Shape Fitting with Outliers*

Sariel Har-Peled† Yusu Wang‡

August 27, 2003

Abstract

Given a set $\mathcal{H}$ of n hyperplanes in $\mathbb{R}^d$, we present an algorithm that ε-approximates the extent between the top and bottom k levels of the arrangement of $\mathcal{H}$ in time $O(n+(k/\varepsilon)^c)$, where c is a constant depending on d. The algorithm relies on computing a subset of $\mathcal{H}$ of size $O(k/\varepsilon^{d-1})$, in near linear time, such that the k-level of the arrangement of the subset approximates that of the original arrangement. Using this algorithm, we propose efficient approximation algorithms for shape fitting with outliers for various shapes. This is the first algorithms to handle outliers efficiently for the shape fitting problems considered.

1 Introduction

Shape fitting, a fundamental problem in computational geometry, computer vision, machine learning, data mining, and many other areas, is concerned with finding the best shape fitting a given input. For example, a widely used shape-fitting problem asks for a shape that best fits a set of points P under some "fitting" criterion. Choices of such shapes include points, lines, hyperplanes, spheres, etc. One typical criterion for measuring how well a shape γ fits P, denoted as $\mu(P, \gamma)$, is the maximum distance between a point of P and its nearest point on γ, i.e., $\mu(P, \gamma) = \max_{p \in P} \min_{q \in \gamma} d(p, q)$. One would then like to find the extent measure of P, defined as $\mu(P) = \min_\gamma \mu(P, \gamma)$, where the minimum is taken over a family of shapes. For example, the problem of finding the point (resp. line) that fits P best is the same as finding the minimum radius sphere (resp. cylinder) enclosing P, and the problem of finding the hyperplane (resp. sphere, cylinder) that fits P best is the same as finding the smallest width slab (resp. spherical shell, cylindrical shell) containing P.

The exact algorithms for shape fitting are generally expensive, e.g., the best-known algorithms for computing the smallest volume bounding box or tetrahedron containing P in

*The latest version of this paper is available from [HW02]. A preliminary version of this paper appeared in *Proc. 19th Annu. ACM Sympos. Comput. Geom.*, pages 29–38, 2003. S.H. is supported by NSF CAREER award CR-0132901. Y.W is supported by NSF grants ITR–333–1050 and CCR–02–04118.

†Department of Computer Science, DCL 2111; University of Illinois; 1304 West Springfield Ave.; Urbana, IL 61801; USA. Email: sariel@cs.uiuc.edu.

‡Department of Computer Science, Duke University, Durham, NC 27708-0129, U.S.A. Email: wys@cs.duke.edu.

$\mathbb{R}^3$ require $O(n^3)$ time . Consequently, attention has shifted to developing approximation algorithms for computing extent measures. Recently, Agarwal *et al.* [AH01, AHV03, HV01] provided a general technique for approximate shape fitting in low dimensions. Their technique relied on using linearization to reduce the problem into convex shape fitting problem, and then solve it by known convex shape approximation techniques. The most striking features of this technique are its wide applicability to numerous shape fitting problems, and that the resulting running time is $O(n + 1/\varepsilon^c)$, where n is the input size, c is a constant that depends on the problem at hand, and ε is the quality of approximation requested.

However, in the real world, noise in the input is omnipresent and one has to assume that some of the input points are noise, and as such should be ignored. To a certain extent, the problem of handling outliers in computational geometry is still open, see [Cha02, EE94, Mat95b] for relevant results. While those results provide relatively efficient solutions, most of them are restricted to two and three dimensions (where intuitively, the k-level has relatively low complexity), and are restricted in the type of problems that they can solve. For additional results about outliers in general, see [CKMN01, DV01].

In this paper, we present approximation algorithms that can handle outliers efficiently in two and higher dimensions for a broad collection of shape fitting problems. For example, given a set P of n points in $\mathbb{R}^d$, we can find approximately the smallest cylinder that contains at least $n - k$ points of P in $O(n + (k/\varepsilon)^c)$ time, where c is a constant depending only on d. [1]. Note that for a fixed ε, and a moderately small k, this is a linear time algorithm. Our algorithm can solve all the shape fitting problems considered in [AH01, AHV03, HV01] efficiently under the additional constraint of k outliers. For most of those problems, this is the first efficient approximation algorithm to handle outliers.

Interestingly enough, our algorithm relies on computing a *coreset* to the given point set in *linear* time. Namely, we compute a subset $\mathfrak{C}$ of the points of cardinality $k/\varepsilon^{O(1)}$, such that instead of solving the problem on the original point set, one can solve it on $\mathfrak{C}$. In particular, in some cases, we just plug in the (exact) algorithms of [Cha02, DLSS95, EE94, Mat95b] on $\mathfrak{C}$, and get an efficient approximation algorithm.

To facilitate this, we investigate a related question, which is interesting in its own right: Given a set $\mathcal{H}$ of n hyperplanes in $\mathbb{R}^d$ and a subset $X \subseteq \mathbb{R}^{d-1}$, how to compute efficiently the shortest vertical segment, with a vertical projection in X, that stabs (at least) $n - k$ of the hyperplanes of $\mathcal{H}$. In approximate form, one would like to find a vertical segment, with a vertical projection in X, that stabs at least $n - k$ of the hyperplanes of $\mathcal{H}$ and is of length at most a factor $(1 + \varepsilon)$ of that of the optimal segment. (Note that linear programming with violations [Mat95b] can be applied directly when $X = \mathbb{R}^{d-1}$ or X is a convex polytope, as the problem can then be formulated as a LP-type problem. For our applications, however, X arises from our linearization technique, and is not necessarily a convex set. In particular, the set X is a semi-algebraic set of constant complexity) The direct approach for solving this problem, is by computing the bottom/top k levels of $\mathcal{A}(\mathcal{H})$, and enumerating by brute force all possible such vertical segments (with vertical projection in X), where $\mathcal{A}(\mathcal{H})$ denote the arrangement of hyperplanes of $\mathcal{H}$. The approach implemented in this fashion is doomed, as it leads to an inefficient algorithm with running time (roughly) of $O(n^d k^d)$. At the same time,

[1]In all our discussions, we consider the dimension to be a small constant, and as such the $O(\cdot)$ notation hides constants that depends solely on d.

since the k-level of $\mathcal{A}(\mathcal{H})$ is not necessarily convex for $k > 1$, we can no longer take advantage of known convex shape approximation techniques as was done in [AH01, AHV03, HV01].

Instead, we study the question of approximating the first/last k levels directly. Note, that the *combinatorial* approximation of k levels is relatively well understood, see [Mat95b, AES99, GG02], and can be performed by random sampling or cuttings. (In the combinatorial settings we are interested in finding a curve γ that lies close to the k-level under the crossing metric. Namely, any vertical segment connecting a point on γ to the k-level crosses at most εn lines.) However, our notion of approximation is stronger as it combines both the geometry and the combinatorial structure of the levels. In particular, we show that there is a coreset for this problem; namely, one can compute a subset of $O(k/\varepsilon^{d-1})$ hyperplanes, such that the rth level of the coreset ε-approximates the r-level of the arrangement $\mathcal{A}(\mathcal{H})$, for $1 \leq r \leq k$. Here the approximation is the euclidean distance between the level and the approximate level, compared with the length of the shortest vertical segment that connects the k-level with the $n-k$ level of the arrangement $\mathcal{A}(\mathcal{H})$. This can be done in $O(n+k/\varepsilon^{d-1})$ time. (All algorithms in this paper are randomized and their results are correct with high probability.)

We can apply the above results to a large number of shape fitting problems, including those considered in [AH01, AHV03, HV01]. To name a few, we can approximate the following measures with k outliers: diameter, width, projection width, minimum enclosing ball, minimum-width annulus, minimum-volume spherical shell, minimum-width cylindrical shell, and also those measures for moving points. Note, that most of those optimization problems fall outside the paradigm of linear programming, and as such can not be solved using linear programming with violations. We can further extend our technique to handle insertions and deletions with poly-logarithmic time per update.

In Section 3, we present an algorithm to compute a small coreset for a set of hyperplanes in $\mathbb{R}^d$. We extend the results for a set of polynomials or their roots in Section 4. In Section 5, we present our results for various shape fitting problems with outliers. Finally, we conclude in Section 6.

2 Preliminaries

Throughout the paper, we refer to the x_d-parallel direction in $\mathbb{R}^d$ as *vertical*. Given a point $x = (x_1, \ldots, x_{d-1})$ in $\mathbb{R}^{d-1}$, let (x, x_d) denote the point $(x_1, \ldots, x_{d-1}, x_d)$ in $\mathbb{R}^d$. Each point $x \in \mathbb{R}^d$ is also a vector in $\mathbb{R}^d$. Given a geometric object A, $A + x$ represents the object obtained by translating each point in A by x.

A *surface* is a subset of $\mathbb{R}^d$ that intersects any vertical line in a single point. Let A and B be either a point, a hyperplane, or a surface in $\mathbb{R}^d$. We say that A lie *above* (resp. *below*) B, denoted by $A \succeq B$ (resp. $A \preceq B$), if for any vertical line ℓ intersecting both A and B, we have that $x_d \geq y_d$ (resp. $x_d \leq y_d$), where $(x_1, \ldots, x_{d-1}, x_d) = A \cap \ell$ and $(x_1, \ldots, x_{d-1}, y_d) = B \cap \ell$.

Definition 2.1 For a set of n hyperplanes $\mathcal{H}$ in $\mathbb{R}^d$, the *level* of a point $x \in \mathbb{R}^d$ in the arrangement $\mathcal{A}(\mathcal{H})$ is the number of hyperplanes of $\mathcal{H}$ lying vertically below x. For $k = 0, \ldots, n-1$, let $\mathbf{L}_{\mathcal{H},k}$ represent the surface which is closure of all points on the hyperplanes of $\mathcal{H}$ whose level is k. We define the *top k-level* of $\mathcal{H}$ to be $\mathbf{U}_{\mathcal{H},k} = \mathbf{L}_{\mathcal{H},n-k-1}$, for $k = 0, \ldots, n-1$.

Let $\mathbf{L}_{\mathcal{H},\leq k}$ and $\mathbf{U}_{\mathcal{H},\leq k}$ denote $\cup_{i=0}^{k}\mathbf{L}_{\mathcal{H},i}$ and $\cup_{i=0}^{k}\mathbf{U}_{\mathcal{H},i}$, respectively. Note that both $\mathbf{L}_{\mathcal{H},k}$ and $\mathbf{L}_{\mathcal{H},\leq k}$ are subsets of the arrangement of $\mathcal{H}$. For $x \in \mathbb{R}^{d-1}$, we slightly abuse notations and define $\mathbf{L}_{\mathcal{H},k}(x)$ to be the value x_d such that $(x, x_d) \in \mathbf{L}_{\mathcal{H},k}$.

Definition 2.2 The (k, r)-*extent* $\mathcal{H}|_r^k : \mathbb{R}^{d-1} \to \mathbb{R}$ is defined as the vertical distance between the r-level and the top k-level of $\mathcal{A}(\mathcal{H})$, i.e., for any $x \in \mathbb{R}^{d-1}$,

$$\mathcal{H}|_r^k(x) = \mathbf{U}_{\mathcal{H},k}(x) - \mathbf{L}_{\mathcal{H},r}(x).$$

The k-*extent of* $\mathcal{H}$ is the (k, k)-extent of $\mathcal{H}$, and is denoted by $\mathcal{E}_{\mathcal{H},k} = \mathcal{H}|_k^k$.

Definition 2.3 Given a parameter $\varepsilon > 0$, a function $\mathcal{E}_\varepsilon : \mathbb{R}^{d-1} \to \mathbb{R}$ is an ε-*approximation* to $\mathcal{H}|_r^k(\cdot)$ if, for any $x \in \mathbb{R}^{d-1}$, we have: $(1 - \varepsilon)\mathcal{H}|_r^k(x) \leq \mathcal{E}_\varepsilon(x) \leq \mathcal{H}|_r^k(x)$.

Our approximation relies on the idea that one can find a small subset of hyperplanes, which are "crucial" as far as the top/bottom k levels are concerned.

Definition 2.4 For a point $x \in \mathbb{R}^{d-1}$, a subset of hyperplanes $\mathcal{H}' \subseteq \mathcal{H}$ in $\mathbb{R}^d$ is a (k, ε, δ)-*coreset at* x if the following holds: for any $r \leq k$, we have

$$\mathbf{L}_{\mathcal{H},r}(x) \leq \mathbf{L}_{\mathcal{H}',r}(x) \leq \mathbf{L}_{\mathcal{H},r}(x) + \varepsilon\mathcal{E}_{\mathcal{H},k}(x) + \delta,$$

and

$$\mathbf{U}_{\mathcal{H},r}(x) - \varepsilon\mathcal{E}_{\mathcal{H},k}(x) - \delta \leq \mathbf{U}_{\mathcal{H}',r}(x) \leq \mathbf{U}_{\mathcal{H},r}(x).$$

If $\mathcal{H}'$ is a (k, ε, δ)-coreset at all points of $\mathbb{R}^{d-1}$, then it is (k, ε, δ)-*coreset* for $\mathcal{H}$.

We omit δ from the above notation when $\delta = 0$. One of the main contributions of this paper is showing the existence of (k, ε)-coresets of size $O(k/\varepsilon^{d-1})$ for a set of hyperplanes in $\mathbb{R}^d$ for any given $k, \varepsilon > 0$. Note that if $\mathcal{H}'$ is a $(k, \varepsilon/2)$-coreset for $\mathcal{H}$, then the (r, t)-extent of $\mathcal{H}'$ ε-approximates that of $\mathcal{H}$ for any $r, t \leq k$. Therefore, in order to compute an ε-approximation of $\mathcal{E}_{\mathcal{H},k}$, one just needs to compute $\mathbf{L}_{\mathcal{H}',k}$ and $\mathbf{U}_{\mathcal{H}',k}$, and the resulting $\mathcal{E}_{\mathcal{H}',k}$ is the required approximation.

3 (k, ε)-Coresets

In this section, given a set $\mathcal{H} = \{h_1, \ldots, h_n\}$ of hyperplanes in $\mathbb{R}^d$, and parameters $\varepsilon > 0$ and $1 \leq k \leq n$, we show how to compute a small (k, ε)-coreset for $\mathcal{H}$. Let $\Delta^{opt}(\mathcal{H}, k)$ denote the shortest vertical distance between the k-level and the top k-level of $\mathcal{A}(\mathcal{H})$, i.e., $\Delta^{opt}(\mathcal{H}, k) = \min_{x \in \mathbb{R}^{d-1}} \mathcal{E}_{\mathcal{H},k}(x)$. The algorithm first computes a set of $O(1)$ vertical segments, the length of each one is at most $\Delta^{opt}(\mathcal{H}, k)$, such that, with high probability, all but $O(k)$ hyperplanes of $\mathcal{H}$ intersects at least one of those segments. Next, we subdivide $\mathcal{H}$ into $O(1/\varepsilon)$ disjoint subsets, and show that the union of the $(k, \varepsilon, \varepsilon\Delta^{opt}(\mathcal{H}, k))$-coreset of each subset form a (k, ε)-coreset for $\mathcal{H}$. Furthermore, we show that a coreset of size $O(k/\varepsilon^{d-2})$ can be computed efficiently for each subset, resulting in a coreset of size $O(k/\varepsilon^{d-1})$ for $\mathcal{H}$.

3.1 Short segments

Lemma 3.1 *Let $\mathcal{J}$ be a set of n hyperplanes in $\mathbb{R}^d$. In $O(n)$ time, one can compute a set S of $O(1)$ vertical segments, such that with high probability: (i) all the segments of S are no longer than $\Delta^{opt}(\mathcal{J}, k)$, and (ii) $|\mathcal{J}_0| = O(k)$, where $\mathcal{J}_0$ denote the set of all hyperplanes of $\mathcal{J}$* not *stabbed by the segments of S.*

Proof: The algorithm is iterative. Let $\mathcal{Q}_1 = \mathcal{J}$, $S_1 = \emptyset$. For $i > 1$, let S_i be the set of vertical segments computed in the beginning of the ith iteration. Let $\mathcal{Q}_i = \mathcal{J} \setminus S_i$, and $m_i = |\mathcal{Q}_i|$, where $\mathcal{J} \setminus S_i$ denote the set of all hyperplanes of $\mathcal{J}$ not intersecting any segment of S_i. That is, $\mathcal{Q}_i$ is the set of hyperplanes not yet handled by the algorithm.

If $m_i = O(k)$ we are done, and set $\mathcal{J}_0$ to be $\mathcal{Q}_i$. Otherwise, if $m_i \le n^{1/(3d)}$, we compute the k-level and the top k-level of the arrangement $\mathcal{A}(\mathcal{Q}_i)$, and the shortest vertical segment between those two levels. Let h_i^* denote this segment, then $|h_i^*| \le \Delta^{opt}(\mathcal{J}, k)$. Set $S_{i+1} = S_i \cup \{h_i^*\}$, and observe that S_{i+1} is the required set, as $|\mathcal{J} \setminus S_{i+1}| = 2k$. Clearly, a naive implementation of the algorithm for this step would take $O(m_i^{2d+1}) = O(n)$ time.

Otherwise, if $m_i > n^{1/(3d)}$, then let $\mathcal{R}_i$ be a random sample from $\mathcal{Q}_i$ of size $O(n^{1/3d} \log n)$. Note that if we consider the range space $\Sigma = (\mathcal{J}, \mathcal{X})$, where $\mathcal{X}$ consists of all subsets of $X \subseteq \mathcal{J}$ where there is a vertical segment s such that $X = s \cap \mathcal{J}$, where $s \cap \mathcal{J}$ denote the hyperplanes of $\mathcal{J}$ that intersect s. The size of $\mathcal{X}$ is bounded by $O(n^{d+1})$. Furthermore, the range space Σ has a bounded Vapnik-Chervonenkis dimension (VC-dimension). Therefore, by the ε-sample theorem [AS00], with hight probability, $\mathcal{R}_i$ is an τ-sample of $\mathcal{Q}_i$, where $\tau = 1/n^{1/6d}$. That is, for any vertical segment s, with high probability, we have

$$\frac{|\mathcal{Q}_i \cap s|}{|\mathcal{Q}_i|} - \tau \le \frac{|\mathcal{R}_i \cap s|}{|\mathcal{R}_i|} \le \frac{|\mathcal{Q}_i \cap s|}{|\mathcal{Q}_i|} + \tau,$$

where $X \cap s$ denote the set of hyperplanes of X that intersects s. Compute the arrangement $\mathcal{A}(\mathcal{R}_i)$, and compute the shortest vertical segment h_i^* between the K_i-level and the top K_i-level of $\mathcal{A}(\mathcal{R}_i)$, where

$$K_i = \left\lceil |\mathcal{R}_i| \left(\frac{k}{|\mathcal{Q}_i|} + \tau \right) \right\rceil.$$

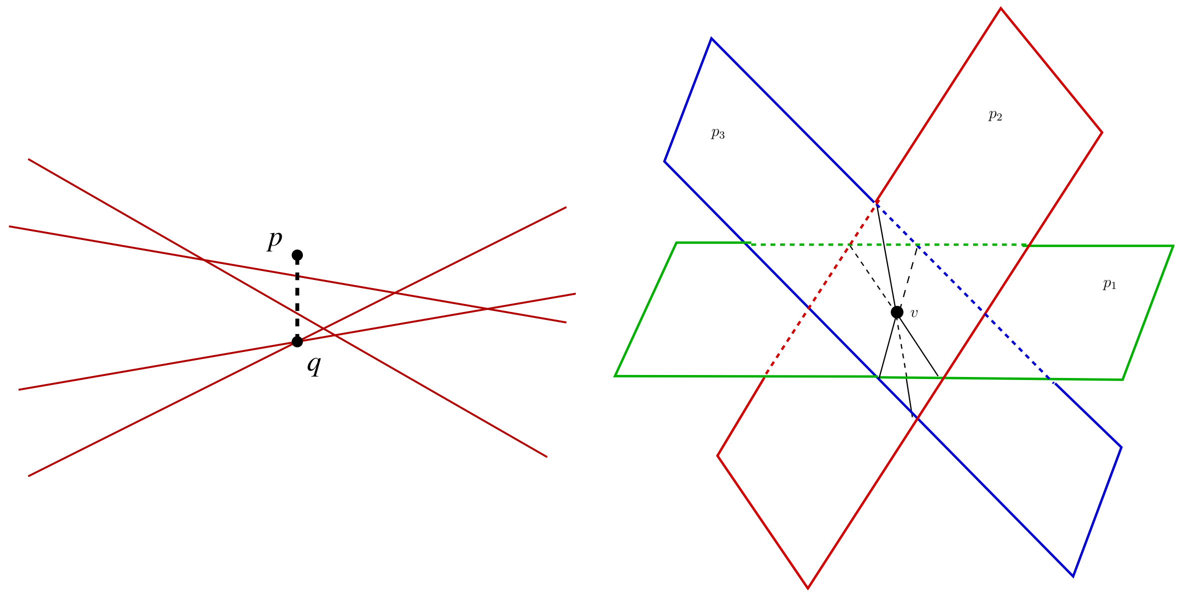
Since $\mathcal{R}_i$ is an τ-sample of $\mathcal{Q}_i$, we have, with high probability, that

$$\mathbf{L}_{\mathcal{Q}_i,k} \preceq \mathbf{L}_{\mathcal{R}_i,K_i} \preceq \mathbf{U}_{\mathcal{R}_i,K_i} \preceq \mathbf{U}_{\mathcal{Q}_i,k}.$$

Thus, with high probability, the segment h_i^* is shorter than $\Delta^{opt}(\mathcal{J}, k)$, and h_i^* intersects at least $m_i(1 - 4\tau) - 2k$ hyperplanes of $\mathcal{Q}_i$. Let $S_{i+1} = S_i \cup \{h_i^*\}$, and let $\mathcal{Q}_{i+1} = \mathcal{Q}_i \setminus \{h_i^*\}$. This step takes $O\Big(|\mathcal{R}_i|^{2d+1}\Big) = O(n)$ time.

We are left with the task of bounding the number of iterations of the algorithm. Clearly, with high probability, $m_{i+1} \le 4\tau m_i + 2k$. Therefore, for $m_i \ge 4k/\tau$, we have $m_{i+1} \le 8\tau m_i \le 8m_i/n^{1/6d}$. On the other hand, if $m_i \le 4k/\tau$ then $m_{i+1} \le 6k$ and the algorithm stops in the next iteration. We conclude that the number of iterations performed by the algorithm is $O\big(\log_{1/\tau} n\big) = O(6d) = O(1)$ with high probability, which implies the lemma. ■

The algorithm of Lemma 3.1 can be derandomized using deterministic construction of ε-samples [Cha00]. This results in a somewhat slower algorithm. Note, that the algorithm of Lemma 3.1 can be modified to output together with each vertical segment of S the corresponding set of hyperplanes of $\mathcal{J}$ that intersect it within the same time bound.



(a) (b)

Figure 1: (a) A δ-sheaf in $\mathbb{R}^2$ with axis pq. (b) A sheaf in $\mathbb{R}^3$ with focal point v.

3.2 Disjoint Subsets

Apply Lemma 3.1 to $\mathcal{H}$. Let S denote the set of segments generated, and $\mathcal{H}_0$ be the set of hyperplanes of $\mathcal{H}$ not stabbed by any segment of S. Break every segment of S into $\lceil 4/\varepsilon \rceil$ equal length subsegment. For any such subsegment s, we have $|s| \leq \varepsilon \Delta^{opt}(\mathcal{H}, k)/4$. Let $S' = \{s_1, \ldots, s_m\}$ denote the resulting set of segments, where $m = O(1/\varepsilon)$ with high probability. We distribute the hyperplanes of $\mathcal{H} \setminus \mathcal{H}_0$ into the sets associated with the subsegment of S' that they intersect. If a hyperplane is stabbed by more than one segment from S', it is only distributed to (an arbitrary) one of them. Let $\mathcal{H}_i$ be the resulting set of hyperplanes associated with segment s_i, for $i = 1, \ldots, m$. By construction $|\mathcal{H}| = \sum_{i=0}^{m} |\mathcal{H}_i|$, and $|\mathcal{H}_0| = O(k)$. This redistribution process can be easily performed in linear time by using the floor function, and by observing that each of the hyperplanes of $\mathcal{H} \setminus \mathcal{H}_0$ is already associated with one of the segments of S.

Definition 3.2 A set of hyperplanes $\mathcal{J}$ is a δ-*sheaf*, if there exists a vertical segment s of length δ, such that all the hyperplanes of $\mathcal{J}$ stabs s. The vertical segment s is the *axis* of $\mathcal{J}$.[2] A 0-sheaf is referred to as a *sheaf*, in this case all hyperplanes in $\mathcal{J}$ pass through a common point, which is referred to as the *focal point* of $\mathcal{J}$. See Figure 1.

Using the above splitting approach, together with Lemma 3.1, implies the following theorem.

Theorem 3.3 *Given a set $\mathcal{J}$ of n hyperplanes in $\mathbb{R}^d$, and parameters $k, \varepsilon > 0$, one can compute, in $O(n + 1/\varepsilon)$ time, a partition of $\mathcal{J}$ into $m = O(1/\varepsilon)$ sets, $\mathcal{J}_0, \ldots, \mathcal{J}_m$, such that, with high probability, $|\mathcal{J}_0| = O(k)$, and $\mathcal{J}_i$ is an $(\varepsilon \Delta^{opt}(\mathcal{J}, k))$-sheaf, for $i = 1, \ldots, m$.*

[2]We will refer to it as the *axis of evil* when appropriate.

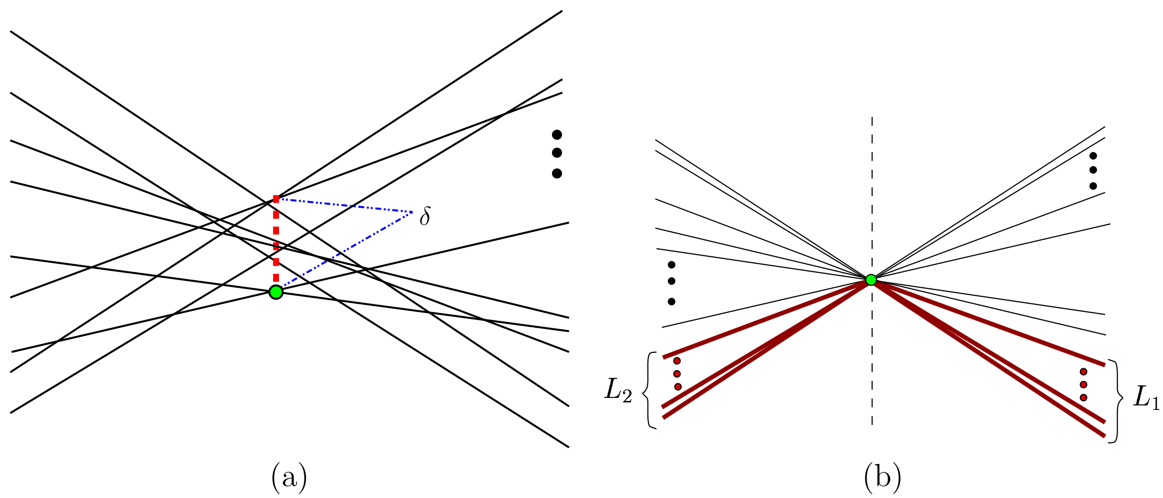


Figure 2: (a) L is a δ-sheaf. By shifting all lines downward we have $\underline{L}$ in (b). $\underline{L}'$ is composed of L_1: set of k lines with smallest slope; and L_2: set of k lines with largest slope. Thick segments form the $\leq k$-levels of L.

Lemma 3.4 *Let $\mathcal{J} = \cup_{i=0}^{m} \mathcal{J}_i$, and $\mathcal{J}' = \cup_{i=0}^{m} \mathcal{J}'_i$, where $\mathcal{J}'_i \subseteq \mathcal{J}_i$ is a (k, ε, δ)-coreset for $\mathcal{J}_i$, for $i = 0, \ldots, m$. Then $\mathcal{J}'$ is a (k, ε, δ)-coreset for $\mathcal{J}$.*

Proof: Fix any $r \leq k$, and consider any point $x \in \mathbb{R}^{d-1}$. Let $p = (x, \mathbf{L}_{\mathcal{J},r}(x))$, and let $i_0, \ldots, i_m$ be the number of hyperplanes lying below p in $\mathcal{J}_0, \ldots, \mathcal{J}_m$, respectively, such that $i_0 + \cdots + i_m = r$. Observe that $\mathbf{L}_{\mathcal{J},r}(x) = \max_j \mathbf{L}_{\mathcal{J}_j, i_j}(x)$, and that for any subset $\mathcal{Q} \subseteq \mathcal{J}$, we have $\mathbf{L}_{\mathcal{J},r}(x) \leq \mathbf{L}_{\mathcal{Q},r}(x)$ and $\mathbf{U}_{\mathcal{J},r}(x) \geq \mathbf{U}_{\mathcal{Q},r}(x)$, implying that $\mathcal{E}_{\mathcal{Q},k}(x) \leq \mathcal{E}_{\mathcal{J},k}(x)$. Therefore,

$$
\begin{aligned}
\mathbf{L}_{\mathcal{J},r}(x) &\leq \mathbf{L}_{\mathcal{J}',r}(x) \leq \max_{j=1}^{m} \mathbf{L}_{\mathcal{J}'_j, i_j}(x) \\
&\leq \max_{j=1}^{m} \big(\mathbf{L}_{\mathcal{J}_j, i_j}(x) + \varepsilon \mathcal{E}_{\mathcal{J}_j,k}(x) + \delta \big) \\
&\leq \left(\max_{j=1}^{m} \mathbf{L}_{\mathcal{J}_j, i_j}(x) \right) + \varepsilon \mathcal{E}_{\mathcal{J},k}(x) + \delta \\
&= \mathbf{L}_{\mathcal{J},r}(x) + \varepsilon \mathcal{E}_{\mathcal{J},k}(x) + \delta.
\end{aligned}
$$

A symmetric argument proves the claim for $\mathbf{U}_{\mathcal{J},r}(x)$. ∎

The above lemma implies that if we can compute a (k, ε, δ)-coreset of a small size for each $\mathcal{H}_i$ then we can easily compute such a coreset for $\mathcal{H}$.

Lemma 3.5 *For a set $\mathcal{J}$ of n hyperplanes in $\mathbb{R}^d$ and a parameter $\delta > 0$, let $\underline{\mathcal{J}}$ be the set of hyperplanes resulting by translating each hyperplane h of $\mathcal{J}$ downward by a vertical distance δ_h, where $0 \leq \delta_h \leq \delta$. Then for any $x \in \mathbb{R}^{d-1}$ and $1 \leq r, k \leq n$, we have*

(i) $\mathbf{L}_{\mathcal{J},r}(x) - \delta \leq \mathbf{L}_{\underline{\mathcal{J}},r}(x) \leq \mathbf{L}_{\mathcal{J},r}(x)$ *and* $\mathbf{U}_{\mathcal{J},r}(x) - \delta \leq \mathbf{U}_{\underline{\mathcal{J}},r}(x) \leq \mathbf{U}_{\mathcal{J},r}(x)$.

(ii) $\mathcal{E}_{\mathcal{J},k}(x) - \delta \leq \mathcal{E}_{\underline{\mathcal{J}},k}(x) \leq \mathcal{E}_{\mathcal{J},k}(x) + \delta$.

Proof: The second part follows immediately from the first one. As for the first claim, observe that $\mathbf{L}_{\underline{\mathcal{J}},r}(x) \leq \mathbf{L}_{\mathcal{J},r}(x)$ holds since all the planes of $\underline{\mathcal{J}}$ are copies of hyperplanes of $\mathcal{J}$ which were translated downward.

Next, fix any $x \in \mathbb{R}^{d-1}$, and let $h_1, \ldots, h_{n-r}$ be the hyperplanes of $\mathcal{J}$ lying above $\mathbf{L}_{\mathcal{J},r}(x)$, and let $\underline{h}_0, \ldots, \underline{h}_{n-r}$ be the corresponding hyperplanes in $\underline{\mathcal{J}}$. Clearly, $\underline{h}_0, \ldots, \underline{h}_{n-r}$ lie above point $(x, \mathbf{L}_{\mathcal{J}}(x) - \delta)$, as each hyperplane was translated down by distance at most δ. Thus $\mathbf{L}_{\mathcal{J},r}(x) - \delta \leq \mathbf{L}_{\underline{\mathcal{J}},r}(x)$. ∎

Observation 3.6 *Given a set $\mathcal{H}$ of hyperplanes in $\mathbb{R}^d$, and a (k, ε, δ)-coreset $\mathcal{H}'$ for $\mathcal{H}$. The set $\mathcal{H}'$ is a (k, μ)-coreset for $\mathcal{H}$, where $\mu = \varepsilon + \delta / \Delta^{opt}(\mathcal{H}, k)$.*

3.3 The Two Dimensional Case

Lemma 3.7 *Let L be a set of n lines in the plane which is a δ-sheaf. Given $k > 0$, one can compute, in $O(n + k)$ time, a subset $L' \subseteq L$ such that (i) $|L'| = 2k$ and (ii) L' is a $(k, 0, \delta)$-coreset for L.*

Proof: Translate each line in L downward so that it passes through the point q, where q is the bottom endpoint of the axis of the δ-sheaf. Call the resulting set $\underline{L}$; $\underline{L}$ is a sheaf in the plane with focal point q. Let $\underline{L}'$ be the union of the set of k lines of $\underline{L}$ with the largest slope, and the set of k lines of $\underline{L}$ with the smallest slope. It is easy to verify that $\underline{L}'$ is a $(k, 0)$-coreset for $\underline{L}$ (i.e., $\mathcal{A}(\underline{L})$ and $\mathcal{A}(\underline{L}')$ have the same top and bottom k levels), and $|\underline{L}'| = 2k$. See Figure 2.

Let $L' \subseteq L$ be the set of original lines that corresponds to the lines of $\underline{L}'$. By Lemma 3.5, for any $x \in \mathbb{R}$ and $0 \leq r \leq k$, we have

$$\mathbf{L}_{L,r}(x) \quad \leq \quad \mathbf{L}_{L',r}(x) \leq \mathbf{L}_{\underline{L}',r}(x) + \delta = \mathbf{L}_{\underline{L},r}(x) + \delta \leq \mathbf{L}_{L,r}(x) + \delta,$$

and

$$\mathbf{U}_{L,r}(x) \quad \geq \quad \mathbf{U}_{L',r}(x) \geq \mathbf{U}_{\underline{L}',r}(x) = \mathbf{U}_{\underline{L},r}(x) \geq \mathbf{U}_{L,r}(x) - \delta.$$

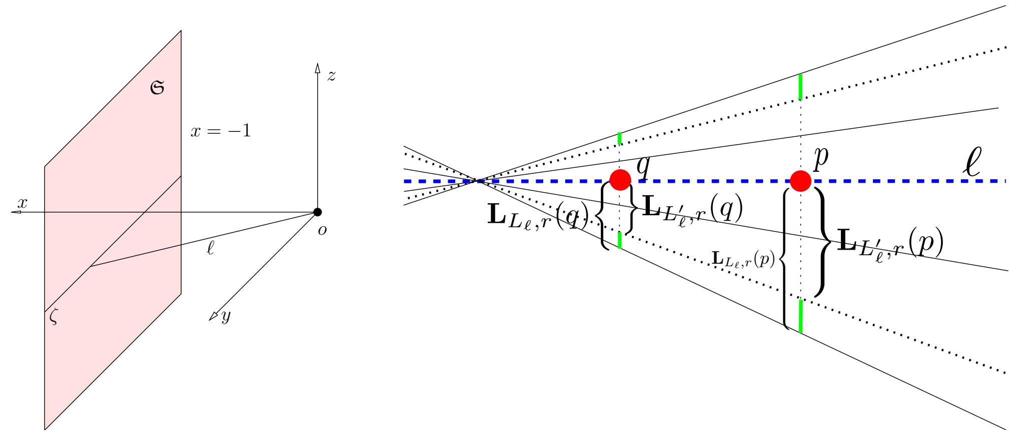
∎

Theorem 3.8 *Given a set L of n lines in the plane, and parameters $k, \varepsilon > 0$, one can compute, in $O(n + k/\varepsilon)$ time, with high probability, a set $L' \subseteq L$ such that: (i) L' is a (k, ε)-coreset for L, and (ii) $|L'| = O(k/\varepsilon)$.*

Proof: Apply Theorem 3.3 to L, and let $L = \cup_{i=0}^{m} L_i$ be the resulting set, where $m = O(1/\varepsilon)$. With high probability, the sets $L_1, \ldots, L_m$ are δ-sheaves, and L_0 has size $O(k)$, where $\delta = \varepsilon \Delta^{opt}(L, k)$. Apply Lemma 3.7 to each L_i, for $i = 1, \ldots, m$, and let $L'_i \subseteq L_i \subseteq L$ be the resulting $(k, 0, \delta)$-coreset. It then follows from Lemma 3.4 that $L' = L_0 \cup \bigcup_{i=1}^{m} L'_i$ is a $(k, 0, \delta)$-coreset for L. By Observation 3.6 the set L' is a (k, ε)-coreset for L. ∎

3.4 The Three Dimensional Case

Lemma 3.9 *Given a sheaf $\mathcal{J}$ of n planes in three dimensions with its focal point at the origin o, and parameters k, $0 < \varepsilon < 1$, one can compute, in $O(n + k/\varepsilon)$ time, a subset $\mathcal{J}'$, such that, with high probability, $\mathcal{J}'$ is a (k, ε)-coreset for $\mathcal{J}$, and $|\mathcal{J}'| = O(k/\varepsilon)$.*



(a) (b)

Figure 3: (a) Illustration of the plane $\mathfrak{S}$ and line ζ; (b) The set of lines L'_ℓ (dashed lines) is a (k,ε)-coreset for the set of lines L_ℓ, if and only if for an arbitrary point p on ℓ, which is different from the origin, L'_ℓ is a (k,ε)-coreset at p.

Proof: Consider the plane $\mathfrak{S} \equiv (x=-1) \equiv \left\{(-1,y,z) \,\middle|\, y,z \in \mathbb{R}\right\}$. Let $\zeta \subseteq \mathfrak{S}$ be the line

$$(x=-1) \cap (z=0) \equiv \left\{(-1,y,0) \,\middle|\, y \in \mathbb{R}\right\},$$

that is, the intersection between plane $\mathfrak{S}$ and the xy-plane (see Figure 3 (a)). Consider the set L_ζ of lines obtained by intersecting $\mathfrak{S}$ with the planes of $\mathcal{J}$. By Theorem 3.8, one can compute a subset $L'_\zeta \subseteq L_\zeta$ of size $O(k/\varepsilon)$ which is a (k,ε)-coreset for L_ζ. Let $\mathcal{J}'$ be set of planes of $\mathcal{J}$ corresponding to L'_ζ. We claim that $\mathcal{J}'$ is a (k,ε)-coreset for $\mathcal{J}$ at any point in the xy-plane, which implies that it is a (k,ε)-coreset for $\mathcal{J}$.

Indeed, let ℓ be any line on the xy-plane that passes through the origin, and h_ℓ the vertical plane perpendicular to the xy-plane with ℓ being its intersection with the xy-plane. Let L_ℓ be the set of lines formed by the intersection of h_ℓ and planes of $\mathcal{J}$. Clearly, L_ℓ form a sheaf structure on h_ℓ, as illustrated in Figure 3 (b).

We claim that if $\mathcal{J}' \subset \mathcal{J}$ is a (k,ε)-coreset at p, for any $p \in \ell$ (different from the origin), then it is (k,ε)-coreset at all points in ℓ. Indeed, first observe that $\mathbf{L}_{L_\ell,r}(p) = \mathbf{L}_{\mathcal{J},r}(p)$ and $\mathbf{L}_{L'_\ell,r}(p) = \mathbf{L}_{\mathcal{J}',r}(p)$, where L'_ℓ is the set of lines induced by the intersection of h_ℓ with planes of $\mathcal{J}'$. Now since $\mathcal{J}'$ is a (k,ε)-coreset at p, we have

$$\mathbf{L}_{L_\ell,r}(p) \le \mathbf{L}_{L'_\ell,r}(p) \le \mathbf{L}_{L_\ell,r}(p) + \varepsilon \mathcal{E}_{L_\ell,k}(p)$$

for any $1 \le r \le k$, which implies that, for any $q \in \ell$ (without loss of generality, we assume that q and p are on the same side of the origin on ℓ), we have $\mathbf{L}_{L_\ell,r}(q) \le \mathbf{L}_{L'_\ell,r}(q) \le \mathbf{L}_{L_\ell,r}(q) + \varepsilon \mathcal{E}_{L_\ell,k}(q)$. The last step follows from similarity of triangles since

$$\frac{\mathbf{L}_{L_\ell,r}(p)}{\mathbf{L}_{L_\ell,r}(q)} = \frac{\mathbf{L}_{L'_\ell,r}(p)}{\mathbf{L}_{L'_\ell,r}(q)} = \frac{\mathcal{E}_{L_\ell,k}(p)}{\mathcal{E}_{L_\ell,k}(q)},$$

as illustrated in Figure 3 (b).

Now consider any line ℓ that passes through the origin and assume that it intersects ζ at point u. Recall that L'_ζ is an (k, ε)-coreset for L_ζ, implying that $\mathcal{J}'$ is an (k, ε)-coreset at point u. Hence, by the above discussion, $\mathcal{J}'$ is an (k, ε)-coreset for all points on ℓ. It then follows that $\mathcal{J}'$ is a (k, ε)-coreset for all points on the xy-plane, except the points lying on the y-axis. It is now straightforward to use a limit argument to show that $\mathcal{J}'$ is a (k, ε)-coreset for all the points in the xy-plane, as required.

Note that for the sake of simplicity, we only argued about $\mathbf{L}_{L_\ell, r}$ in the preceding discussion, since the same analysis holds for $\mathbf{U}_{L_\ell, r}$ by symmetry. ■

Lemma 3.10 *Given a δ-sheaf $\mathcal{J}$ of n planes in three dimensions, and parameters k, $0 < \varepsilon < 1$, one can compute, in $O(n + k/\varepsilon)$ time, a subset $\mathcal{J}'$, such that, with high probability, $\mathcal{J}'$ is a $(k, \varepsilon, 2\delta)$-coreset for $\mathcal{J}$, and $|\mathcal{J}'| = O(k/\varepsilon)$.*

Proof: Translate the planes of $\mathcal{J}$ downward by distance at most δ, such that they all pass through o, which is the bottom vertex of the axis of $\mathcal{J}$. Let $\underline{\mathcal{J}}$ denote the resulting sheaf, and assume that the focal point o of $\underline{\mathcal{J}}$ is in the origin. By Lemma 3.9, one can compute a subset $\underline{\mathcal{J}'} \subseteq \underline{\mathcal{J}}$, which is a $(k.\varepsilon)$-coreset for $\underline{\mathcal{J}}$, $|\underline{\mathcal{J}'}| = O(k/\varepsilon)$, and this set can be computed in $O(n + k/\varepsilon)$ time.

Let $\mathcal{J}'$ be the set of original planes that corresponds to the planes of $\underline{\mathcal{J}'}$. Then $\mathcal{J}'$ is a $(k, \varepsilon, 2\delta)$-coreset for $\mathcal{J}$. Indeed, by Lemma 3.5, for any $p \in \mathbb{R}^2$,

$$\begin{aligned} \mathbf{L}_{\mathcal{J},r}(p) &\le \mathbf{L}_{\mathcal{J}',r}(p) \le \mathbf{L}_{\underline{\mathcal{J}'},r}(p) + \delta \le \mathbf{L}_{\underline{\mathcal{J}},r}(p) + \varepsilon \mathcal{E}_{\underline{\mathcal{J}},k}(p) + \delta \\ &\le \mathbf{L}_{\mathcal{J},r}(p) + \varepsilon \mathcal{E}_{\underline{\mathcal{J}},k}(p) + \delta \le \mathbf{L}_{\mathcal{J},r}(p) + \varepsilon \mathcal{E}_{\mathcal{J},k}(p) + 2\delta, \end{aligned}$$

since $\mathcal{E}_{\underline{\mathcal{J}},k}(p) \le \mathcal{E}_{\mathcal{J},k}(p) + \delta$ and $\varepsilon < 1$. ■

Theorem 3.11 *Given a set $\mathcal{H}$ of n planes in three dimensions, and parameters k, ε, one can compute, in $O(n + k/\varepsilon^2)$ time, a subset $\mathcal{H}' \subseteq \mathcal{H}$, such that, with high probability, it holds: (i) $\mathcal{H}'$ is a (k, ε)-coreset for $\mathcal{H}$, and (ii) $|\mathcal{H}'| = O(k/\varepsilon^2)$.*

Proof: Apply Theorem 3.3 to $\mathcal{H}$ with $\varepsilon/4$ and k. This results in a partition of $\mathcal{H}$ into sets $\mathcal{H}_0, \mathcal{H}_1, \ldots, \mathcal{H}_m \subseteq \mathcal{H}$, where $\mathcal{H}_1, \ldots, \mathcal{H}_m$ are δ-sheafs, $\delta = (\varepsilon/4)\Delta^{opt}(\mathcal{H}, k)$, $m = O(1/\varepsilon)$, and $|\mathcal{H}_0| = O(k)$ with high probability..

Applying Lemma 3.10 to $\mathcal{H}_i$, for $i = 1, \ldots, m$, results in a $(k, \varepsilon/2, 2\delta)$-coreset $\mathcal{H}'_i$ for $\mathcal{H}_i$, where $|\mathcal{H}'_i| = O(k/\varepsilon)$.

Let $\mathcal{H}' = \mathcal{H}_0 \cup \bigcup_{i=1}^m \mathcal{H}'_i$. By Lemma 3.4, $\mathcal{H}'$ is a $(k, \varepsilon/2, 2\delta)$-coreset for $\mathcal{H}$, which in turn implies that $\mathcal{H}'$ is a (k, ε)-coreset for $\mathcal{H}$, by Observation 3.6.

Finally, $|\mathcal{H}'| = |\mathcal{H}_0| + \sum_{i=1}^m |\mathcal{H}'_i| = O(k + (1/\varepsilon)(k/\varepsilon)) = O(k/\varepsilon^2)$. The overall running time is $O(n + 1/\varepsilon + n + k/\varepsilon^2) = O(n + k/\varepsilon^2)$. ■

3.5 The Higher Dimensional Case

Let $f(n, k, \varepsilon, d)$ denote a bound on the size of (k, ε)-coreset of n hyperplanes in $\mathbb{R}^d$. We know that $f(n, k, \varepsilon, 2) = O(k/\varepsilon)$ and $f(n, k, \varepsilon, 3) = O(k/\varepsilon^2)$ by Theorem 3.8 and Theorem 3.11, respectively. Let $T(n, k, \varepsilon, d)$ denote the time needed to compute such a set.

The proof of the following lemma is a direct extension of the proof of Lemma 3.10.

Lemma 3.12 *Given a δ-sheaf $\mathcal{J}$ of n hyperplanes in $\mathbb{R}^d$, and parameters $k, \varepsilon > 0$, then, one can compute, in $O(T(n, k, \varepsilon, d-1))$ time, a subset $\mathcal{J}'$, such that $\mathcal{J}'$ is a $(k, \varepsilon, 2\delta)$-coreset for $\mathcal{J}$, and $|\mathcal{J}'| = f(n, k, \varepsilon, d-1)$.*

Theorem 3.13 *Given a set $\mathcal{H}$ of n hyperplanes in $\mathbb{R}^d$, and parameters $k, \varepsilon > 0$, with high probability, one can compute, in $T(n, k, \varepsilon, d) = O(n + k/\varepsilon^{d-1})$ time, a subset $\mathcal{H}'$, such that (i) $\mathcal{H}'$ is a (k, ε)-coreset for $\mathcal{H}$ and (ii) $f(n, k, \varepsilon, d) = |\mathcal{H}'| = O(k/\varepsilon^{d-1})$.*

Proof: The proof of the theorem is a straightforward extension of the proof of Theorem 3.11, and as such we only verify the bounds on the coreset size and running time. We have:

$$f(n, k, \varepsilon, d) = O(k) + O\left(\frac{1}{\varepsilon} f(n, k, \varepsilon, d-1)\right) = O\left(\frac{k}{\varepsilon^{d-1}}\right).$$

As for the running time, we have:

$$T(n, k, \varepsilon, d) = O\left(n + \frac{1}{\varepsilon}\right) + \sum_{i=1}^{O(1/\varepsilon)} T(n_i, k, \varepsilon, d-1),$$

where $\sum_i n_i \leq n$ and $T(n, k, \varepsilon, 3) = O(n + k/\varepsilon^2)$, by Theorem 3.11. Thus, $T(n, k, \varepsilon, d) = O(n + k/\varepsilon^{d-1})$, as claimed. ■

4 Polynomials and Their Roots

In this section, we extend previous results to the extent of a family of polynomials and their roots.

4.1 Polynomials

Let $\mathcal{F} = \{f_1, \ldots, f_n\}$ be a family of d-variate polynomials and let $u_1, \ldots, u_d$ be the variables over which the functions of $\mathcal{F}$ are defined. Each f_i corresponds to a surface in $\mathbb{R}^{d+1}$: For example, any d-variate linear function can be considered as a hyperplane in $\mathbb{R}^{d+1}$ (and vice versa). We extend, in the natural way, the definitions of k-level, (k, r)-extent, ε-approximation, and (k, ε, δ)-coreset for the arrangement $\mathcal{A}(\mathcal{F})$.

Each monomial over $u_1, \ldots, u_d$ appearing in $\mathcal{F}$ can be mapped to a distinct variable x_i. Let $x_1, \ldots, x_s$ be the resulting variables. As such $\mathcal{F}$ can be linearized into a set $\mathcal{H} = \{h_1, \ldots, h_n\}$ of linear functions over $\mathbb{R}^s$. In particular, $\mathcal{H}$ is a set of n hyperplanes in $\mathbb{R}^{s+1}$. Note that the surface induced by f_i in $\mathbb{R}^{d+1}$ corresponds only to a subset of the surface of h_i in $\mathbb{R}^{s+1}$. This technique is called *linearization*, and has been widely used in fields such as machine learning [CS00] and computational geometry [AM94].

For example, consider a family of polynomials $\mathcal{F} = \{f_1, \ldots, f_n\}$, where $f_i(x, y) = a_i(x^2 + y^2) + b_i x + c_i y + d_i$, and $a_i, b_i, c_i, d_i \in \mathbb{R}$, for $i = 1, \ldots, n$. This family of polynomials defined over $\mathbb{R}^2$, can be linearized to a family of linear functions defined over $\mathbb{R}^3$, by $h_i(x, y, z) = a_i z + b_i x + c_i y + d_i$, and setting $\mathcal{H} = \{h_1, \ldots, h_n\}$. Clearly, $\mathcal{H}$ is a set of hyperplanes in $\mathbb{R}^4$, and $f_i(x, y) = h_i(x, y, x^2 + y^2)$. Thus, for any point $(x, y) \in \mathbb{R}^2$, instead of evaluating $\mathcal{F}$ on

(x, y), we can evaluate $\mathcal{H}$ on $\eta(x, y) = (x, y, x^2 + y^2)$, where $\eta(x, y)$ is the *linearization image* of (x, y). The advantage of this linearization is that $\mathcal{H}$, being a family of linear functions, is now easier to handle than $\mathcal{F}$. There is a general technique for finding the best possible linearization (i.e., a mapping η with the target dimension as small as possible), see [AM94] for details. Observe, that $X = \eta(\mathbb{R}^2)$ is a *subset* of $\mathbb{R}^3$ (this is the "standard" paraboloid), and we are interested in the value of $\mathcal{H}$ only on points belonging to X. In particular, the set X is not necessarily convex. The set X resulting from the linearization is a semi-algebraic set of constant complexity, and as such basic manipulation operations of X can be performed in constant time.

Note that for each $1 \leq i \leq n$, $f_i(p) = h_i(\eta(p))$ for $p \in \mathbb{R}^d$. As such, if $\mathcal{H}' \subseteq \mathcal{H}$ is a (k, ε)-coreset for $\mathcal{H}$, then clearly the corresponding subset in $\mathcal{F}$ is a (k, ε)-coreset for $\mathcal{F}$. The following theorem is a restatement of Theorem 3.13 in this settings.

Theorem 4.1 *Given a family of d-variate polynomials $\mathcal{F} = \{f_1, \ldots, f_n\}$, and parameters k and ε, one can compute, in $O(n + k/\varepsilon^s)$ time, a subset $\mathcal{F}' \subseteq \mathcal{F}$ of $O(k/\varepsilon^s)$ polynomials, such that $\mathcal{F}'$ is a (k, ε)-coreset for $\mathcal{F}$, with high probability. Here s is the number of different monomials present in the polynomials of $\mathcal{F}$.*

4.2 Roots of Polynomials

It turns out that for roots of polynomials the definition of coreset (Definition 2.4) is somewhat too restrictive. However, we can get a similar notion to a coreset by slightly relaxing the requirements.

Definition 4.2 Let $\mathcal{F}$ be a set of non-negative functions defined over $\mathbb{R}^d$. A subset $\mathcal{F}' \subseteq \mathcal{F}$ is (k, ε)-*sensitive* if for any $r \leq k$ and $x \in \mathbb{R}^d$, we have

$$\mathbf{L}_{\mathcal{F},r}(x) \leq \mathbf{L}_{\mathcal{F}',r}(x) \leq \mathbf{L}_{\mathcal{F},r}(x) + \frac{\varepsilon}{2}\mathcal{F}|_r^k(x); \quad \text{and}$$

$$\mathbf{U}_{\mathcal{F},r}(x) - \frac{\varepsilon}{2}\mathcal{F}|_k^r(x) \leq \mathbf{U}_{\mathcal{F}',r}(x) \leq \mathbf{U}_{\mathcal{F},r}(x).$$

Note that any (k, ε)-coreset is $(k, 2\varepsilon)$-sensitive (while the inverse is not necessarily true).

Lemma 4.3 *If $\mathcal{F}'$ is (k, ε)-sensitive for $\mathcal{F}$, then*

$$(1 - \varepsilon)\, \mathcal{F}|_r^t(x) \leq \mathcal{F}'|_r^t(x) \leq \mathcal{F}|_r^t(x),$$

for any $t, r \leq k$ and $x \in \mathbb{R}^d$.

Proof:
$$\begin{aligned}
\mathcal{F}|_r^t(x) \geq \mathcal{F}'|_r^t(x) &= \mathbf{U}_{\mathcal{F}',t}(x) - \mathbf{L}_{\mathcal{F}',r}(x) \\
&\geq \left(\mathbf{U}_{\mathcal{F},t}(x) - \frac{\varepsilon}{2}\mathcal{F}|_k^t(x)\right) - \left(\mathbf{L}_{\mathcal{F},r}(x) + \frac{\varepsilon}{2}\mathcal{F}|_r^k(x)\right) \\
&= \left(\mathbf{U}_{\mathcal{F},t}(x) - \mathbf{L}_{\mathcal{F},r}(x)\right) - \frac{\varepsilon}{2}\left(\mathcal{F}|_k^t(x) + \mathcal{F}|_r^k(x)\right) \\
&\geq \mathcal{F}|_r^t(x) - \varepsilon\mathcal{F}|_r^t(x) = (1 - \varepsilon)\, \mathcal{F}|_r^t(x).
\end{aligned}$$
■

Theorem 4.4 *Let $\mathcal{F} = \left\{ f_1^{1/2}, \ldots, f_n^{1/2} \right\}$ be a family of d-variate functions defined over $\mathbb{R}^d$, where f_i is a d-variate polynomial, for $i = 1, \ldots, n$. Given k and $0 < \varepsilon < 1$, one can compute, in $O(n + k/\varepsilon^{2s})$ time, a subset $\mathcal{F}' \subseteq \mathcal{F}$, such that, with high probability, $\mathcal{F}'$ is (k, ε)-sensitive for $\mathcal{F}$, and $|\mathcal{F}'| = O(k/\varepsilon^{2s})$, where s is the number of distinct monomials present in the polynomials of $\mathcal{F}$.*

Proof: Let $\mathcal{G} = \left\{ f_i \,\middle|\, f_i^{1/2} \in \mathcal{F} \right\}$ denote the set of d-variate polynomials which are the square of the functions in $\mathcal{F}$. Let $\mathcal{H}$ be the set of hyperplanes in $\mathbb{R}^{s+1}$ obtained by linearization of f_i's, and let $\eta : \mathbb{R}^d \to \mathbb{R}^s$ be the linearization used. Given any $k, \varepsilon > 0$, one can compute in $O(n + k/\delta^s)$ time, a (k, δ)-coreset $\mathcal{H}' \subseteq \mathcal{H}$ of size $O(k/\delta^s)$ for $\mathcal{H}$, where $\delta = \varepsilon^2/32$. Clearly, $\mathcal{H}'$ is $(k, 2\delta)$-sensitive for $\mathcal{H}$.

Let $\mathcal{G}' \subseteq \mathcal{G}$ be the set of functions of $\mathcal{G}$ that corresponds to $\mathcal{H}'$. Let α be any point in $\mathbb{R}^d$ and $x = \eta(\alpha) \in \mathbb{R}^s$ be the linearized image of α. Let $a = \mathbf{L}_{\mathcal{G},r}(\alpha) = \mathbf{L}_{\mathcal{H},r}(x)$, $A = \mathbf{L}_{\mathcal{G}',r}(x) = \mathbf{L}_{\mathcal{H}',r}(x)$, $b = \mathbf{U}_{\mathcal{G},k}(\alpha) = \mathbf{U}_{\mathcal{H},k}(x)$. By the definition of $(k, 2\delta)$-sensitivity, we have that $A - a \le \delta\, \mathcal{G}|_r^k(x) = \delta(b - a)$.

Now, if $\sqrt{A} + \sqrt{a} \le 2(\delta/\varepsilon)(\sqrt{b} + \sqrt{a})$ then $\sqrt{A} + \sqrt{a} \le 4\sqrt{b}\delta/\varepsilon \le \sqrt{b}\varepsilon/8$. Thus,

$$
\begin{aligned}
\sqrt{A} - \sqrt{a} &\le \sqrt{A} + \sqrt{a} \le \frac{\sqrt{b}\varepsilon}{8} \le \left(\frac{\varepsilon}{2} - \frac{\varepsilon}{8}\right)\sqrt{b} = \frac{\varepsilon}{2}\sqrt{b} - \frac{\varepsilon}{8}\sqrt{b} \le \frac{\varepsilon}{2}\sqrt{b} - \sqrt{a} \\
&\le \frac{\varepsilon}{2}(\sqrt{b} - \sqrt{a}),
\end{aligned}
$$

since $\sqrt{a} \le \sqrt{A} + \sqrt{a} \le \sqrt{b}\varepsilon/8$.

Otherwise, $\sqrt{A} + \sqrt{a} \ge 2(\delta/\varepsilon)(\sqrt{b} + \sqrt{a})$ and

$$
\begin{aligned}
\mathbf{L}_{\mathcal{F}',r}(\alpha) - \mathbf{L}_{\mathcal{F},r}(\alpha) &= \sqrt{A} - \sqrt{a} = \frac{A - a}{\sqrt{A} + \sqrt{a}} \le \frac{\delta(b - a)}{\sqrt{A} + \sqrt{a}} \le \frac{\delta(b - a)}{2(\delta/\varepsilon)(\sqrt{b} + \sqrt{a})} \\
&= (\varepsilon/2)(\sqrt{b} - \sqrt{a}) \\
&\le \frac{\varepsilon}{2}\mathcal{F}|_r^k(\alpha).
\end{aligned}
$$

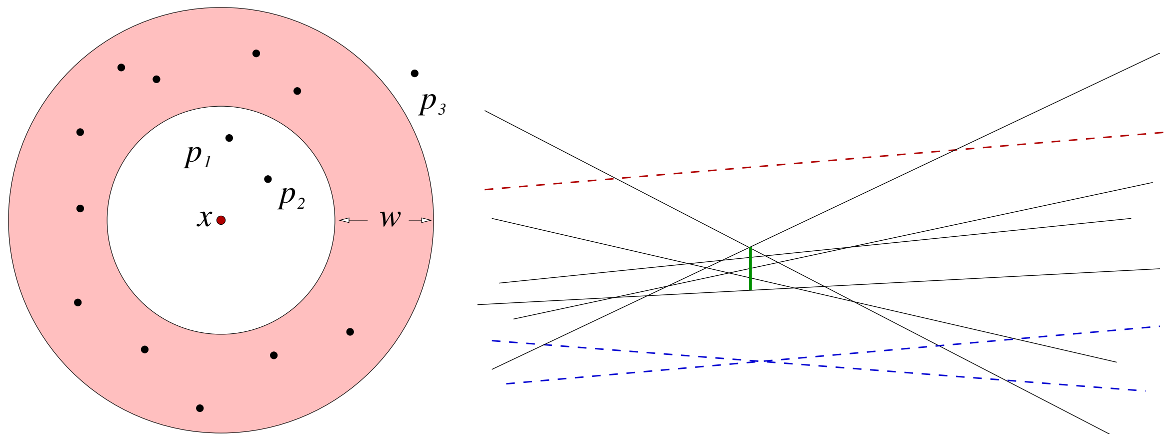
■

5 Applications

In this section, we shortly present some of the results that follow from our technique for various shape fitting problems with outliers. Some of the other results mentioned in the introduction follows by a straightforward application of our techniques, as demonstrated in this section. To avoid tedious repetition we only present the more interesting ones here.

The general framework is as follows: First, reduce the problem of finding the best fitting shape into that of computing the smallest extent over a family of polynomials or roots of polynomials. Next, apply the following theorems to approximate the smallest extent with outliers. The details are described later in this section.

Theorem 5.1 *Given a $(k, \varepsilon/2)$-sensitive subset $\mathcal{F}'$ of a family $\mathcal{F}$ of d-variate functions, where $0 < \varepsilon < 1$. One can compute*



(a) (b)

Figure 4: (a) Annulus with center x (dark region) has a width w. p_1, p_2 and p_3 are outliers. (b) After linearization, thick vertical segment is the shortest extent with 3 outliers. with dotted lines corresponding to outliers p_1, p_2 and p_3.

(i) *A point* $x' \in \mathbb{R}^d$ *and* $\delta' = \mathcal{E}_{\mathcal{F},k}(x')$*, such that* $\Delta^{opt}(\mathcal{F}, k) \leq \delta' \leq (1+\varepsilon)\Delta^{opt}(\mathcal{F}, k)$.

(ii) *A point* $\widehat{x} \in \mathbb{R}^d$ *and* $\widehat{\delta} = \mathcal{F}|_{\widehat{t}}^{\widehat{r}}(\widehat{x})$*, such that* $\widehat{r} + \widehat{t} = k$ *and* $\omega^* \leq \widehat{\delta} \leq (1+\varepsilon)\omega^*$*, where* $\omega^* = \min_{x \in \mathbb{R}^d, r+t=k} \mathcal{F}|_t^r(x)$.

The running time is $O(n + \mathcal{C}(\mathcal{F}')^2)$*, where* $n = |\mathcal{F}|$ *and* $\mathcal{C}(\mathcal{F}')$ *is the complexity of the arrangement* $\mathcal{A}(\mathcal{F}')$.

Proof: We prove the first claim, as the second claim follows from a similar argument.

Each function in $\mathcal{F}$ induces a surface in $\mathbb{R}^{d+1}$. Set $x' \in \mathbb{R}^d$ to be the point such that $\mathcal{E}_{\mathcal{F}',k}(x') = \Delta^{opt}(\mathcal{F}', k) = \min_{x \in \mathbb{R}^d} \mathcal{E}_{\mathcal{F}',k}(x)$. Note that $\delta' = \mathcal{E}_{\mathcal{F},k}(x')$ can be computed in linear time once x' is specified; while x' can be obtained by computing the levels $\mathbf{L}_{\mathcal{F}',\leq k}$ and $\mathbf{U}_{\mathcal{F}',\leq k}$ and then searching, by brute force, for the shortest extent in the arrangement $\mathcal{A}(\mathcal{F}')$. It is easy to verify that this can be done in $O(n + \mathcal{C}(\mathcal{F}')^2)$ time.

Let $x^* \in \mathbb{R}^d$ be the point realizing $\mathcal{E}_{\mathcal{F},k}(x^*) = \Delta^{opt}(\mathcal{F}, k)$. Since $\mathcal{F}'$ is a $(k, \varepsilon/2)$-sensitive, and $\varepsilon < 1$, then by Lemma 4.3 we have

$$\begin{aligned} \Delta^{opt}(\mathcal{F}, k) &\leq \delta' = \mathcal{E}_{\mathcal{F},k}(x') \leq \frac{\mathcal{E}_{\mathcal{F}',k}(x')}{1 - \varepsilon/2} \\ &\leq (1+\varepsilon)\mathcal{E}_{\mathcal{F}',k}(x') \leq (1+\varepsilon)\mathcal{E}_{\mathcal{F}',k}(x^*) \\ &\leq (1+\varepsilon)\mathcal{E}_{\mathcal{F},k}(x^*) = (1+\varepsilon)\Delta^{opt}(\mathcal{F}, k). \end{aligned}$$

■

Note that the above theorem implies that we can not only approximate the value of the smallest extend, but also find an instance that achieves the approximate value. The running time of Theorem 5.1 can be slightly improved by being more careful in the analysis, and by applying the Clarkson-Shor technique [CS89] to bound more tightly the number of pairs of facets of the arrangement that should be considered, when computing $\Delta^{opt}(\mathcal{F}', k)$.

All the results in the remainder of this section hold with high probability, as they all rely on the randomized algorithm of Lemma 3.1.

5.1 Min-width annulus/spherical shell

Given a set $P = \{p_1, \ldots, p_n\}$ of points in $\mathbb{R}^d$, and parameters $k, \varepsilon > 0$, let $\omega_{opt,k}$ denote the width of the thinnest annulus containing all but k points in P. See Figure 4. We would like to compute the annulus that contains at least $n-k$ points of P, and whose width is at most $(1+\varepsilon)\omega_{opt,k}$. More precisely, let $d(x,p)$ denote the distance between two points $x, p \in \mathbb{R}^d$, and

$$\mu(x, P, k) = \min_{P' \subseteq P, |P'| = n-k} \left(\max_{p \in P'} d(x,p) - \min_{p \in P'} d(x,p) \right).$$

We omit k from the notation when $k = 0$. We have $\omega_{opt,k} = \min_{x \in \mathbb{R}^d} \mu(x, P, k)$, and we wish to compute a subset $P^* \subseteq P$ of size at least $n-k$ and a point $x^* \in \mathbb{R}^d$ such that $\mu(x^*, P^*) \leq (1+\varepsilon)\omega_{opt,k}$.

Let $p_i = (\xi_1, \ldots, \xi_d)$ be a point from the set P, and let $x = (x_1, \ldots, x_d)$ be an arbitrary point in $\mathbb{R}^d$. Under the euclidean distance, we define

$$f_i(x) = d(p_i, x) = \sqrt{\sum_{1 \leq j \leq d} x_j^2 - 2 \sum_{1 \leq j \leq d} \xi_j x_j + \sum_{1 \leq j \leq d} \xi_j^2}.$$

Let $\mathcal{F} = \{f_1, \ldots, f_n\}$ denote the set of functions defined by the points of P.

Observe that $\omega_{opt,k}$ is in fact the shortest (r,t)-extent for $\mathcal{F}$ such that $r + t = k$. By Theorem 4.4, we can compute, in $O(n + k/\varepsilon^{2d})$ time, a subset $\mathcal{F}' \subseteq \mathcal{F}$ which is a (k,ε)-sensitive for $\mathcal{F}$ and $|\mathcal{F}'| = O(k/\varepsilon^{2d})$. Since the image of each function of $\mathcal{F}$ is a cone in $\mathbb{R}^{d+1}$, one can thus approximate $\omega_{opt,k}$ in

$$O\Big(n + |\mathcal{F}'|^{2(d+1)}\Big) = O\left(n + \frac{k^{2(d+1)}}{\varepsilon^{4d(d+1)}}\right)$$

time, using the algorithm of Theorem 5.1.

Theorem 5.2 *Given a set P of n points and parameters $k, \varepsilon > 0$, one can compute an $x^* \in \mathbb{R}^d$, and $P^* \subseteq P$ where $|P^*| \geq n-k$, such that $\mu(x^*, P^*) \leq (1+\varepsilon)\omega_{opt,k}$ in $O(n + k^{2d+1}/\varepsilon^{4d(d+1)})$ time.*

5.2 Min-volume annulus/spherical shell

Let $\mathcal{B}(x, R, r)$ denote the shell between two concentric balls, centered at point x, with radius R and r, respectively. That is, $\mathcal{B}(x, R, r)$ is composed of all points inside the outer ball, and outside the inner ball. Given a set of points $P \in \mathbb{R}^d$, we would like to find a spherical shell whose volume is minimized, and which contains at least $n-k$ points of P, over all $x \in \mathbb{R}^d$ and $R, r \in \mathbb{R}$. Let $\mathcal{SS}_{opt,k}(P)$ denote the volume of this spherical shell. This problem is easier than the min-width variant, as it can be reduced to linear programming (this is folklore, it is also described in [AAHS00]).

The current fastest algorithm for the exact problem, in two dimensions, with k outliers is due to Chan [Cha02] and it works in $O\big(n \log k + k^{11/4} n^{1/4} \log^c k\big)$ time.

Theorem 5.3 *Given a set P of n points and parameters $k, \varepsilon > 0$, one can compute a set P, such that (i) $P' \subseteq P$, (ii) P is a (k, ε)-coreset for the minimum area spherical shell measure, and (iii) $|P'| = O(k/\varepsilon^d)$.*

In particular, one can compute a spherical shell whose volume ε-approximates $\mathcal{SS}_{opt,k}(P)$ in $O(n + k^{2(d+1)}/\varepsilon^{2d(d+1)})$ time, for $d > 2$, and in $O\Big(n + \frac{k}{\varepsilon^2}\log\frac{k}{\varepsilon} + \frac{k^3}{\varepsilon^{1/2}}\log^c k\Big)$ time, for $d = 2$, where c is a constant.

Proof: Using linearization, computing $\mathcal{SS}_{opt}(P)$ is reduced to computing the extent of n hyperplanes in $\mathbb{R}^{d+1}$ [AHV03]. Thus, by Theorem 3.13, there is a subset P' of P, which is (k, ε)-sensitive for the measure $\mathcal{SS}_{opt}(\cdot)$, and $|P'| = O(k/\varepsilon^d)$.

For $d > 2$, the running time is just the result of applying Theorem 5.1. As for $d = 2$, we use the algorithm of Chan [Cha02] on the subset P'. ■

5.3 Minimum-Width Cylindrical Shell

Given a line ℓ in $\mathbb{R}^d$ and two real numbers $0 \le r \le R$, the cylindrical shell $S(\ell, r, R)$ is the closed region lying between the two co-axial cylinders of radii r and R respectively, with ℓ as their axis. The width of shell $S(\ell, r, R)$ is $R - r$. Given a set of n points P, let $\mathcal{CS}^*$ denote the minimum width among all cylindrical shells that enclose at least $n - k$ points of P. In the *approximate minimum-width cylindrical shell with outliers* problem, we would like to compute a cylindrical shell containing at least $n - k$ points from P whose width is at most $(1 + \varepsilon)\mathcal{CS}^*$.

Theorem 5.4 *Given a set P of n points in $\mathbb{R}^d$ and parameters $k, \varepsilon > 0$, one can compute in $O(n + k^{cd^2}/\varepsilon^{c'd^4})$ time, a line ℓ^* and $P^* \subseteq P$ where $|P^*| \ge n - k$, such that the minimum width of the cylindrical shell containing P^* with axis ℓ^* is smaller than $(1+\varepsilon)\mathcal{CS}^*$. Here c, c' are constants independent of d, ε, and n.*

Proof: Using linearization, computing $\mathcal{CS}^*$ is reduced to computing the extent of n hyperplanes in $\mathbb{R}^{O(d^2)}$ [AHV03]. Thus, by Theorem 3.13, there is a subset P' of P, which is (k, ε)-sensitive for the width measure, and $|P'| = O\Big(k/\varepsilon^{O(d^2)}\Big)$. The running time is just the result of applying Theorem 5.1 to this set. ■

5.4 Moving Points

The same technique can also be extended to approximate various measures of moving points. More precisely, given a set P of n points moving in $\mathbb{R}^d$, for each $p \in P$, let $p(t) = (\xi_1(t), \ldots, \xi_d(t))$ denote the position of point p at time t, and let $P(t)$ denote the set P at time t. We say that the motion of P is *algebraic* of degree ν if the functions $\xi_1(t), \ldots, \xi_d(t)$ are polynomials of degree at most ν for all the points of P. For simplicity, we assume in the remaining part that $\nu = 1$, i.e., each point moves along a straight line.

Let $\mu_k(P)$ denote some measure of the point set P, such as width (with at most k outliers). For any $\varepsilon > 0$, we say that a subset $Q \subseteq P$ ε-approximates P with respect to μ_k if at any time t, $(1 - \varepsilon)\mu_k(P(t)) \le \mu_k(Q(t)) \le \mu_k(P(t))$. Arguing as in [AHV03], we get the following result.

Theorem 5.5 *Given a point-set P with n linearly-moving points, and parameters $k, \varepsilon > 0$, one can compute, in $O(n + k/\varepsilon^c)$ time, a subset $Q \subseteq P$ of size $O(k/\varepsilon^c)$, such that Q is an ε-approximation to P under measures with k outliers such as: diameter, width, minimum-radius enclosing ball, projection width, min-width annulus, min width-cylindrical shell. Here c is a constant that depends on the dimension d and the measure considered.*

5.5 Handling Updates – Insertions and Deletions

By building a balanced binary tree over the given point set, and maintaining in each node the coreset of all the points in its subtree, one can maintain a coreset under insertions and deletions, where every insertion and deletion is handled in poly-logarithmic time. The following theorem follows in a plug and play fashion by combining the techniques of [AH01] with our techniques.

Theorem 5.6 *Let P be a set of points in $\mathbb{R}^d$, k and $\varepsilon > 0$ be parameters, and $\mu(\cdot)$ one of the measures discussed before. Suppose that for any $Q \subseteq P$, we can compute an a (k, ε)-sensitive (for $\mu(\cdot)$) subset S of Q of size $O(k/\varepsilon^\rho)$ in $O(|Q| + f(\varepsilon))$ time. Then, we can maintain a (k, ε)-sensitive subset of P of size $O(k/\varepsilon^\rho)$ under insertion/deletion operations in $O(((\log n)/\varepsilon)^\rho + f(\varepsilon/\log n) \log n)$ time per update.*

Note, that Theorem 5.6 implies that we can maintain the approximate optimal solution for *all* of the measures discussed above, under insertions and deletions, with k outliers. The time to handle an update is $O(\mathrm{poly}(k, 1/\varepsilon, \log n))$, where $\mathrm{poly}(\cdots)$ is a constant degree polynomial in its parameters, where the constant is a function of d and the measure being approximated.

6 Conclusions

We presented a general approximation algorithm for shape fitting that can handle outliers efficiently with near linear running time when the number of outliers is small. The main contribution of this paper is the proof that there exists a small coreset for various shape fitting problems when considering outliers. The authors believe that the existence of such a small coreset is quite surprising.

6.1 Comparison to Linear Programming with Violations

It is beneficial to compare our result to the algorithms known for linear programming with violations. Mulmuley [Mul91] showed using the Clarkson-Shor technique that in an arrangement of n hyperplanes in d dimensions, there are at most $O(k^d)$ *local* minima in the top k-level. The argument relies on picking a random sample of n/k hyperplanes, and showing that the probability that a local minima in the top k-level is the minimum of the sample (i.e., the lowest point on the upper envelope of the sample of hyperplanes) is $\Omega(1/k^d)$, and since there is only one local minima in the upper envelope of the sample, it follows that the global number of local minima in the first top k-levels is $O(k^d)$. Matoušek [Mat95b] solves the linear programming with violations by generating all those local minima explicitly (in

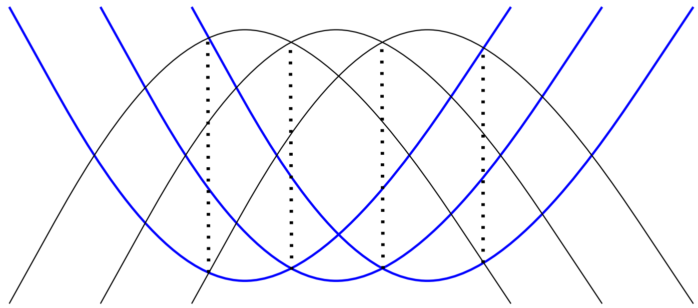


Figure 5: The graphs of a set of functions $\mathcal{F}$, with each being a parabola. Dotted vertical segments depict the local minima of 0-extent in the arrangement of $\mathcal{F}$.

amortized $O(n)$ time for each one by solving instances of linear programming), and returning the best one.

A straightforward adaption of this argument, shows that there are $O(k^{d+1})$ minima to the k-extent function $\mathcal{E}_{\mathcal{H},k}$. In our setting, since we are interested only in the extent in a subspace $X \subseteq \mathbb{R}^{d-1}$, it is no longer true that such a bound (that depends only on k) holds on the number of minima of the extent when restricted to X. Specifically, it is easy to find examples where the number of local minima of the k-extent is $\Omega(n)$ when restricted to X. We briefly describe such an example below. Imagine a set of functions $\mathcal{F}$ where each $f \in \mathcal{F}$ is a parabola $f(x) = ax^2 + bx + c$ (see Figure 5). The number of local minima of the 0-extent is $\Omega(n)$ as indicated by dotted vertical segment in Figure 5. Now we linearize each f into a plane $h(x, y) = ay + bx + c$ in $\mathbb{R}^3$, with the linearization image being $\eta(x) = (x, x^2)$; call the resulting set of planes $\mathcal{H}$. The set X, in this case, is the image of the linearization, namely $X = \eta(\mathbb{R}) = \left\{ (x, x^2) \,\middle|\, x \in \mathbb{R} \right\}$. Observe, that the number of local minima of the 0-extent along $\eta(\mathbb{R})$ for $\mathcal{H}$ remains the same. Therefore, while the number of minima of the 0-extent with respect to $\mathbb{R}^2$ is $O(1)$ (i.e., there is one minimum in the upper and lower envelop of the arrangement of $\mathcal{H}$), the number of minima when restricted to X is $\Omega(n)$. Furthermore, the problem of finding the shortest vertical segment when restricted to X (even without outliers) is no longer an LP-type problem, and as such we can no longer move efficiently from one minimum to another using Matoušek's approach.

Nevertheless, this gives a new interpretation of our result. It implies that if we are only interested in approximation, the number of approximate local minima is bounded by a polynomial in k (while the exact one is *not*). Furthermore, our coreset result yields a direct way to generate those local minima.

6.2 General Discussion

A limitation of our technique is that the number of outliers has to be moderately small (i.e., $O(n^{1/2d})$) to achieve near linear running time. Erickson and Seidel [ES95, Eri99] show that to determine whether a set of n points in $\mathbb{R}^d$ contains $d + 1$ points on a common hyperplane requires $\Omega(n^d)$ sideness queries. (Given $d + 1$ points $p_0, \ldots, p_d$, the *sideness query* decides which side the point p_0 lies to the oriented hyperplane defined by $p_1, \ldots, p_n$.) This lower bound holds in a decision tree model of computation in which every decision is based on the result of a sideness query, which they argue is a reasonable model. Their result implies that when the number of outliers is huge (i.e., $k = n - d - 1$), a near-linear time approximation algorithm cannot be achieved. We leave the problem of improving the running time of our algorithm as an open question for further research.

This raises the question whether one can find "universal" approximation algorithms with outliers. Namely, approximation algorithms that have near linear running time independent of the number of outliers. Currently, the only case the authors are aware of that such an algorithm exists, is for the case of minimum enclosing circle with k outliers, see [Mat95a] for the currently fastest algorithm known for this problem.

Of course, using standard ε-sampling techniques, one can achieve *combinatorial* approximation of k levels for relatively large k efficiently. However, this provides a different and *weaker* type of approximation than the one presented in the paper. In particular, our results imply that we can compute an approximation when the number of outliers is fixed, while the geometric measure is approximated. Using ε-sampling implies a result where the number of outliers is approximated. This is acceptable only when the number of outliers is relatively large.

Another limitation of our technique (which is inherent to almost all the work about outliers [DHS01]) is that the number of outliers, k, has to be specified in the input. In practice, there are situations when this information is not known beforehand. How to formulate a reasonable problem without k being the input, and then how to solve it, are among future directions of future research.

Finally, note that, for $k = 0$, this paper provides an alternative construction to the techniques of [AHV03]. The approximation techniques presented here are simpler to implement but the resulting coresets (not surprisingly) are larger.

Acknowledgment

The authors would like to thank the anonymous referees for their useful comments.

References

[AAHS00] 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.

[AES99] P. K. Agarwal, A. Efrat, and M. Sharir. Vertical decomposition of shallow levels in 3-dimensional arrangements and its applications. *SIAM J. Comput.*, 29:912–953, 1999.

[AH01] P. K. Agarwal and S. Har-Peled. Maintaining the approximate extent measures of moving points. In *Proc. 12th ACM-SIAM Sympos. Discrete Algorithms*, pages 148–157, 2001.

[AHV03] P. K. Agarwal, S. Har-Peled, and K. R. Varadarajan. Approximating extent measures of points. *J. Assoc. Comput. Mach.*, 2003. to appear.

[AM94] P. K. Agarwal and J. Matoušek. On range searching with semialgebraic sets. *Discrete Comput. Geom.*, 11:393–418, 1994.

[AS00] N. Alon and J. H. Spencer. *The probabilistic method*. Wiley Inter-Science, 2nd edition, 2000.

[Cha00] B. Chazelle. *The Discrepancy Method*. Cambridge University Press, 2000.

[Cha02] T. M. Chan. Low-dimensional linear programming with violations. In *Proc. 43th Annu. IEEE Sympos. Found. Comput. Sci.*, pages 570–579, 2002.

[CKMN01] 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.

[CS89] K. L. Clarkson and P. W. Shor. Applications of random sampling in computational geometry, II. *Discrete Comput. Geom.*, 4:387–421, 1989.

[CS00] N. Cristianini and J. Shaw-Taylor. *Support Vector Machines*. Cambridge Press, 2000.

[DHS01] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification*. Wiley-Interscience, New York, 2nd edition, 2001.

[DLSS95] A. Datta, H.-P. Lenhof, C. Schwarz, and M. Smid. Static and dynamic algorithms for k-point clustering problems. *J. Algorithms*, 19:474–503, 1995.

[DV01] J. Dunagan and S. Vempala. Optimal outlier removal in high-dimensional spaces. In *Proc. 33rd Annu. ACM Sympos. Theory Comput.*, pages 627–636, 2001.

[EE94] D. Eppstein and J. Erickson. Iterated nearest neighbors and finding minimal polytopes. *Discrete Comput. Geom.*, 11:321–350, 1994.

[Eri99] J. Erickson. New lower bounds for convex hull problems in odd dimensions. *SIAM J. Comput.*, 28:1198–1214, 1999.

[ES95] J. Erickson and R. Seidel. Better lower bounds on detecting affine and spherical degeneracies. *Discrete Comput. Geom.*, 13:41–57, 1995.

[GG02] J. Gao and L. J. Guibas. Staying in the middle: approximate medians in $\mathbb{R}^1$ and $\mathbb{R}^2$ for moving points. menuscript, 2002.

[HV01] S. Har-Peled and K. R. Varadarajan. Approximate shape fitting via linearization. In *Proc. 42nd Annu. IEEE Sympos. Found. Comput. Sci.*, pages 66–73, 2001.

[HW02] S. Har-Peled and Y. Wang. Shape fitting with outliers. `http://ww.uiuc.edu/~sariel/research/papers/02/outliers/`, 2002.

[Mat95a] J. Matoušek. On enclosing k points by a circle. *Inform. Process. Lett.*, 53:217–221, 1995.

[Mat95b] J. Matoušek. On geometric optimization with few violated constraints. *Discrete Comput. Geom.*, 14:365–384, 1995.

[Mul91] K. Mulmuley. On levels in arrangements and Voronoi diagrams. *Discrete Comput. Geom.*, 6:307–338, 1991.