Process.*, 23:182–206, 1983.

[5] C. Gurwitz. Weighted median algorithms for L_1 approximation. *BIT*, 30:301–310, 1990.

[6] D. B. Johnson and T. Mizoguchi. Selecting the kth element in $X + Y$ and $X_1 + X_2 + \ldots + X_m$. *SIAM Journal on Computing*, 7:147–153, 1978.

[7] R. M. Karp and S-Y. R. Li Two special cases of the assignment problem. *Discrete Mathematics*, 13:129–142, 1975.

[8] A. Reiser. A linear selection algorithm for sets of elements with weights. *Information Processing Letters*, 7:159–162, 1978.

[9] M. I. Shamos. Geometry and statistics. In J. F. Traub, editor, *Algorithms and Complexity: New Directions and recent Results*, pages 251–280. Academic Press, 1976.

[10] M. Werman, S. Peleg, R. Melter, and T. Y. Kong. Bipartite graph matching for points on a line or a circle. *Journal of Algorithms*, 7:277–284, 1986.

[11] M. Werman, S. Peleg, and A. Rosenfeld. A distance metric for multidimensional histograms. Technical Report 90, CAR-TR-90, Center for Automation Research, University of Maryland, 1984.

Table I

j	c_j	$F(c_j)$	d_j	$\lvert F(c_j)-0\rvert$	$\lvert F(c_j)-1\rvert$	$\lvert F(c_j)-(-1)\rvert$	$\lvert F(c_j)-(-2)\rvert$
1	0	0	$2\pi-5$	0	1	1	2
2	1	1	1	1	0	2	3
3	1	1	0	1	0	2	3
4	2	-1	1	1	2	0	1
5	3	-2	1	2	3	1	0
6	4	-1	1	1	2	0	1
7	5	-2	1	2	3	1	0
8	5	-2	0	2	3	1	0
				7	$2\pi+5$	$2\pi-1$	$4\pi-5$
						min	