Computational Geometry 46 (2013) 203–212

Contents lists available at SciVerse ScienceDirect

Computational Geometry: Theory and Applications

www.elsevier.com/locate/comgeo

Covering a bichromatic point set with two disjoint monochromatic disks

S. Cabello [a,1], J.M. Díaz-Báñez [b,*,2], P. Pérez-Lantero [c,3]

[a] *Department of Mathematics, FMF, University of Ljubljana, Slovenia*
[b] *Departamento Matemática Aplicada II, Universidad de Sevilla, Spain*
[c] *Escuela de Ingeniería Civil Informática, Universidad de Valparaíso, Chile*

ARTICLE INFO

Article history:
Received 18 February 2011
Accepted 27 June 2012
Available online 4 July 2012
Communicated by K. Clarkson

Keywords:
Geometric optimization
Covering problem
Approximation algorithms
Disks

ABSTRACT

Let P be a set of n points in the plane in general position such that its elements are colored red or blue. We study the problem of finding two disjoint disks D_r and D_b such that D_r covers only red points, D_b covers only blue points, and the number of elements of P contained in $D_r \cup D_b$ is maximized. We prove that this problem can be solved in $O(n^{11/3} \operatorname{polylog} n)$ time. We also present a randomized algorithm that with high probability returns a $(1-\varepsilon)$-approximation to the optimal solution in $O(n^{4/3}\varepsilon^{-6} \operatorname{polylog} n)$ time.

1. Introduction

In data mining and classification problems, a natural method for analyzing data is to select prototypes representing different data classes. A standard technique for achieving this is to perform cluster analysis on the training data [11,16]. The clusters can be obtained by using simple geometric shapes such as disks or boxes. Aronov and Har-Peled [2], Eckstein et al. [12], and Liu et al. [18] considered disks and axis-aligned boxes for the selection problem. Aronov and Har-Peled [2] studied the following problem: *Given a bicolored set of n points in the plane, find a disk that contains the maximum number of red points without containing any blue point*. They propose an algorithm to solve this problem optimally in $O(n^2 \log n)$ time, and also provide a $(1-\varepsilon)$-approximation algorithm that needs near-linear time. This type of classification is *asymmetric* in the way red and blue points are treated. In this paper, we consider a *symmetric two-class* version, where we want to find a witness set for each color. We next formalize the problem.

Let P be a set of n points in the plane such that its elements are colored red or blue. Denote by R (resp. B) the set of red (resp. blue) elements of P. We say that $Y \subset \mathbb{R}^2$ is *red* (resp. *blue*) if Y contains only red (resp. blue) elements of P, and Y is monochromatic if it is either red or blue. In this paper we study the following problem:

The two disjoint disks problem (2DD-problem): find a red disk D_r and a blue disk D_b such that D_r and D_b are disjoint and $|D_r \cap R| + |D_b \cap B|$ is maximized.

* Corresponding author.
E-mail addresses: sergio.cabello@fmf.uni-lj.si (S. Cabello), dbanez@us.es (J.M. Díaz-Báñez), pablo.perez@uv.cl (P. Pérez-Lantero).

[1] Partially supported by the Slovenian Research Agency, program P1-0297.
[2] Partially supported by Project FEDER MEC MTM2009-08625, and the ESF EUROCORES programme EuroGIGA-ComPoSe IP04-MICINN Project EUI-EURC-2011-4306.
[3] Partially supported by Project FEDER MEC MTM2009-08625, and grant FONDECYT 11110069.

0925-7721/$ – see front matter
http://dx.doi.org/10.1016/j.comgeo.2012.06.002

We provide algorithms to solve the 2DD-problem optimally and approximately. It is easy to see that the 2DD-problem can be solved optimally in $O(n^5)$ time. We reduce this running time to $O(n^{11/3}\operatorname{polylog} n)$. This result is described in Section 2. We also provide a randomized approximation scheme that, with probability at least $1-O(1/n)$, returns a $(1-\varepsilon)$-approximation to the optimal solution in $O(n^{4/3}\varepsilon^{-6}\operatorname{polylog} n)$ time. Under the assumption that at least a constant fraction of the points is covered in the optimal solution, the $(1-\varepsilon)$-approximation can be obtained in $O(n\varepsilon^{-13}\operatorname{polylog} n)$ time. This approximation algorithm is described in Section 3. In our algorithms, we did not try to improve the exponents of ε or $\log n$.

We next discuss variants of the 2DD-problem that have been considered previously. If in the 2DD-problem we do not restrict D_r and D_b to be disjoint, then we can solve the problem considering each color separately. First, we find the disk covering the maximum number of red points and no blue point, and after that, the disk that contains the maximum number of blue points and no red point. These two problems can be solved optimally in $O(n^2\log n)$ time, or a $(1-\varepsilon)$-approximation can be obtained in near-linear time with high probability [2].

Another variant of the 2DD-problem that permits intersection of the disks but penalizes it is the problem of finding disks D_r and D_b that maximize $|(D_r\setminus D_b)\cap R|+|(D_b\setminus D_r)\cap B|$. This criterion is considered for axis-aligned boxes in [5]. An optimal solution for this variant can be found in $O(n^2)$ time using a key observation and known approaches. Namely, suppose that (D_r, D_b) is an optimal solution, and denote by ℓ the radical axis of D_r and D_b. Let π_1 and π_2 be the open half-planes bounded by ℓ such that $((D_r\setminus D_b)\cap R)\subset\pi_1$ and $((D_b\setminus D_r)\cap B)\subset\pi_2$. Note that $(D_r\setminus D_b)\cap R=\pi_1\cap R$ and $(D_b\setminus D_r)\cap R=\pi_2\cap B$ because the pair (D_r, D_b) is optimal. Since both half-planes bounded by the line ℓ are disks with infinite radii, the problem reduces to finding a line ℓ such that the number of red points to one side of ℓ plus the number of blue points to the other side is maximized. This latter problem is known as the *Weak Separation Problem* [7,13,15] and can be solved in $O(n^2)$ time in the worst case [15]. Moreover, it was proven in [4] that the Weak Separation Problem is 3SUM-hard [14].

Another variant of the 2DD-problem is the problem of finding two unit disks D_r and D_b with disjoint interiors, but not necessarily monochromatic, such that $|D_r\cap R|+|D_b\cap B|$ is maximized. Note that in this variant there are two differences with the 2DD-problem: the disks are unitary and do not need to be monochromatic. This variant was considered in [6], where an $O(n^{8/3}\log^2 n)$-time algorithm is described.

Notation. Given two points p and q we denote: by $\overline{pq}$ the straight line segment joining p and q, by $\ell(p,q)$ the straight line containing both p and q, by $\operatorname{bis}(p,q)$ the line perpendicular to $\overline{pq}$ passing through the midpoint of $\overline{pq}$ (i.e. the bisector of p and q), and by $D(p,q)$ the disk centered at p with radius equal to the length of $\overline{pq}$. Given a region $S\subset\mathbb{R}^2$, let δS denote the boundary of S.

General position. We assume general position, that is, there are no four cocircular points in P, neither three collinear points. We relax the definition of our problem by allowing the boundary of the red disk (resp. blue disk) to contain one blue point (resp. one red point). A solution to the relaxed problem induces a solution to the original one by shrinking the disks slightly.

2. An exact algorithm

In this section we provide an exact $O(n^{11/3}\operatorname{polylog} n)$-time algorithm to solve the 2DD-problem. First, we will show that we only need to consider a certain type of solutions, thus obtaining a discretization of the problem. Secondly, we will consider a decision version of the problem, where we want to decide if there exists a solution covering a prescribed number of points of each color. Finally, we will discuss how to find an optimal solution to the 2DD-problem.

2.1. Discretization

It is not hard to see that a simple discretization of our problem in which the boundary of each disk contains three points of P, or two diametrically opposed points, is not possible. In order to obtain an appropriate discretization we will use the following lemmas.

Lemma 2.1. *If the points p, q, o, p', q', and o' are such that $o\in\operatorname{bis}(p,q)$, $o'\in\operatorname{bis}(p',q')$, and $D(o,p)\cap D(o',p')=\emptyset$, then both o and o' can be moved simultaneously along* $\operatorname{bis}(p,q)$ *and* $\operatorname{bis}(p',q')$*, respectively, so that at every moment $D(o,p)\cap D(o',p')=\emptyset$, until o or o' reaches infinity.*

Proof. Notice that both p' and q' are in the same half-plane bounded by $\ell(p,q)$, or both p and q are in the same half-plane bounded by $\ell(p',q')$. Assume w.l.o.g. the former case. We prove now that such a movement exists so that o reaches infinity. Denote by π_1 and π_2 the open half-planes bounded by $\ell(p,q)$ and suppose w.l.o.g. that $p',q'\in\pi_1$. Refer to Fig. 1(a). Let h_o be the half-line starting at o such that $h_o\subset\operatorname{bis}(p,q)$ and $h_o\cap\pi_2$ is unbounded. Analogously define $h_{o'}$ as the half-line starting at o' such that $h_{o'}\subset\operatorname{bis}(p',q')$ and $h_{o'}\cap\pi_1$ is unbounded. If $D(o',p')\subset\pi_1$ then the result follows by moving only o through h_o in direction to infinity. Otherwise, let $\varepsilon>0$ be a small enough value and $u(o')\in h_o$ be the furthest point from o such that the disk $D(u(o'),p)$ is exterior tangent to $D(o',p')$ (see Fig. 1(b)). We first move o in direction to $u(o')$ until the distance between o and $u(o')$ is equal to ε. After that we simultaneously move o' through $h_{o'}$ in direction to infinity and o