Computational Geometry 61 (2017) 24–37

Contents lists available at ScienceDirect

Computational Geometry: Theory and Applications

www.elsevier.com/locate/comgeo

Separability of imprecise points ☆

Farnaz Sheikhi [a], Ali Mohades [a,*,1], Mark de Berg [b,1], Ali D. Mehrabi [b]

[a] Computer Science Department, Amirkabir University of Technology, Iran
[b] Department of Computer Science, Eindhoven University of Technology, the Netherlands

ARTICLE INFO

Article history:
Received 27 March 2015
Accepted 18 October 2016
Available online xxxx

Keywords:
Imprecise points
Separator

ABSTRACT

An imprecise point p in the plane is a point represented by an imprecision region $\mathcal{I}_p$ indicating the set of possible locations of the point p. We study separability problems for a set R of red imprecise points and a set B of blue imprecise points, where the imprecision regions are axis-parallel rectangles and each point $p \in R \cup B$ is drawn uniformly at random from $\mathcal{I}_p$. Our results include algorithms for finding *certain separators* (which separate R from B with probability 1), *possible separators* (which separate R from B with non-zero probability), *most likely* separators (which separate R from B with maximum probability), and *maximal* separators (which maximize the expected number of correctly classified points).

1. Introduction

Separability problems are a natural class of problems arising in the analysis of categorical geometric data. In a bichromatic separability problem one is given a set of n points in $\mathbb{R}^d$, each of which is categorized as either *red* or *blue*, and the goal is to decide whether the red points can be separated from the blue points by a *separator* from a given class of geometric objects. If a separator exists, one may also be interested in finding all separators or the separator minimizing a given cost function. In this paper we are interested in separability problems on *imprecise points*, that is, points of which the location is not know precisely. In the remainder of this introduction we first give an overview of the known results on separability problems for normal (that is, precise) points, then we discuss some existing work from computational geometry on imprecise points, and finally we state our results on separability problems for imprecise points.

Separability problems. For separators in the form of a hyperplane, the decision version of the separability problem can be solved by linear programming in $O(n)$ time, when the dimension d is a fixed constant, as was observed by Megiddo [23] already 30 years ago. Since then various classes of separators have been studied, mostly for the 2-dimensional version of the problem. O'Rourke et al. [27] studied separability problems for circular separators. They proposed an $O(n)$-time algorithm for deciding whether such a separator exists. Furthermore, they showed that the smallest and the largest separating circle (if they exist) can be found optimally in $O(n)$ and $O(n \log n)$ time, respectively. The problem of finding a convex polygon with minimum number of edges separating the two point sets was solved by Edelsbrunner and Preparata [10] in $O(n \log n)$

☆ A preliminary version appeared in the 14th Scandinavian Symposium and Workshop on Algorithm Theory (SWAT) 2014.

* Corresponding author.
E-mail address: mohades@aut.ac.ir (A. Mohades).

[1] M. de Berg and A. D. Mehrabi were supported by the Netherlands Organization for Scientific Research (NWO) under grants 024.002.003 and 612.001.118, respectively.

http://dx.doi.org/10.1016/j.comgeo.2016.10.001
0925-7721/

time. Fekete [12] showed that separating two point sets by a simple polygon with minimum number of edges is NP-hard, and a polynomial-time $O(\log n)$-approximation algorithm was presented by Mitchell [24].

Other results on the separability problem concern strips and wedges as separators [16,2,29]. Hurtado et al. [16] showed that deciding whether two point sets can be separated by a strip or a wedge can be done in $O(n \log n)$ time, which is optimal [2]. Hurtado et al. [16] also showed that all orientations of the separating strips can be found in $O(n \log n)$ time; after this, the separating strip with the minimum and maximum width can be computed in $O(n)$ and $O(n \log n)$ time, respectively. Moreover, they showed that the separating wedge with the minimum and maximum angle can be computed in $O(n \log n)$ time. A thorough study of these types of separators is presented by Seara [29].

Inspired by the reconstruction of buildings from LIDAR[2] data, Van Kreveld et al. [17] presented an $O(n \log n)$-time algorithm to compute all orientations for which a separating rectangle exists. This was extended to L-shaped separators by Sheikhi et al. [30], who presented a worst-case optimal $O(n^2)$-time algorithm to compute all orientations for which an L-shaped separator exists. They also gave an output-sensitive algorithm running in $O(n^{8/5+\varepsilon} + k \log k)$ time, where k is the number of reported orientations and $\varepsilon > 0$ is any fixed constant.

Obviously it is not always possible to separate the given point sets by a separator of the given type. Houle [14,15] therefore introduced *weak separability*, where the goal is to maximize the number of correctly classified points. For example, for linear separability the weak separability problem asks for a line ℓ that maximizes the sum of the number of red points to the right of ℓ and the number of blue points to the left of ℓ. (A separator that correctly classifies all points is then called a *strong separator*.) For linear separators, Houle [14] showed that the weak separability problem can be solved in $O(n^2)$ time. Following that, Everett et al. [11] solved this problem in $O(nk \log k + n \log n)$ time, where k is the number of misclassified points. Chan [6] presented an $O((n + k^2) \log n)$-expected time algorithm for this problem, and Bereg et al. [4] showed that the problem is 3SUM-hard. Cortés et al. [7] presented an $O(n^2 \log n)$-time algorithm for the weak separability problem for a strip. Aronov et al. [3] studied the problem of measuring the quality of a weak separator according to how much work it would take (based on several distance functions) to move the misclassified points across the separator in order to make it a strong separator.

Imprecise points. In data-analysis problems involving geometric data, the data is typically obtained by GPS, LIDAR, or some other imprecise measuring technology. Ideally, one would like to take this imprecision into account when analyzing the data. Within the computational-geometry literature, several models have been proposed to handle imprecision [8,18,25]. The most popular models associate with each data point p an *imprecision region* $\mathcal{I}_p$, which indicates the set of possible locations of p. Typical choices for the imprecision regions are disks [28], axis-parallel rectangles or squares [18], and horizontal segments [18]. Horizontal segments model the situation where there is imprecision in only one of the coordinates, and rectangles or squares model the situation where the coordinates come from independent measurements. A point p with an associated imprecision region $\mathcal{I}_p$ is often called an *imprecise point*. Löffler [18] and Löffler and Van Kreveld [19,20] studied several classical computational-geometry problems on imprecise points. In most problems they wanted to find certain "extremal" structures, such as the largest or smallest possible convex hull. De Berg et al. [5], on the other hand, studied the question whether a given structure is possible (for example, whether a given subset of the points could be the convex hull).

Some recent work on imprecise (or: uncertain) points considers probabilistic aspects. Suri et al. [31] presented two models for this. In the *unipoint model* each point has a fixed location and an associated probability of existence. In other words, one is given a set of potential locations, and at each potential location a point exists with a certain probability. In the *multipoint model*, each point has several potential locations with an associated probability. Suri et al. then studied the problem of finding the *most likely convex hull*, that is the convex hull with the maximum probability of occurrence. In the unipoint model they proposed an $O(n^3)$-time algorithm for dimension $d = 2$, and showed that the problem is NP-hard for $d \geqslant 3$. In the multipoint model the problem is NP-hard even for $d = 2$. Agarwal et al. [1] proposed exact and approximation algorithms to compute the probability of a query point lying inside the convex hull of the input.

Our results. In this paper we study various separability problems for imprecise points. We extend the region-based imprecision model to include probabilistic aspects. More precisely, we assume each point p is drawn from its imprecision region $\mathcal{I}_p$ according to some distribution. In the current paper, we consider the uniform distribution. Throughout the paper we assume all the input imprecision regions are (relatively) open regions. Given a set R of red imprecise points and a set B of blue imprecise points, and a class of separators, we then wish to find

- a *certain separator*, which separates R from B with probability 1;
- a *possible separator*, which separates R from B with non-zero probability;
- a *most likely separator*, which separates R from B with maximum probability;
- a *maximal separator*, which is a weak separator that maximizes the expected number of correctly classified points.

In these problems, we require our separators to be strict, that is, we do not allow points on the separator. Thus (since the imprecision regions are open) a certain separator is a separator such that for any choice of points from the imprecision

[2] LIDAR stands for LIght Detection and Ranging, a popular remote-sensing technology.