Separability with Outliers

Sariel Har-Peled* Vladlen Koltun†

September 12, 2005

Abstract

We develop exact and approximate algorithms for computing optimal separators and measuring the extent to which two point sets in d-dimensional space are separated, with respect to different classes of separators and various extent measures. This class of geometric problems generalizes two widely studied problem families, namely separability and the computation of statistical estimators.

1 Introduction

Consider a family $\mathcal{F}$ of surfaces in $\mathbb{R}^d$, for $d \geq 2$, called *separators*, such that every $f \in \mathcal{F}$ partitions $\mathbb{R}^d$ into at least two connected components, some of which are labelled *inside* f, while the rest are said to be *outside* f. For instance, if $\mathcal{F}$ is a set of hyperplanes, the open halfspace lexicographically below $f \in \mathcal{F}$ is said to be inside f, while the other halfspace is outside. As another example, if $\mathcal{F}$ is a set of spheres, points inside $f \in \mathcal{F}$ are inside while the other connected component of $\mathbb{R}^d \setminus f$ is outside. Given such a family $\mathcal{F}$ and two sets of points $\mathcal{R}$ and $\mathcal{B}$ in $\mathbb{R}^d$ (said to be *red* and *blue*, respectively), such that $|\mathcal{R}| = |\mathcal{B}| = n$, the *separability* problem asks for finding a separator $f \in \mathcal{F}$, if one exists, such that all the blue (resp., red) points are inside (resp., outside) f.

The study of the separability problem is motivated in part by the fundamental classification problem in machine learning: Given two sets of objects (the *training sets*), construct a *predictor* that will facilitate a rapid classification of a new object as belonging to one of the sets. For the separability problem, the training sets are sets of points, and the computed separator is a good candidate predictor for classifying all other points in the space.

*Department of Computer Science; University of Illinois; 201 N. Goodwin Avenue; Urbana, IL, 61801, USA; sariel@cs.uiuc.edu; http://www.uiuc.edu/~sariel/. Work on this paper has been supported by a NSF CAREER award CCR-0132901.

†Computer Science Department, 353 Serra Mall, Gates 464, Stanford University, Stanford, CA 94305, USA; http://cs.stanford.edu/~vladlen

Numerous statistical techniques have been developed for the classification problem, such as Support Vector Machines [17, 27]. The separability problem has also been studied from the combinatorial point of view, particularly in computational geometry, with the aim of developing algorithms with guaranteed correctness and bounds on worst-case running time. In this context it is customary to assume that the dimension d is constant.

For hyperplane and sphere separators linear-time algorithms are known [32, 34, 36]. Algorithms have been found and hardness results have been developed for a number of types of separators for $d = 2$, such as (boundaries of) strips, wedges, double-wedges [8, 29, 28], convex polygons and simple polygons [22, 24, 33]. Separability for slab, wedge, prism and cone separators in $d = 3$ has received recent attention, alongside other types of separators in 3-space, and exact algorithms with running times ranging from near-linear to near-$O(n^8)$ were presented [3, 30].

Previous works on the separability problem in computational geometry deal with finding a separator that *completely* separates the red and blue points, as described above. However, in practical applications it is likely that the point sets will be *almost*, but not completely separable, in a sense that is made precise below, due to noise, sampling and round-off errors. In such scenario the above algorithms simply report that the point sets are not separable and terminate. In this paper we develop separability algorithms that ensure robustness to inaccuracies in the input by computing the *best* separator (when the point sets are separable, this is simply a complete separator as above).

Specifically, for $p, q \in \mathbb{R}^d$ and $\mathcal{S}, \mathcal{T} \subseteq \mathbb{R}^d$ define $d(p, q) = \| p-q \|$, $d(p, \mathcal{T}) = \min_{q \in \mathcal{T}} \| p - q \|$, and $d(\mathcal{S}, \mathcal{T}) = \min_{p \in \mathcal{S}, q \in \mathcal{T}} \| p - q \|$. Consider the following measures for estimating the extent of a point set $\mathcal{S} \in \mathbb{R}^3$ with respect to a separator $f \in \mathcal{F}$.

- The combinatorial measure: $\mathcal{M}_c(f, \mathcal{S}) = |\, \mathcal{S} \,|$. (This quantity does not depend on f for a fixed $\mathcal{S}$.)
- The $\mathcal{L}_\infty$ measure: $\mathcal{M}_\infty(f, \mathcal{S}) = \max_{p \in \mathcal{S}} d(f, p)$.
- The $\mathcal{L}_1$ measure: $\mathcal{M}_1(f, \mathcal{S}) = \sum_{p \in \mathcal{S}} d(f, p)$.
- The $\mathcal{L}_2$ measure: $\mathcal{M}_2(f, \mathcal{S}) = \sum_{p \in \mathcal{S}} (d(f, p))^2$.

For $\mathcal{R}$, $\mathcal{B}$ and $f \in \mathcal{F}$ as above, let a red (resp., blue) *outlier* be defined as a point of $\mathcal{R}$ (resp., of $\mathcal{B}$) that lies inside (resp., outside) f. Let $\mathcal{R}_f \subseteq \mathcal{R}$ (resp., $\mathcal{B}_f \subseteq \mathcal{B}$) be the set of red (resp., blue) outliers and define $\mathcal{O}_f := \mathcal{R}_f \cup \mathcal{B}_f$. In this paper we develop algorithms that, given $\mathcal{R}$, $\mathcal{B}$ and $\mathcal{F}$, compute a separator $f \in \mathcal{F}$ that minimizes (exactly or approximately) one of the above extent measures of $\mathcal{O}_f$ with respect to f. That is, the separator minimizes the number of outliers, the distance of the farthest outlier from the separator, the sum of the distances of the outliers from the separator, or the sum of the squares of these distances. We call this class of problems *separability with outliers*. We concentrate in this initial study on hyperplane and sphere separators. We present exact algorithms that apply in any $d = O(1)$.

Since exact algorithms quickly become impractical as the dimension increases, we present approximation algorithms for all the studied problems. Most of these are candidates for efficient implementation. Specifically, we show that:

- Let $\mathcal{F}$ be the set of hyperplanes (resp., spheres) in $\mathbb{R}^d$. The separator $f^* \in \mathcal{F}$ that minimizes $\mathcal{M}_c(f, \mathcal{O}_f)$ for $f \in \mathcal{F}$ can be computed in time $O(nk^{d+1} \log k)$ (resp., $O(nk^{d+2} \log k)$), where $k = \mathcal{M}_c(f^*, \mathcal{O}_{f^*})$. For any $\varepsilon > 0$, a separator $f \in \mathcal{F}$, such that $\mathcal{M}_c(f, \mathcal{O}_f) \leq (1+\varepsilon)\mathcal{M}_c(f^*, \mathcal{O}_{f^*})$ can be computed in time $O\big(n(\varepsilon^{-2} \log n)^{d+1}\big)$ (resp., $O\big(n(\varepsilon^{-2} \log n)^{d+2}\big)$).
- Let $\mathcal{F}$ be the set of hyperplanes (resp., spheres) in $\mathbb{R}^d$. The separator $f^* \in \mathcal{F}$ that minimizes $\mathcal{M}_\infty(f, \mathcal{O}_f)$ for $f \in \mathcal{F}$ can be computed in time $O\big(n^{\lceil d/2 \rceil}\big)$ (resp., $O\big(n^{\lfloor d/2 \rfloor + 1}\big)$). For any $\varepsilon > 0$, a separator $f \in \mathcal{F}$, such that $\mathcal{M}_\infty(f, \mathcal{O}_f) \leq (1+\varepsilon)\mathcal{M}_\infty(f^*, \mathcal{O}_{f^*})$ can be computed in time $O\big(n/\varepsilon^{(d-1)/2}\big)$ (resp., $O\big(n/\varepsilon^{(d-1)/2} + 1/\varepsilon^{4d}\big)$).
- Let $\mathcal{F}$ be the set of hyperplanes in $\mathbb{R}^d$. The separator $f^* \in \mathcal{F}$ that minimizes $\mathcal{M}_1(f, \mathcal{O}_f)$ (resp., $\mathcal{M}_2(f, \mathcal{O}_f)$) for $f \in \mathcal{F}$ can be computed in time $O(n^d)$ (resp., $O(n^{d+1})$). For any $\varepsilon > 0$, a separator $f \in \mathcal{F}$, such that $\mathcal{M}_1(f, \mathcal{O}_f) \leq (1+\varepsilon)\mathcal{M}_1(f^*, \mathcal{O}_{f^*})$ (resp., $\mathcal{M}_2(f, \mathcal{O}_f) \leq (1+\varepsilon)\mathcal{M}_2(f^*, \mathcal{O}_{f^*})$) can be computed in time $O\big(n/\varepsilon^{(d-1)/2}\big)$ (resp., $O\big(n/\varepsilon^{d/2}\big)$).
- Let $\mathcal{F}$ be the set of spheres in $\mathbb{R}^d$. For any $\varepsilon > 0$, a separator $f \in \mathcal{F}$, such that $\mathcal{M}_1(f, \mathcal{O}_f) \leq (1+\varepsilon)\mathcal{M}_1(f^*, \mathcal{O}_{f^*})$ (resp., $\mathcal{M}_2(f, \mathcal{O}_f) \leq (1+\varepsilon)\mathcal{M}_2(f^*, \mathcal{O}_{f^*})$) can be computed in time $O\Big(\big(\frac{n}{\varepsilon} \log \frac{n}{\varepsilon}\big)^{d+1}\Big)$.

The running time for most of these algorithms can be improved by specialized techniques when the dimension is small (e.g., $d \leq 3$). We do not describe exact algorithms for finding optimal sphere separators with respect to the $\mathcal{L}_1$ and $\mathcal{L}_2$ measures, since computing such a separator requires analytic evaluation of a function that is a sum of terms, each term being an absolute value of a difference of a square root of a polynomial and a variable. Such analytic evaluation is not known to be possible.

To our knowledge, the only instance of separability with outliers that was studied before is line separability with respect to the combinatorial measure in the plane [13, 23]. The optimal line separator in this setting can be computed in time $O((n + k^2) \log k)$, where k denotes the number of outliers achieved by the optimum.

Motivated by the practical considerations expressed above, a number of previous results in computational geometry provide algorithms that are insensitive to outliers. Problems that were studied in this context include linear programming [13, 23, 31], shape fitting [26], and facility location [15].

Separability with outliers has an interesting connection to the computation of statistical estimators, which in turn is intimately related to shape fitting. Given a point set $\mathcal{S}$ and a family $\mathcal{F}$ of *estimators*, consider the problem of finding the estimator $f \in \mathcal{F}$ that minimizes one of the above extent measures of $\mathcal{S}$ with respect to f. Motivated by applications in statistical analysis and computational metrology, this class of problems has been intensively studied in recent years, specifically for hyperplane, sphere, line, and cylinder estimators [2, 4, 5, 7, 12, 14, 19, 25, 26]. It is also a special case of separability with outliers. Indeed, setting $\mathcal{R} = \mathcal{B} = \mathcal{S}$ yields precisely $\mathcal{O}_f = \mathcal{S}$ for any $f \in \mathcal{F}$, equating the computation of the optimal separator of $\mathcal{R}$ and $\mathcal{B}$ to the computation of the optimal estimator of $\mathcal{S}$. Some of the results obtained in this paper indeed use the algorithmic machinery developed for the computation of statistical estimators, appropriately extending and modifying it to handle separability with outliers.

2 The Combinatorial Measure

Throughout this paper, let $\mathcal{R}, \mathcal{B} \in \mathbb{R}^d$ be two collections of n points each.

2.1 Hyperplane Combinatorial Separators

Theorem 2.1 *Let $\mathcal{F}$ be the set of hyperplanes in $\mathbb{R}^d$. The separator $f^* \in \mathcal{F}$ that minimizes $\mathcal{M}_c(f, \mathcal{O}_f)$ for $f \in \mathcal{F}$ can be computed in time $O(nk_{\text{opt}}^{d+1} \log k_{\text{opt}})$, where $k_{\text{opt}} = \mathcal{M}_c(f^*, \mathcal{O}_{f^*})$.*

Proof: For $f \in \mathcal{F}$, a point $x = (x_1, x_2, \ldots, x_d) \in \mathbb{R}^d$ lies on h, if $a_1x_1 + \ldots + a_dx_d + a_{d+1} = 0$, where $a_1, \ldots, a_{d+1}$ are the coefficients of h. With a slight abuse of notation we denote $a_1x_1 + \ldots + a_dx_d + a_{d+1}$ by $f(x)$. Note that although the parametric space defining h is $d+1$ dimensional, one parameter can be eliminated by enforcing $a_{d+1} = 1$.

In the above parametric space, every point of $\mathcal{R} \cup \mathcal{B}$ defines a linear inequality. A point corresponding to a separator f violates exactly those inequalities that correspond to points in $\mathcal{O}_f$. Given the linear program formed by these inequality constraints, we are looking for a point that minimizes the number of violated constraints. Stated differently, we seek the minimum number k_{opt} of constraints that need to be removed to make the linear program feasible.

Given k, we can decide in time $O(nk^{d+1})$ whether $k_{\text{opt}} \leq k$ [31]. Performing an exponential search with $k = 2^0, 2^1, 2^2, \ldots$ results in a constant factor approximation of k_{opt}. A binary search then locates the exact optimum. ∎

Remark: Using the improved decision procedure of Chan for $d = 2, 3$ [13], the running time in Theorem 2.1 can be improved to $O((n+k_{\text{opt}}^2) \log^2 n)$ for $d = 2$ and to $O(n \log^2 n + k_{\text{opt}}^{11/4} n^{1/4} \operatorname{polylog} n)$ for $d = 3$.

Theorem 2.2 *Let $\mathcal{F}$ be the set of hyperplanes in $\mathbb{R}^d$ and let $f^* \in \mathcal{F}$ be the separator that minimizes $\mathcal{M}_c(f, \mathcal{O}_f)$ for $f \in \mathcal{F}$. For any $\varepsilon > 0$, a separator $f \in \mathcal{F}$, such that $\mathcal{M}_c(f, \mathcal{O}_f) \leq$*

$(1+\varepsilon)\mathcal{M}_c(f^*, \mathcal{O}_{f^*})$ *can be computed in time* $O\big(n(\varepsilon^{-2}\log n)^{d+1}\big)$.

Proof: We use the above reduction to the problem of estimating the number of constraints that need to be removed so that a linear program in d dimensions becomes feasible. A recent result of Aronov and Har-Peled [9] shows that this number can be approximated within a factor of $(1+\varepsilon)$ in time $O\big(n(\varepsilon^{-2}\log n)^{d+1}\big)$. ∎

2.2 Sphere Combinatorial Separators

Theorem 2.3 *Let* $\mathcal{F}$ *be the set of spheres in* $\mathbb{R}^d$*. The separator* $f^* \in \mathcal{F}$ *that minimizes* $\mathcal{M}_c(f, \mathcal{O}_f)$ *for* $f \in \mathcal{F}$ *can be computed in time* $O(nk_{\mathrm{opt}}^{d+2}\log k_{\mathrm{opt}})$*, where* $k_{\mathrm{opt}} = \mathcal{M}_c(f^*, \mathcal{O}_{f^*})$*. For any* $\varepsilon > 0$*, a separator* $f \in \mathcal{F}$*, such that* $\mathcal{M}_c(f, \mathcal{O}_f) \leq (1+\varepsilon)\mathcal{M}_c(f^*, \mathcal{O}_{f^*})$ *can be computed in time* $O\big(n(\varepsilon^{-2}\log n)^{d+2}\big)$.

Proof: Apply the standard "lifting transformation" [20] that maps every ball B in $\mathbb{R}^d$ to a hyperplane B^* in $\mathbb{R}^{d+1}$ and a point p in $\mathbb{R}^d$ to a point p^* on the standard paraboloid in $\mathbb{R}^{d+1}$, such that $p \in B$ (resp., $p \in \partial B$, $p \notin B$) if and only if p^* lies below (resp., on, above) B^*. This reduces the problem to finding the (approximately) optimal hyperplane separator in $\mathbb{R}^{d+1}$ and the results follow from Theorems 2.1 and 2.2. ∎

Remark: Using the improved decision procedure of Chan [13], the running time in Theorem 2.3 can be improved to $O(n\log^2 n + k_{\mathrm{opt}}^{11/4} n^{1/4} \operatorname{polylog} n)$ for $d = 2$.

3 The $\mathcal{L}_\infty$ Measure

3.1 Hyperplane $\mathcal{L}_\infty$-separators

Theorem 3.1 *Let* $\mathcal{F}$ *be the set of hyperplanes in* $\mathbb{R}^d$*,* $d \geq 3$*. The separator* $f^* \in \mathcal{F}$ *that minimizes* $\mathcal{M}_\infty(f, \mathcal{O}_f)$ *for* $f \in \mathcal{F}$ *can be computed in time* $O\big(n^{\lceil d/2 \rceil}\big)$.

Proof: We follow the approach of Chan [12] to computing the width of a point set. For a separator $f \in \mathcal{F}$, parameterize it as $f = (x_f, a_f)$, for $x_f \in \mathbb{R}^d$, $a_f \in \mathbb{R}$, such that $f = \{p \in \mathbb{R}^d \mid p \cdot x_f = a_f\}$. Consider the minimal slab S_f bounded by hyperplanes parallel to f that contains $\mathcal{O}_f$. It can be parameterized as $S_f = \{p \in \mathbb{R}^d \mid b_f \leq p \cdot x_f \leq c_f\}$, such that $b_f, c_f \in \mathbb{R}$ and $2a_f = b_f + c_f$. Since x_f, b_f and c_f are dependent, c_f can be eliminated: $S_f = \{p \in \mathbb{R}^d \mid b_f \leq p \cdot x_f \leq b_f + 1\}$. Observe that $\mathcal{M}_\infty(f, \mathcal{O}_f) = 1/(2\,\|x_f\|)$, where $1/\,\|x_f\|$ is the width of S_f. We compute the separator f^* that minimizes $1/\,\|x_f\|$. Consider the following programming problem:

$$\begin{array}{llll} \min & 1/\,\|x\|, & & \\ \text{s.t.} & \forall p \in \mathcal{B} & x \cdot p \leq b+1, \\ & \forall q \in \mathcal{R} & x \cdot q \geq b, \\ & b \in \mathbb{R}, x \in \mathbb{R}^d & & \end{array}$$

The optimal solution vector (b^*, x^*) provides exactly the parameters b_{f^*} and x_{f^*}, respectively, of the optimal separator. The problem is a non-convex optimization problem within a convex polytope defined by $2n$ linear constraints in dimension $d+1$. It can thus be solved in time $O(n^{\lceil d/2 \rceil})$ by constructing the polytope and explicitly computing the value of the optimization function at each boundary feature [16]. ∎

Remark: The running time in Theorem 3.1 can be improved to $O(n^{3/2+\varepsilon})$ for any $\varepsilon > 0$ when $d = 3$, using the randomized techniques of [6], and is $O(n \log n)$ when $d = 2$. We omit the details.

Theorem 3.2 *Let $\mathcal{F}$ be the set of hyperplanes in $\mathbb{R}^d$ and let $f^* \in \mathcal{F}$ be the separator that minimizes $\mathcal{M}_\infty(f, \mathcal{O}_f)$ for $f \in \mathcal{F}$. For any $\varepsilon > 0$, a separator $f \in \mathcal{F}$, such that $\mathcal{M}_\infty(f, \mathcal{O}_f) \leq (1+\varepsilon)\mathcal{M}_\infty(f^*, \mathcal{O}_{f^*})$ can be computed in time $O(n/\varepsilon^{(d-1)/2})$.*

Proof: We can test in time $O(n)$ whether $\mathcal{B}$ and $\mathcal{R}$ are separable by a hyperplane f, in which case $\mathcal{M}_\infty(f, \mathcal{O}_f) = 0$ [32]. Assume therefore that this is not the case. Let $\text{conv}(\mathcal{S})$ denote the convex hull of a point set $\mathcal{S} \in \mathbb{R}^d$. Assume without loss of generality that $O \in \text{conv}(\mathcal{B}) \cap \text{conv}(\mathcal{R})$, where O denotes the origin. Let the *penetration depth* $\pi(\text{conv}(\mathcal{B}), \text{conv}(\mathcal{R}))$ denote the smallest distance by which $\text{conv}(\mathcal{B})$ can be translated so that $\text{conv}(\mathcal{B})$ and $\text{conv}(\mathcal{R})$ are interior-disjoint [6]. Denote the optimal translation vector by $t \in \mathbb{R}^d$. It is easy to see that there is a unique hyperplane H_r, that separates the interiors of $\text{conv}(\mathcal{B}+t)$ and $\text{conv}(\mathcal{R})$. Consider also $H_b = H_r - t$ and $H_m = H_r - \frac{t}{2}$. (H_m is thus the mid-hyperplane between H_r and H_b.) Observe that H_m is the optimal separator f^*. We concentrate on computing the translation t that approximately realizes the penetration depth $\pi(\text{conv}(\mathcal{B}), \text{conv}(\mathcal{R}))$, i.e., if t^* denotes the optimal translation, $\|t\| \leq (1+\varepsilon)\ \|t^*\|$ and $\text{conv}(\mathcal{B}+t) \cap \text{conv}(\mathcal{R}) = \emptyset$. Given t we can find the corresponding separator in time $O(n)$. We first observe that given a *direction* $\delta \in \mathbb{S}^{d-1}$, we can compute the shortest separating translation t_δ in direction δ in $O(n)$ time. Indeed, assume without loss of generality that $\delta = (0, \ldots, 0, 1)$ is the positive vertical direction. Consider the linear program

$$\begin{array}{llll} \min & e, & & \\ \text{s.t.} & \forall b \in \mathcal{B} & (b + e\delta) \cdot f \geq g, \\ & \forall r \in \mathcal{R} & r \cdot f \leq g, \\ & e, g \in \mathbb{R}, f \in \mathbb{R}^d & & \end{array}$$

The optimum in this program is the value of t_δ and can be found in time $O(n)$ [32]. Observe that if $\gamma, \delta \in \mathbb{S}^{d-1}$ are two directions at angle at most $\sqrt{\varepsilon}$ from each other, then $(1-\varepsilon)t_\delta \leq$

$t_\gamma \leq (1+\varepsilon)t_\delta$ [6]. This implies that if $\Delta \subseteq \mathbb{S}^{d-1}$ is a $\sqrt{\varepsilon}$-net on $\mathbb{S}^{d-1}$ and t^* denotes the optimal translation, $\min_{\delta\in\Delta} \| t_\delta \| \leq (1+\varepsilon) \| t^* \|$. There exists such a $\sqrt{\varepsilon}$-net Δ of size $O\big(1/\varepsilon^{(d-1)/2}\big)$ [18]. Executing the above algorithm for every $\delta \in \Delta$ implies the theorem. ■

3.2 Sphere $\mathcal{L}_\infty$-separators

The algorithms in this section follow the approach of Chan [12] to computing the minimum-width spherical shell.

Theorem 3.3 *Let $\mathcal{F}$ be the set of spheres in $\mathbb{R}^d$. The separator $f^* \in \mathcal{F}$ that minimizes $\mathcal{M}_\infty(f, \mathcal{O}_f)$ for $f \in \mathcal{F}$ can be computed in time $O\big(n^{\lfloor d/2 \rfloor + 1}\big)$.*

Proof: For a separator $f \in \mathcal{F}$, denote its center by $x_f \in \mathbb{R}^d$ and its radius by $r_f \in \mathbb{R}$. Consider the minimum-width spherical shell enclosed by two spheres centered at x_f that contains $\mathcal{O}_f$. Denote its inner and outer radii by y_f and z_f, respectively. Note that $\mathcal{M}_\infty(f, \mathcal{O}_f) = (z_f - y_f)/2$. We compute the separator f^* that minimizes $z_{f^*} - y_{f^*}$. Consider the following programming problem:

$$\begin{array}{llr} \min & z - y, & \\ \text{s.t.} & \forall b \in \mathcal{B} & \| b - x \| \leq z, \\ & \forall r \in \mathcal{R} & \| r - x \| \geq y, \\ & y, z \in \mathbb{R}, x \in \mathbb{R}^d & \end{array}$$

The optimal solution vector (y^*, z^*, x^*) provides the radius $(z_{f^*} + y_{f^*})/2$ and center x_{f^*} of the optimal separator. By a change of variables, the problem is easily seen to be a non-convex optimization problem within a convex polytope defined by $2n$ linear constraints in dimension $d+2$. It can be solved in time $O\big(n^{\lfloor d/2 \rfloor + 1}\big)$ by constructing the polytope and explicitly computing the value of the optimization function at each boundary feature [16]. ■

Theorem 3.4 *Let $\mathcal{F}$ be the set of spheres in $\mathbb{R}^d$ and let $f^* \in \mathcal{F}$ be the separator that minimizes $\mathcal{M}_\infty(f, \mathcal{O}_f)$ for $f \in \mathcal{F}$. For any $\varepsilon > 0$, a separator $f \in \mathcal{F}$, such that $\mathcal{M}_\infty(f, \mathcal{O}_f) \leq (1+\varepsilon)\mathcal{M}_\infty(f^*, \mathcal{O}_{f^*})$ can be computed in time $O\big(n/\varepsilon^{(d-1)/2} + 1/\varepsilon^{4d}\big)$.*

Proof: Omitted due to space limitations. ■

Remark: In Theorems 3.4 and 3.2, more advanced arguments can be used to decouple the $1/\varepsilon$ factor in the bound from n, yielding a running time of the form $O\big(n + 1/\varepsilon^{O(d)}\big)$, which can be refined for small d. We omit the details.

4 The $\mathcal{L}_1$ and $\mathcal{L}_2$ Measures

4.1 Hyperplane $\mathcal{L}_1$-separators

Theorem 4.1 *Let $\mathcal{F}$ be the set of hyperplanes in $\mathbb{R}^d$. The separator $f^* \in \mathcal{F}$ that minimizes $\mathcal{M}_1(f, \mathcal{O}_f)$ for $f \in \mathcal{F}$ can be computed in time $O(n^d)$.*

Proof: Observe that the optimal separator divides $R \cup B$ into two equal sets. Our algorithm considers all such partitions of $R \cup B$. After the standard duality transformation, these partitions correspond precisely to the faces of the median level of the arrangement $\mathcal{A}$ of the hyperplanes dual to the points $R \cup B$. In fact, it can be shown that the optimal separator corresponds to a *vertex* of this level. We construct $\mathcal{A}$ and traverse the vertices of the median level using breadth-first search on the 1-skeleton of the arrangement [21]. Throughout the traversal we maintain the value of $\mathcal{M}_1(f, \mathcal{O}_f)$. Since at each step only $O(1)$ points change their relation to the separator, the value of the measure can be maintained in $O(1)$ time per step. Upon the completion of the traversal we output the minimal value of $\mathcal{M}_1(f, \mathcal{O}_f)$ encountered during the process. We omit the easy details. ∎

Remark: The running time in Theorem 4.1 can be improved to $O(n^{4/3} \log n)$ when $d = 2$ using the dynamic convex hull algorithm of [10] to construct the median level in the arrangement. When $d = 3$, the running time is $O(n^{5/2+\varepsilon})$ for any $\varepsilon > 0$ using combinatorial and algorithmic results concerning median levels in three dimensions [11, 35]. Unfortunately there are no substantially sub-$O(n^d)$ bounds on the complexity of the median level of an arrangement in $\mathbb{R}^d$ for $d \geq 4$ (see [1] for the state of the art). Obtaining such a bound is an interesting open problem.

The following is an easy consequence of the techniques of Yamamoto et al [37].

Theorem 4.2 *Let $\mathcal{F}$ be the set of hyperplanes in $\mathbb{R}^d$ and let $f^* \in \mathcal{F}$ be the separator that minimizes $\mathcal{M}_1(f, \mathcal{O}_f)$ for $f \in \mathcal{F}$. For any $\varepsilon > 0$, a separator $f \in \mathcal{F}$, such that $\mathcal{M}_1(f, \mathcal{O}_f) \leq (1+\varepsilon)\mathcal{M}_1(f^*, \mathcal{O}_{f^*})$ can be computed in time $O(n/\varepsilon^{(d-1)/2})$.*

Proof: Parameterize $f \in \mathcal{F}$ by $x_f \in \mathbb{R}^d$, such that $f = \{p \in \mathbb{R}^d \mid p \cdot x_f = 1\}$. Let $\alpha \in \mathbb{S}^{d-1}$ be the vertical direction. For $x = (x_1, \ldots, x_d) \in \mathbb{R}^d$, define the vertical distance $d^\alpha(f, p) = \mid p \cdot x_f - 1 \mid$. Define the vertical $\mathcal{L}_1$-measure of f with respect to $S \subseteq \mathbb{R}^d$ as $\mathcal{M}_1^\alpha(f, \mathcal{S}) = \sum_{p \in \mathcal{S}} d^\alpha(f, p)$.

Lemma 4.3 *The separator $f^\alpha \in \mathcal{F}$ that minimizes $\mathcal{M}_1^\alpha(f, \mathcal{O}_f)$ for $f \in \mathcal{F}$ can be computed in time $O(n)$.*

Proof: Consider a point $r \in \mathcal{R}$. Observe that its contribution to $\mathcal{M}_1^\alpha(f, \mathcal{O}_f)$ is $C_r(x_f) = \max(1 - r \cdot x_f, 0)$. On the other hand, the contribution of $b \in \mathcal{B}$ to $\mathcal{M}_1^\alpha(f, \mathcal{O}_f)$ is $C_b(x_f) = \max(b \cdot x_f - 1, 0)$. We seek to minimize the function $\sum_{p \in \mathcal{B} \cup \mathcal{R}} C_p(x_f)$ over the space of separators $x_f \in \mathbb{R}^d$. We are thus searching for the minimum of the sum of piecewise linear convex functions over $\mathbb{R}^d$. This can be computed in time $O(n)$ using the pruning technique of Megiddo [32, 37]. ■

Note that the algorithm of Lemma 4.3 can compute the separator that minimizes the directional distance in any direction $\gamma \in \mathbb{S}^{d-1}$ by first applying a linear transformation to the space that transforms γ to be vertical. Observe that if $\gamma, \delta \in \mathbb{S}^{d-1}$ are two directions at angle at most $\sqrt{\varepsilon}$ from each other, $(1-\varepsilon)\mathcal{M}_1^{\gamma}(f^{\gamma}, \mathcal{O}_{f^{\gamma}}) \leq \mathcal{M}_1^{\delta}(f^{\delta}, \mathcal{O}_{f^{\delta}}) \leq (1+\varepsilon)\mathcal{M}_1^{\gamma}(f^{\gamma}, \mathcal{O}_{f^{\gamma}})$. (We omit the simple proof.) Note also that if $\beta \in \mathbb{S}^{d-1}$ denotes the direction normal to f^*, then $\mathcal{M}_1^{\beta}(f^{\beta}, \mathcal{O}_{f^{\beta}}) = \mathcal{M}_1(f^*, \mathcal{O}_{f^*})$. This implies that if $\Delta \subseteq \mathbb{S}^{d-1}$ is a $\sqrt{\varepsilon}$-net on $\mathbb{S}^{d-1}$, $\min_{\delta \in \Delta} \mathcal{M}_1^{\delta}(f^{\delta}, \mathcal{O}_{f^{\delta}}) \leq (1+\varepsilon)\mathcal{M}_1(f^*, \mathcal{O}_{f^*})$. There exists such a $\sqrt{\varepsilon}$-net Δ of size $O\big(1/\varepsilon^{(d-1)/2}\big)$ [18]. Executing the algorithm of Lemma 4.3 for every $\delta \in \Delta$ implies the theorem. ■

Remark: The algorithm of Theorem 4.2 immediately yields a linear (in n) ε-approximation algorithm for the computation of an $\mathcal{L}_1$ linear estimator of a point set in $\mathbb{R}^d$. To our knowledge this is the first such algorithm for this basic problem in statistics (see [37] for an exact algorithm in the plane).

4.2 Hyperplane $\mathcal{L}_2$-separators

Theorem 4.4 *Let $\mathcal{F}$ be the set of hyperplanes in $\mathbb{R}^d$. The separator $f^* \in \mathcal{F}$ that minimizes $\mathcal{M}_2(f, \mathcal{O}_f)$ for $f \in \mathcal{F}$ can be computed in time $O(n^{d+1})$.*

Proof: Consider the arrangement $\mathcal{A}(\mathcal{B}^* \cup \mathcal{R}^*)$ of the set of hyperplanes dual to the points $\mathcal{B} \cup \mathcal{R}$. Every point in a specific face C of $\mathcal{A}(\mathcal{B}^* \cup \mathcal{R}^*)$ corresponds to a hyperplane f that induces a fixed partition of $\mathcal{B} \cup \mathcal{R}$ that is the same for all such f. Thus the set of outliers O_f is fixed and the measure $\mathcal{M}_2(f, \mathcal{O}_f)$ is given by a function common to all such f. This function is a sum of $|O_f|$ quadratic functions and is thus a quadratic function. Furthermore, since the partitions induced by neighboring faces of $\mathcal{A}(\mathcal{B}^* \cup \mathcal{R}^*)$ differ by at most a constant number of points, the above measure function for a specific cell can be computed in constant time given the measure function for a neighboring cell. This leads to the following algorithm.

Construct the arrangement $\mathcal{A}(\mathcal{B}^* \cup \mathcal{R}^*)$ and traverse all its faces of all dimensions. At each step of the traversal, update in constant time the measure function and compute the point that minimizes the function. Determine in linear time [32] whether this optimum point lies in the interior of the current face. If so, compare the value at the optimum to the minimal value so far and update the minimum if necessary. Otherwise the optimal point cannot lie in the current arrangement face. When the traversal is completed the algorithm outputs the minimal value and the hyperplane that corresponds to the dual point that achieves the minimal value. The running time follows since we traverse $O(n^d)$ arrangement faces and spend $O(n)$ time in each. ■

Theorem 4.5 *Let $\mathcal{F}$ be the set of hyperplanes in $\mathbb{R}^d$ and let $f^* \in \mathcal{F}$ be the separator that minimizes $\mathcal{M}_2(f, \mathcal{O}_f)$ for $f \in \mathcal{F}$. For any $\varepsilon > 0$, a separator $f \in \mathcal{F}$, such that $\mathcal{M}_2(f, \mathcal{O}_f) \leq$*

$(1+\varepsilon)\mathcal{M}_2(f^*, \mathcal{O}_{f^*})$ *can be computed in time* $O(n/\varepsilon^{d/2})$.

Proof: Here, we want to find the hyperplane that approximately minimizes the sum of square distances of the misclassified points. This can be achieved by a direct adaptation of the techniques of Theorem 4.2. We omit the details. ■

4.3 Sphere $\mathcal{L}_1$- and $\mathcal{L}_2$-separators

Theorem 4.6 *Let* $\mathcal{F}$ *be the set of spheres in* $\mathbb{R}^d$ *and let* $f^* \in \mathcal{F}$ *be the separator that minimizes* $\mathcal{M}_1(f, \mathcal{O}_f)$ *for* $f \in \mathcal{F}$. *For any* $\varepsilon > 0$, *a separator* $f \in \mathcal{F}$, *such that* $\mathcal{M}_1(f, \mathcal{O}_f) \leq (1+\varepsilon)\mathcal{M}_1(f^*, \mathcal{O}_{f^*})$ *can be computed in time* $O\Big(\big(\frac{n}{\varepsilon}\log\frac{n}{\varepsilon}\big)^{d+1}\Big)$. *The same result can be achieved for the* $\mathcal{M}_2$ *measure.*

Proof: Throughout the proof we primarily discuss the case of the $\mathcal{M}_1$ measure. The algorithm can be naturally adapted to the $\mathcal{M}_2$ measure.

We begin by using Theorem 3.3 to compute an optimal $\mathcal{L}_\infty$ sphere separator f^∞ in time $O\big(n^{\lfloor d/2 \rfloor + 1}\big)$ and define $r = \mathcal{M}_\infty(f^\infty, \mathcal{O}_{f^\infty})$. This provides an n-approximation to the optimal $\mathcal{L}_1$-separator. Indeed, for any sphere s in $\mathbb{R}^d$ there is a point in $\mathcal{O}_s$ at distance at least r from s. On the other hand, all points of $\mathcal{O}_{f^\infty}$ are at distance at most r from f^∞. Thus $r \leq \mathcal{M}_1(f^*, \mathcal{O}_{f^*}) \leq nr$. Define $U = nr$ and note that $\mathcal{M}_1(f^*, \mathcal{O}_{f^*}) \leq U \leq n\mathcal{M}_1(f^*, \mathcal{O}_{f^*})$.

Place around each point of $\mathcal{B} \cup \mathcal{R}$ a ball of radius $r_i = \frac{\varepsilon}{2}(U/n^2)(1+\frac{\varepsilon}{2})^i$, for $i = 0, \ldots, M$, where $M = 4\left\lceil \log_{1+\varepsilon/2}(n^2/\varepsilon) \right\rceil = O(\frac{1}{\varepsilon}\log\frac{n}{\varepsilon})$. Let $\mathcal{D}$ denote the resulting set of $O(nM) = O(\frac{n}{\varepsilon}\log\frac{n}{\varepsilon})$ balls.

Consider a sphere f. Clearly, if f fails to touch any ball of $\mathcal{D}$ around a point $p \in \mathcal{O}_f$, then $d(p, C) \geq r_M \geq Un^2$. In particular, we can ignore such spheres, since they are too expensive to be a good approximation to $\mathcal{M}_1(f^*, \mathcal{O}_{f^*})$. Furthermore, let $r_{\mathcal{D}}(f, p)$ denote the radius of the smallest ball in $\mathcal{D}$ that is centered at p and intersects f. Let $\alpha(f) = \sum_{p \in \mathcal{O}_f} r_{\mathcal{D}}(f, p)$. We have that

$$\begin{aligned} \mathcal{M}_1(f, \mathcal{O}_f) &= \sum_{p \in \mathcal{O}_f} d(f,p) \leq \alpha(f) = \sum_{p \in \mathcal{O}_f} r_{\mathcal{D}}(f,p) \leq \sum_{p \in \mathcal{O}_f} \Big(r_0 + (1+\frac{\varepsilon}{2}) d(f,p) \Big) \\ &= n r_0 + (1+\frac{\varepsilon}{2}) \sum_{p \in \mathcal{O}_f} d(f,p) \leq (1+\varepsilon)\mathcal{M}_1(f, \mathcal{O}_f). \end{aligned}$$

However, the value of $\alpha(f)$ is uniquely defined by the set of balls of $\mathcal{D}$ that the sphere f intersects and the points of $\mathcal{B} \cup \mathcal{R}$ that the ball defined by f contains. We now show how to enumerate all such sets efficiently. We encode a ball B in $\mathbb{R}^d$ by a cone in $\mathbb{R}^{d+1}$ with axis parallel to the x_{d+1}-axis, such that its intersection with the hyperplane $x_{d+1} = 0$ is B. Thus a ball centered at $\mathbf{x} \in \mathbb{R}^d$ with radius r is encoded by a cone with apex $(\mathbf{x}, -r)$, denoted by $\mathrm{cone}(\mathbf{x}, -r)$. A *sphere* S centered at $\mathbf{x}$ with radius r is encoded by the point $(\mathbf{x}, r)$. Consider the set of spheres intersecting B. Clearly, this is the set of points lying above the hyperplane $x_{d+1} = 0$, above cone$(\mathbf{x}, -r)$ and below cone$(\mathbf{x}, r)$. Also, the set of spheres that contain a point $\mathbf{x} \in \mathbb{R}^d$ is the set of points lying above cone$(\mathbf{x}, 0)$. Thus, consider the arrangement of the $2\,|\mathcal{D}|+n$ cones induced in this way by the balls of $\mathcal{D}$ and the points of $\mathcal{B}\cup\mathcal{R}$, together with the hyperplane $x_{d+1} = 0$. Clearly, all the spheres defined by points inside a particular face in this arrangement intersect the same set of balls of $\mathcal{D}$ and enclose the same set of points of $\mathcal{B}\cup\mathcal{R}$. We can thus find an approximate $\mathcal{L}_1$ sphere separator simply by traversing the above arrangement. The complexity of the arrangement is $O((2\,|\mathcal{D}| + n)^{d+1}) = O\left(\left(\frac{n}{\varepsilon}\log\frac{n}{\varepsilon}\right)^{d+1}\right)$ and the traversal can be performed in the same asymptotic time. This implies the theorem. ∎

5 Discussion and Open Problems

This paper is a step towards a better understanding of separability with outliers. A wide range of interesting problems remain. Can algorithms presented in this paper be improved? In particular, are there ε-approximation algorithms whose running time is near-linear in n and polynomial in $1/\varepsilon$ and d?

Can we use the ideas of this paper to handle other types of separators, such as slabs, cylinders, cones, and prisms? We can also consider the family of all convex bodies as separators.

A number of successes in designing approximation algorithms for shape fitting and extent approximation have been achieves by proving that there always exists a small *coreset* that approximately captures the extent of the point set. Can we demonstrate the existence of coresets for separability with outliers (i.e., small sets that witness the extent to which two point sets penetrating each other) or prove that they do not always exist?

It would also be interesting to consider algorithms that maintain (approximately) minimal separators as the point sets change, either as points are inserted and deleted (dynamic setting) or as points move continuously (kinetic setting)? The problem can also be considered in the streaming model of computation.

Acknowledgements

The authors wish to thank Pankaj Agarwal for suggesting the problem of separability with outliers and for initial discussions.

References

[1] P. K. Agarwal, B. Aronov, T. M. Chan, and M. Sharir. On levels in arrangements of lines, segments, planes, and triangles. *Discrete Comput. Geom.*, 19:315–331, 1998.

[2] P. K. Agarwal, B. Aronov, S. Har-Peled, and M. Sharir. Approximation and exact algorithms for minimum-width annuli and shells. *Discrete Comput. Geom.*, 24(4):687–705, 2000.

[3] P. K. Agarwal, B. Aronov, and V. Koltun. Efficient algorithms for bichromatic separability. *ACM Transactions on Algorithms*, 2005. to appear.

[4] P. K. Agarwal, B. Aronov, and M. Sharir. Line traversals of balls and smallest enclosing cylinders in three dimensions. *Discrete Comput. Geom.*, 21:373–388, 1999.

[5] P. K. Agarwal, B. Aronov, and M. Sharir. Exact and approximation algorithms for minimum-width cylindrical shells. *Discrete Comput. Geom.*, 26(3):307–320, 2001.

[6] P. K. Agarwal, L. J. Guibas, S. Har-Peled, A. Rabinovitch, and M. Sharir. Penetration depth of two convex polytopes in 3D. *Nordic J. Comput.*, 7(3):227–240, 2000.

[7] P. K. Agarwal, S. Har-Peled, and K. R. Varadarajan. Approximating extent measures of points. *J. Assoc. Comput. Mach.*, 51:606–635, 2004.

[8] E. Arkin, F. Hurtado, J. Mitchell, C. Seara, and S. Skiena. Some lower bounds on geometric separability problems. In *11th Fall Workshop on Computational Geometry*, 2001.

[9] B. Aronov and S. Har-Peled. On approximating the depth and related problems. In *Proc. 16th ACM-SIAM Sympos. Discrete Algorithms*, 2005.

[10] G. S. Brodal and R. Jacob. Dynamic planar convex hull. In *Proc. 43th Annu. IEEE Sympos. Found. Comput. Sci.*, pages 617–626, 2002.

[11] T. M. Chan. Output-sensitive results on convex hulls, extreme points, and related problems. *Discrete Comput. Geom.*, 16:369–387, 1996.

[12] T. M. Chan. Approximating the diameter, width, smallest enclosing cylinder and minimum-width annulus. *Internat. J. Comput. Geom. Appl.*, 12(2):67–85, 2002.

[13] T. M. Chan. Low-dimensional linear programming with violations. In *Proc. 43th Annu. IEEE Sympos. Found. Comput. Sci.*, pages 570–579, 2002.

[14] T. M. Chan. Faster core-set constructions and data stream algorithms in fixed dimensions. In *Proc. 20th Annu. ACM Sympos. Comput. Geom.*, pages 152–159, 2004.

[15] M. Charikar, S. Khuller, D. M. Mount, and G. Narasimhan. Algorithms for facility location problems with outliers. In *Proc. 12th ACM-SIAM Sympos. Discrete Algorithms*, pages 642–651, 2001.

[16] B. Chazelle. An optimal convex hull algorithm in any fixed dimension. *Discrete Comput. Geom.*, 10:377–409, 1993.

[17] N. Cristianini and J. Shaw-Taylor. *Support Vector Machines*. Cambridge University Press, 2000.

[18] R. M. Dudley. Metric entropy of some classes of sets with differentiable boundaries. *J. Approx. Theory*, 10:227–236, 1974.

[19] C. A. Duncan, M. T. Goodrich, and E. A. Ramos. Efficient approximation and optimization algorithms for computational metrology. In *Proc. 8th ACM-SIAM Sympos. Discrete Algorithms*, pages 121–130, 1997.

[20] H. Edelsbrunner. *Algorithms in Combinatorial Geometry*, volume 10 of *EATCS Monographs on Theoretical Computer Science*. Springer-Verlag, Heidelberg, West Germany, 1987.

[21] H. Edelsbrunner, J. O'Rourke, and R. Seidel. Constructing arrangements of lines and hyperplanes with applications. *SIAM J. Comput.*, 15:341–363, 1986.

[22] H. Edelsbrunner and F. P. Preparata. Minimum polygonal separation. *Inform. Comput.*, 77:218–232, 1988.

[23] H. Everett, J.-M. Robert, and M. van Kreveld. An optimal algorithm for the ($\leq k$)-levels, with applications to separation and transversal problems. *Internat. J. Comput. Geom. Appl.*, 6:247–261, 1996.

[24] S. P. Fekete. On the complexity of min-link red-blue separation. Manuscript, Department of Applied Mathematics, SUNY Stony Brook, NY, 1992.

[25] S. Har-Peled and K. Varadarajan. High-dimensional shape fitting in linear time. In *Proc. 19th Annu. ACM Sympos. Comput. Geom.*, pages 39–47, 2003.

[26] S. Har-Peled and Y. Wang. Shape fitting with outliers. *SIAM J. Comput.*, 33(2):269–285, 2004.

[27] T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning.* Springer-Verlag, Berlin, Germany, 2001.

[28] F. Hurtado, M. Mora, P. A. Ramos, and C. Seara. Two problems on separability with lines and polygonals. In *Proc. 15th European Workshop on Computational Geometry*, pages 33–35, 1999.

[29] F. Hurtado, M. Noy, P. A. Ramos, and C. Seara. Separating objects in the plane with wedges and strips. *Discrete Appl. Math.*, 109:109–138, 2001.

[30] F. Hurtado, C. Seara, and S. Sethia. Red-blue separability problems in 3d. In *Proc. 3rd Int. Conf. Comput. Sci. and Its Appl.*, pages 766–775, 2003.

[31] J. Matoušek. On geometric optimization with few violated constraints. *Discrete Comput. Geom.*, 14:365–384, 1995.

[32] N. Megiddo. Linear programming in linear time when the dimension is fixed. *J. Assoc. Comput. Mach.*, 31:114–127, 1984.

[33] J. S. B. Mitchell. Approximation algorithms for geometric separation problems. Technical report, Department of Applied Mathematics, SUNY Stony Brook, NY, July 1993.

[34] J. O'Rourke, S. R. Kosaraju, and N. Megiddo. Computing circular separability. *Discrete Comput. Geom.*, 1:105–113, 1986.

[35] M. Sharir, S. Smorodinsky, and G. Tardos. An improved bound for k-sets in three dimensions. *Discrete Comput. Geom.*, 26:195–204, 2001.

[36] V. Vapnik. *The Nature of Statistical Learning Theory.* Springer-Verlag, New York, 1996.

[37] P. Yamamoto, K. Kato, K. Imai, and H. Imai. Algorithms for vertical and orthogonal L_1 linear approximation of points. In *Proc. 4th Annu. ACM Sympos. Comput. Geom.*, pages 352–361. ACM Press, 1988.