No Coreset, No Cry*

Sariel Har-Peled†

February 19, 2010

Abstract

We show that coresets do not exist for the problem of 2-slabs in $\mathbb{R}^3$, thus demonstrating that the natural approach for solving approximately this problem efficiently is infeasible. On the positive side, for a point set P in $\mathbb{R}^3$, we describe a near linear time algorithm for computing a $(1+\varepsilon)$-approximation to the minimum width 2-slab cover of P. This is a first step in providing an efficient approximation algorithm for the problem of covering a point set with k-slabs.

1 Introduction

Geometric optimization in low dimensions is an important problem in computational geometry [AS98]. One of the central problems is to compute the shape best fitting a given point set, where the shape is restricted to belong to a certain family of shapes parameterized by a few parameters, while minimizing a certain quantity of the shape. For example, covering a point set P with minimum width slab, where a *slab* is the region enclosed between two parallel hyperplanes and the width of the slab is the distance between the two hyperplanes (i.e., this is equivalent to computing the width of P). Problems falling under this framework include computing the width and diameter of the point set, covering a point set with minimum volume bounding box, covering with minimum volume ellipsoid, covering with minimum width annulus, and a lot of other problems.

While some of those problems have exact fast solution, at least in low dimension, most of them can be solved only with algorithms that have running time exponential in the number of parameters defining the shape. For example, the fastest algorithm for computing the minimum width slab that covers a point set in $\mathbb{R}^d$ runs in $n^{O(d)}$ time.

It is thus natural to look for an efficient approximation algorithms for those problems. Here, one specifies an approximation parameter $\varepsilon > 0$, and one wish to find a shape which is $(1+\varepsilon)$-approximation to the optimal shape, see [BE97]. For the 1-slab width problem, we wish to find a slab $\mathcal{S}$ that covers a point set P, such that $\mathrm{width}(\mathcal{S}) \leq (1+\varepsilon)\mathrm{width}_{opt}(P, 1)$, where $\mathrm{width}_{opt}(P, 1)$ is the minimum width of a slab covering P.

---

*The full version of this paper is available from http://www.uiuc.edu/~sariel/papers/02/2slab/.

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

In recent years, there was a lot research done on those and similar problems (see [AAS97, AS96, AST94, EFNN89, EGS86, GLRS98, LL91, MSY97, PS85, Riv79, RLW91, RZ92, SY95, SJ99, YC97, Cha02, AAHS00] and references therein) and currently most of them can be solved in $O(n + 1/\varepsilon^c)$ time [AHV04], where c is a constant that depends on the problem at hand.

The problem becomes notably harder when one wish to perform clustering of the point set. Namely, cover the point set by k shapes (k is an integer constant larger than one), while simultaneously minimizing a global parameter of those shapes. For example, the k-slab problem ask to cover the point set by k slabs $(\S_1, \ldots, \S_k)$ of minimum width, where the width of $(\S_1, \ldots, \S_k)$ is $\max_{i=1}^{k} \mathrm{width}(\S_i)$. Problems falling under this category with efficient approximation algorithms include the k-center problem [Gon85, AP02], and the k-cylinder problem [APV02] (here we want to cover the point set with k cylinders of minimum radius).

It is known that in high dimensions, the exact problem is NP-Complete [Meg90] even for covering the point set by a single cylinder, and there is no FPTAS for this problem [HV04]. For the problem of covering the point set with k cylinders of minimum maximum radius, it is known that the problem is NP-Complete even in three dimensions if k is part of the input [MT82], and it can not be approximated in polynomial time unless $P = NP$. The currently fastest approximation algorithm known in high dimensions [BC03, HV02], runs in $d \cdot n^{O(\mathrm{poly}(k, 1/\varepsilon))}$ time, where $\mathrm{poly}(\cdot)$ is a polynomial of constant degree independent of the dimension. On the other hand, if we are interested in finding a single minimum radius m-flat covering the given point set, this can be ε-approximated in $O(nd)$ time [HV04] (here the constant in the O depends exponentially on $1/\varepsilon$), where an m-flat is a m-dimensional affine subspace.

In the *k-slab problem*, we wish to cover the point set with k slabs of minimum width (i.e., we wish to find k affine subspaces each of dimension $d-1$, such that the maximum distance of any point of the input to its closest subspace is minimized). At first, this problem might look somewhat artificial. However, it is related to projective clustering in high dimensions. In the projective clustering problem, we are looking for a cover of the points by k m-flats that have small radius. Such a projective clustering implies that the point set can be indexed as k point sets each of them being only m-dimensional. This is a considerable saving when the dimension is very large, as most efficient indexing structures have exponential dependency on the dimension. Thus, finding such a cover might result in a substantial performance improvements for various database applications. Furthermore, a lot of other clustering problems, that currently it is unknown how to solve them efficiently, can be reduced to this problem or its dual. In particular, problems of covering a point set with: (i) k rings with minimum max width, (ii) k bounding boxes of minimum maximum volume, (iii) k cylindrical shells of minimum maximum radius, and others all fall into this framework (see [AHV04] for the details of the reduction of those problems into this problem).

This and the more general problem of finding efficient approximation algorithm for the minimum radius k m-flat problem are still relatively open (both in low and high dimensions). The most natural approach to attack this problem is to try and use coresets. Those are small subsets of the input, such that if we solve the problem on those subsets, this yield an efficient solution for the original problem. Coresets had recently proved to be very useful in solving several clustering problems, both in low dimensions [AHV04, APV02], and high dimensions [BHI02, BC03, KMY03, HV04]. The surprising facts known about coresets is that their size is sometime dimension independent [BHI02], and they are small even in the presence of outliers [HW04]. (Note however, that the notion of coresets in high-dimensions is slightly weaker.) Interestingly, in low-dimensions, the existence of coresets immediately implies an efficient approximation algorithm for the problem at hand [APV02].

Surprisingly, we show in Section 2, that there is no such coresets for the 2-slab problem, in the worst case. Namely, any subset of the input that provides a good estimate of the width of the coverage of the points by two slabs, for all possible 2-slabs, must contain (almost) the whole point set. Thus, showing that solving this problem efficiently would require a different approach.

On the positive size, in Section 3, we make a first step in the direction of finding an efficient algorithm for this problem, and solve a special case which still has no efficient approximation algorithm and does not have a coreset; namely, the 2-slab problem in three dimensions. Formally, given a set P of n points in three dimensions, and a parameter $\varepsilon > 0$, the algorithm compute, in near linear time, a cover of P by two slabs, where the width of the solution is at most $(1+\varepsilon)\mathrm{width}_{opt}(P, 2)$, where $\mathrm{width}_{opt}(P, 2)$ denotes the width of the optimal (i.e., minimum) cover of P by two slabs.

A natural application of our algorithm is for edge detection in surface reconstruction. Indeed, given a set of points in $\mathbb{R}^3$ sampled from a region of a model that corresponds to an edge, the edge can be detected by best fitting the given points with two planes, that corresponds to the two faces adjacent to the edge. One need to slightly modify our algorithm to deploy it for this case.

The paper is organized as follows: In Section 2 we precisely define the notion of additive and multiplicative coresets, and prove that no small coresets exists for the 2-slab problem. As an additional result, we show that there exists a multiplicative coreset for the k-center problem. In Section 3, we present a a near linear time approximation algorithm for the 2-slab problem in three dimensions. Concluding remarks are given in Section 4.

2 When coresets Do Not Exists

2.1 Definitions

Definition 2.1 Given a set P of n objects (usually points) in $\mathbb{R}^d$, we are interested in the shape fitting problem, of finding the best shape that belongs to a certain family of shapes $\mathcal{F}$ that matches P. For example, the smallest ball that encloses the points of P. The *price function* $\mathrm{radius}(P)$, which returns the radius of the smallest ball enclosing P, measures the quality of this fitting.

In the clustering problem, we are provided with an additional parameter k, and we are interested in finding the best clustering of P into k clusters. Namely, we would like to partition P into k sets, such that the overall price of the clustering is minimized. Formally, we are interested in minimizing

$$\mathrm{rd}_k(P) = \min_{(P_1,\ldots,P_k)\in\mathcal{PW}(P,k)} \max_i \mathrm{radius}(P_i),$$

where $\mathcal{PW}(P,k) = \left\{(P_1,\ldots,P_k) \,\middle|\, \cup_i P_i = P, P_i \cap P_j = \emptyset, \text{ for } i \neq j\right\}$ is the set of all partitions of P into k sets.

For example, we would like to cover P by k balls, such that the radius of maximum radius ball is minimized. This is known as the *k-center clustering* problem (or just *k-center*). The *price function*, in this case, $\mathrm{rd}_k(P)$ is the radius of the maximum radius ball in the optimal solution.

Definition 2.2 Let P be a point set in $\mathbb{R}^d$, $1/2 > \varepsilon > 0$ a parameter.

For a cluster c, let $c(\delta)$ denote the cluster resulting form expanding c by δ. Thus, if c is a ball of radius r, then $c(\delta)$ is a ball of radius $r + \delta$. For a set $\mathcal{C}$ of clusters, let

$$\mathcal{C}(\delta) = \left\{ c(\delta) \,\middle|\, c \in \mathcal{C} \right\},$$

be the *additive expansion operator*; that is, $\mathcal{C}(\delta)$ is a set of clusters resulting form expanding each cluster of $\mathcal{C}$ by δ.

Similarly,

$$(1+\varepsilon)\mathcal{C} = \left\{ (1+\varepsilon)c \,\middle|\, c \in \mathcal{C} \right\},$$

is the *multiplicative expansion operator*, where $(1+\varepsilon)c$ is the cluster resulting from expanding c by a factor of $(1+\varepsilon)$. Namely, if $\mathcal{C}$ is a set of balls, then $(1+\varepsilon)\mathcal{C}$ is a set of balls, where a ball $c \in \mathcal{C}$, corresponds to a ball radius $(1+\varepsilon)\,\mathrm{radius}(c)$ in $(1+\varepsilon)\mathcal{C}$.

A set $Q \subseteq P$ is an (additive) *ε-coreset* of P, in relation to a price function radius, if for any clustering $\mathcal{C}$ of Q, we have that P is covered by $\mathcal{C}(\varepsilon\,\mathrm{radius}(\mathcal{C}))$, where $\mathrm{radius}(C) = \max_{c\in\mathcal{C}} \mathrm{radius}(c)$. Namely, we expand every cluster in the clustering by an ε-fraction of the size of the *largest* cluster in the clustering. Thus, if $\mathcal{C}$ is a set of k balls, then $\mathcal{C}(\varepsilon f(\mathcal{C}))$ is just the set of balls resulting from expanding each ball by εr, where r is the radius of the largest ball.

A set $Q \subseteq P$ is a *multiplicative ε-coreset* of P, if for any clustering $\mathcal{C}$ of Q, we have that P is covered by $(1+\varepsilon)\mathcal{C}$.

Note, that ε-multiplicative coresets are by definition also ε-additive coresets.

Remark 2.3 Let $\mathcal{C}$ be a given clustering, and apply to it a constant length sequence of δ-expansion operations, either additive or multiplicative. Let $\mathcal{C}'$ be the resulting clustering. It is easy to verify that there exists a constant c, such that the clustering $\mathcal{D}$ resulting from $c\delta$-additive expansion of $\mathcal{C}$, is larger than $\mathcal{C}'$. Namely, all the clusters of $\mathcal{C}'$ are contained inside the corresponding clusters of $\mathcal{D}$.

Thus, we can simulate any sequence of expansion operations, by a single additive expansion.

2.2 On multiplicative coresets

2.2.1 Coresets for k-center clustering

The k-center problem, is NP-Complete, and can be approximated up to a factor of two in linear time [Har04]. For a set P of n points in $\mathbb{R}^d$, let $\mathrm{rd}(P, k)$ denote the radius of the optimal clustering. This is the minimum over all covering of P by k balls, of the largest ball in the covering set.

Lemma 2.4 *Let P be a set of n points in $\mathbb{R}^d$, and $\varepsilon > 0$ a parameter. There exists an additive ε-coreset for the k-center problem, and this coreset has $O(k/\varepsilon^d)$ points.*

Proof: Let $\mathcal{C}$ denote the optimal clustering of P. Cover each ball of $\mathcal{C}$ by a grid of side length $\varepsilon r_{opt}/d$, where r_{opt} is the radius of the optimal k-center clustering of P. From each such grid cell, pick one points of P. Clearly, the resulting point set Q is of size $O(k/\varepsilon^d)$ and it is an additive coreset of P. ∎

The following is a minor extension of an argument used in [APV02].

Lemma 2.5 *Let P be a set of n points in $\mathbb{R}^d$, and $\varepsilon > 0$ a parameter. There exists a multiplicative ε-coreset for the k-center problem, and this coreset has $O\left(k!/\varepsilon^{dk}\right)$ points.*

Proof: For $k = 1$, the additive coreset of P is also a multiplicative coreset, and it is of size $O(1/\varepsilon^d)$.

As in the proof of Lemma 2.4, we cover the point set by a grid of radius $\varepsilon r_{opt}/(5d)$, let SQ the set of cells (i.e., cubes) of this grid which contains points of P. Clearly, $|\mathrm{SQ}| = O(k/\varepsilon^d)$.

Let Q be the additive ε-coreset of P. Let $\mathcal{C}$ be any k-center clustering of Q, and let Δ be any cell of SQ.

If Δ intersects all the k balls of $\mathcal{C}$, then one of them must be of radius at least $(1 - \varepsilon/2)\mathrm{rd}(P, k)$. Let c be this ball. Clearly, when we expand c by a factor of $(1 + \varepsilon)$ it would completely cover Δ, and as such it would also cover all the points of $\Delta \cap P$.

Thus, we can assume that Δ intersects at most $k - 1$ balls of $\mathcal{C}$. As such, we can inductively compute an ε-multiplicative coreset of $P \cap \Delta$, for $k - 1$ balls. Let Q_Δ be this set, and let $\mathcal{Q} = Q \cup \bigcup_{\Delta \in \mathrm{SQ}} Q_\Delta$.

Note that $|\mathcal{Q}| = T(k, \varepsilon) = O(k/\varepsilon^d)T(k - 1, \varepsilon) + O(k/\varepsilon^d) = O\left(k!/\varepsilon^{dk}\right)$. The set $\mathcal{Q}$ is the required multiplicative coreset by the above argumentation. ∎

2.2.2 When multiplicative coresets do not exist

Recently, Agarwal *et al.* [APV02] proved that an additive ε-coreset exists for the problem of covering a point set by k-cylinders. We next show that there is no small *multiplicative* coreset for this problem. Note, that for a strip c, the set $c(\delta)$ is the strip with the same center line as c, and of width $w + 2\delta$, where w is the width of c, and similarly, $(1 + \varepsilon)c$ is the strip of width $(1 + \varepsilon)w$ with the same center line as c.

Lemma 2.6 *There exists a point set P in $\mathbb{R}^2$, such that any multiplicative $(1/2)$-coreset of P, must be of size at least $|P| - 2$. Here the coreset is for the problem of covering the point set with 2 strips, such that the width of the wider strip is minimized.*

Proof: Consider the point set $P(m) = \left\{ (1/2^j, 2^j) \;\middle|\; j = 1, \ldots, m \right\}$, where m is an arbitrary parameter. Let Q be a $(1/2)$-coreset of $P = P(n)$.

Let $Q_i^- = Q \cap P(i)$ and $Q_i^+ = Q \setminus Q_i^-$.

If the set Q does not contain the point $p(i) = (1/2^i, 2^i)$, then Q_i^- can be covered by a horizontal strip h^- of width $\leq 2^{i-1}$ that has the x-axis as its lower boundary, and clearly if we expand h^- by a factor of $3/2$, the new $(3/2)h^-$ still will not cover $p(i)$. Similarly, we can cover Q_i^+ by a vertical strip h^+ of width $1/2^{i+1}$ that has the y-axis as its left boundary. Again, if we expand h^+ by a factor of 3/2, the new strip $(3/2)h^+$ will not cover $p(i)$. We conclude, that any multiplicative (1/2)-coreset for P must include all the points $p(2), p(3), \ldots, p(n-1)$.

Thus, no small multiplicative coreset exists for the problem of covering a point set by strips. ■

2.2.3 When small additive coresets do not exist

Definition 2.7 A *slab* $\mathcal{S}$ in $\mathbb{R}^3$ is the close region enclosed between two parallel planes. The *width* of $\mathcal{S}$ is the distance between those two parallel planes. The plane parallel to the two boundary planes and with equal distance to both of them, is the *center* of $\mathcal{S}$.

Definition 2.8 Given a tuple $\Delta = (\S_1, \ldots, \S_k)$ of k slabs in three dimensions, the width of Δ; denoted by $\text{width}(\Delta) = \text{width}(S_1, \ldots, \S_k) = \max_{i=1}^{k} \text{width}(\S_i)$. The *$k$-slab width* of a point set P in $\mathbb{R}^3$, is the width of the set of k slabs that covers P and minimizes the k-slab width. We denote the this minimum width by $\text{width}_{opt}(P, k)$.

In the following, let P be a set of points in $\mathbb{R}^3$, which we want to cover by k-slabs of minimum width.

Lemma 2.9 *There exists a point set P in $\mathbb{R}^3$, such that any additive $(1/2)$-coreset of P, for the 2-slab problem, must be of size at least $|P| - 2$.*

Proof: Let P' be the two-dimensional point set used in Lemma 2.6. Let P be the three-dimensional point set resulting from interpreting any point of P' as a point lying on the plane $z = 0$. Let $\delta = 1/2$.

Let $\mathcal{C}$ be an additive δ-coreset $\mathcal{C}$ of P, for the 2-slab problem, and let $\mathcal{C}'$ be the corresponding subset of P'. We claim that $\mathcal{C}'$ is a multiplicative δ-coreset of P' for the problem of 2-strips cover.

Indeed, let $\S_1, \S_2$ be any two strips that covers $\mathcal{C}'$, and assume that $\S_1$ is wider than $\S_2$. And let $\mathcal{T}_1, \mathcal{T}_2$ be two slabs in three dimensions, *of the same width*, such that $\S_1 = \mathcal{T}_1 \cap h$, and $\S_2 = \mathcal{T}_2 \cap h$, where h is the plane $z = 0$. The existence of such two slabs of equal width can be easily proved, by starting from two slabs $\mathcal{T}_1', \mathcal{T}_2$ which are perpendicular to h, and their intersection with h form $\S_1$ and $\S_2$, respectively. Next, rotate the wider slab, $\mathcal{T}_1'$, such that its width go down, while keeping its intersection with h fixed. We stop as soon as the modified $\mathcal{T}_1'$ has equal width to $\mathcal{T}_2$.

Now, additively expanding $\mathcal{T}_1$ and $\mathcal{T}_2$ corresponds to multiplicatively expanding them, since both of them have the same width. By assumption, the δ-expanded slabs cover P, and as such, they cover P'. Let $\mathcal{V}_1, \mathcal{V}_2$ be the intersection of the expanded slabs with h. Clearly, $\mathcal{V}_1$ and $\mathcal{V}_2$ are just the multiplicative expansion of $\S_1$ and $\S_2$, respectively, by a factor of $(1 + \delta)$. Furthermore, $\mathcal{V}_1$ and $\mathcal{V}_2$ covers the points of P'. We conclude, that $\mathcal{C}'$ is a δ-multiplicative coreset of P', for the 2-strip problem.

However, since $\delta = 1/2$, and by Lemma 2.6, it follows that $|\mathcal{C}| = |\mathcal{C}'| \geq |P| - 2$. ■

Note, that Lemma 2.9 implies that there is no hope of solving the k-slab problem using coreset techniques. Furthermore, it is easy to verify that the above lemma implies (in the dual) that one can not use coresets for the "dual" (and more useful) problem of stabbing hyperplanes with segments.

Problem 2.10 (k-extent) Given a set $\mathcal{H}$ of n hyperplanes in $\mathbb{R}^d$, and a parameter k, find a set of k vertical segments that stabs all the hyperplanes of $\mathcal{H}$, and the length of the longest segment is minimized.

Lemma 2.11 *There exists a set $\mathcal{H}$ of planes in $\mathbb{R}^3$, such that any additive $(1/2)$-coreset of $\mathcal{H}$, for the 2-extent problem, must be of size at least $|P| - 2$.*

3 Algorithm for the 2-Slab in 3D

Our algorithm works, by deploying a decision procedure that decides (approximately) whether or not the point set can be covered by two slabs of width r. This is done by performing the decision when two points are specified which lie on the center plane of one of the slabs (Lemma 3.4). Then, we extend this algorithm, for the decision problem when two points are given, which are "faraway" and lie in the same slab in the optimal solution (Lemma 3.7). This yields the required decision procedure, by enumerating a small number of pairs, one of them guaranteed to be in the same slab in the optimal solution, and to be a long pair (see Lemma 3.8).

Having this decision procedure, we can now solve the problem using a binary search over the possible widths. Naively, this leads to a weakly polynomial algorithm. To improve this, we first show that if we know a pair of points that lie on the center plane of one of the slabs, then we can compute a constant factor approximation in near linear time (Lemma 3.9). Now, we again generate a small set of candidate pairs, and check for each one of them what solution it yields. This results in a constant factor approximation (Lemma 3.10). Combining this constant factor approximation, together with the binary search using the decision procedure, results in the required approximation algorithm (Theorem 3.11).

In describing the algorithm, we use binary search and random sampling to replace parametric search, since it is simpler to describe. Minor improvements in running time are probably possible by using parametric search, and a more careful implementation of the algorithm.

3.1 Preliminaries

Lemma 3.1 ([AHV04]) *Given a set P of n points in $\mathbb{R}^3$, and a parameter $\varepsilon > 0$, one can compute a subset $S \subseteq P$, such that $\mathrm{width}_{opt}(S, 1) \geq (1-\varepsilon)\mathrm{width}_{opt}(P, 1)$, and $|S| = O(1/\varepsilon^2)$. Furthermore, S can be computed in $O(n + 1/\varepsilon^2)$ time.*

Lemma 3.2 ([AHV04]) *Given a set P of n points in $\mathbb{R}^3$, one can maintain an ε-approximate minimum width slab that contains P in $O((\log^3 n)/\varepsilon^2 + 1/\varepsilon^6)$ time per insertion/deletion. After each insertion/deletion the data-structure outputs a slab covering P which is wider by at most a factor of $(1 + \varepsilon)$ than the optimal slab that covers P.*

Fact 3.3 *Given a set P of n points in $\mathbb{R}^3$, one can compute the width of P in $O(n^{3/2+\varepsilon})$ expected time [AS96]. To simplify the exposition, we would use the slower quadratic time algorithm of [HT88].*

3.2 Decision Problem

We are given a set P of n points in $\mathbb{R}^3$, a candidate width r, and a parameter $\varepsilon > 0$. In this section, we describe an algorithm that decides whether one can cover P by two slabs of width at most r. More precisely, we describe an approximate decision procedure for this problem. Namely, if there is a cover of width $\leq r$, it outputs a cover of width at most $\leq (1+\varepsilon)r$. If P can not be covered by two slabs of width $(1+\varepsilon)r$, it outputs that no such cover exists. Otherwise (in this case the point set can not be covered with slabs of width r, but it can be covered by slabs of width $(1+\varepsilon)r$), it output either of those two answers.

Lemma 3.4 *Given r, ε prescribed parameters, a set P of n points in $\mathbb{R}^3$, and a pair $p, q \in P$. Then one can compute a cover $(\mathcal{S}, \mathcal{S}')$ of P by two slabs, such that p, q lie on the center (plane) of $\mathcal{S}$, the width of $\mathcal{S}$ is r, and the width of $\mathcal{S}'$ is $\leq (1+\varepsilon)\rho$, where ρ is the minimum width of $\mathcal{S}'$ under those constraints. The running time of the algorithm is $O\big(n(\log^3 n/\varepsilon^2 + 1/\varepsilon^6)\big)$.*

Proof: We assume that p, q lie on the x-axis. Let $\S(\alpha)$ be the slab of width r with its center plane passing through p, q and this plane has an angle α with the positive direction of the z-axis. Let $P_{in}(\alpha) = \S(\alpha) \cap P$ denote the points of P covered by $\S(\alpha)$, and let $P_{out}(\alpha) = P \setminus P_{in}(\alpha)$ denote the points not covered by it.

Using standard sweeping techniques, we can maintain the sets $P_{in}(\alpha), P_{out}(\alpha)$, for $0 \leq \alpha \leq \pi$. This would require $O(n)$ insertion/deletion operations, and would take $O(n \log n)$ time. Thus, to find the required cover, we need to compute the minimum width cover of $P_{out}(\alpha)$, for $0 \leq \alpha \leq \pi$, as the set $P_{in}(\alpha)$ is always covered by a slab of width r (i.e., $\S(\alpha)$). Using Lemma 3.2, this can be done approximately in $O\big(n(\log^3 n/\varepsilon^2 + 1/\varepsilon^6)\big)$ time. Indeed, let $\mathcal{S}_{out}(\alpha)$ denote the minimum width slab that covers $P_{out}(\alpha)$, and $\rho = \min_\alpha \mathrm{width}(\mathcal{S}_{out}(\alpha))$. Overall, the algorithm computes an α^*, such that $\mathrm{width}(\mathcal{S}_{out}(\alpha^*)) \leq (1+\varepsilon)\rho$. ■

Corollary 3.5 *Given r, ε prescribed parameters, and a pair of points $p, q \in P$ such that $\|pq\| \geq \mathrm{diam}(P)/10$. Let $(\mathcal{S}, \mathcal{S}')$ be a 2-cover of P, such that $\mathrm{width}(\mathcal{S}, \mathcal{S}') = r$ and p, q lie on the center of $\mathcal{S}$. Then, one can compute a 2-cover of P of width at most $(1+\varepsilon)r$. The running time of the algorithm is $O\big(n\big(\log^3 n/\varepsilon^2 + 1/\varepsilon^6\big)\big)$.*

Definition 3.6 Given a sphere Φ in $\mathbb{R}^d$ of radius r, a δ-net of Φ, is a subset U of Φ, such that for any $x \in \Phi$, there exists a point $u \in U$ such that $\|xu\| \leq \delta$.

Given Φ and δ, one can compute a δ-net for Φ of cardinality $O\big((r/\delta)^{d-1}\big)$, in linear time in the size of the δ-net.

Lemma 3.7 *Given r, ε prescribed parameters, a point set P in $\mathbb{R}^3$, and two points $p, q \in P$ such that $\mathrm{width}_{opt}(P, 2) \leq r$, $\|pq\| \geq \mathrm{diam}(P)/10$ and p, q lie in the same slab in the optimal 2-slab cover of P. Then, one can compute a cover of P by two slabs of width $\leq (1+\varepsilon)r$ in $O\big(n\big(\log^3 n/\varepsilon^6 + 1/\varepsilon^{10}\big)\big)$ time.*

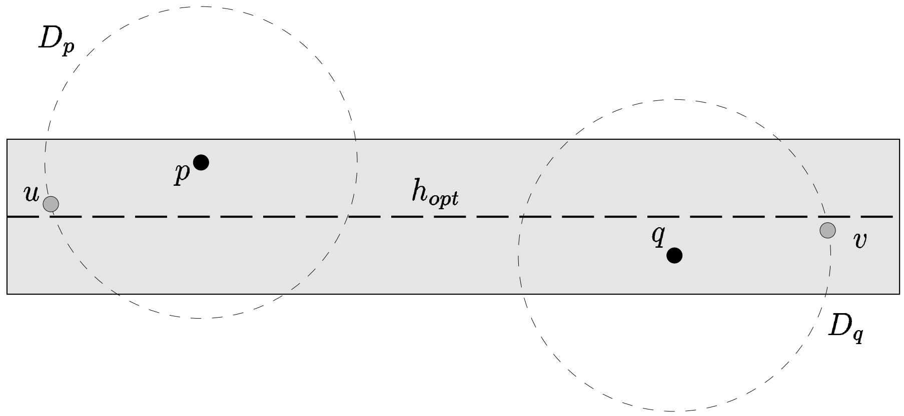

Figure 1: Demonstration of the proof of Lemma 3.7.

Proof: Let D_p, D_q be two spheres of radius $2r$ around p, q, respectively. Spread an $(r\varepsilon/320)$-net over D_p, D_q, and let N_p, N_q denote the resulting sets, respectively. Clearly, $|N_p|, |N_q| = O(1/\varepsilon^2)$. We claim that there must a pair of slabs $(\mathcal{S}, \mathcal{S}')$ that covers P, such that the center plane of $\mathcal{S}$ passes through a pair $p', q' \in N_p \cup N_q$, and $\mathrm{width}(\mathcal{S}, \mathcal{S}') \leq (1+\varepsilon/4)r$.

Indeed, consider the optimal solution $(\mathcal{S}_{opt}, \mathcal{S}'_{opt})$, and assume that p, q lie inside the slab $\mathcal{S}_{opt}$. Let h_{opt} be the center plane of $\mathcal{S}_{opt}$. Let u, v be the pair of points of $N_p \cup N_q$ which are furthest away from each other, and are in distance at most $(r\varepsilon/320)$ from h_{opt}. It is easy to verify that $\|uv\| \geq \mathrm{diam}(P)/10$. Let h_u be the plane resulting from translating h_{opt} so it passes through u.

We next rotate h_u so that it passes through v. Formally, let ℓ be the line in h_u that passes through u and is perpendicular to uv, and let h_{uv} be the plane passing through ℓ and v. We claim that $\mathrm{width}(P \cap \mathcal{S}_{opt}, h_{uv}) \leq (1+\varepsilon/2)r$. Observe that

$$\mathrm{dist}(v, h_u) \leq \mathrm{dist}(v, h_{opt}) + \frac{r\varepsilon}{320} \leq 2\frac{r\varepsilon}{320} = \frac{r\varepsilon}{160}.$$

Note, that the points of P are in distance at most $2r + \mathrm{diam}(P) \leq 3\mathrm{diam}(P)$ from ℓ. Thus, for any point $p \in P$ inside $\S_{opt}$, we have

$$\begin{aligned} \mathrm{dist}(p, h_{uv}) &\leq \mathrm{dist}(p, h_u) + \frac{3\mathrm{diam}(P)}{\|uv\|} \cdot \frac{r\varepsilon}{320} \leq \mathrm{dist}(p, h_{opt}) + \frac{r\varepsilon}{320} + \frac{r\varepsilon}{10} \\ &\leq r/2 + \frac{r\varepsilon}{320} + \frac{r\varepsilon}{10} \leq \frac{r}{2}\left(1 + \frac{\varepsilon}{320} + \frac{\varepsilon}{5}\right) \leq \frac{r}{2}\left(1 + \frac{\varepsilon}{3}\right), \end{aligned}$$

since $\|uv\| \geq \mathrm{diam}(P)/10$. Namely, the slab $\S_{uv}$ having h_{uv} for its center and of width $(1+\varepsilon/3)r$ covers $\mathcal{S}_{opt} \cap P$. In particular, $(\S_{uv}, \mathcal{S}'_{opt})$ is a 2-cover of P with two (far) points of $N_p \cup N_q$ lying on one of its slabs center.

This immediately yields an efficient algorithm for computing this cover. Indeed, for any pair of points $(a, b) \in (N_p \cup N_q) \times (N_p \cup N_q)$ of length $\geq \mathrm{diam}(P)/10$, use the algorithm of Corollary 3.5 to find a pair of slabs of width $(1+\varepsilon/3)(1+\varepsilon/3)r \leq (1+\varepsilon)r$ that covers P and the center of the slab passes through a, b, if such a cover exists. This takes $O\big(n(\log^3 n/\varepsilon^6 + 1/\varepsilon^{10})\big)$ time overall. ■

Lemma 3.8 *Given r, ε prescribed parameters, and a set P of n points in $\mathbb{R}^3$, such that* $\mathrm{width}_{opt}(P, 2) \le r$, *then one can compute a cover of P by two slabs of width $\le (1+\varepsilon)r$ in* $O\big(n\big(\log^3 n/\varepsilon^8 + 1/\varepsilon^{12}\big)\big)$ *time.*

Proof: Compute two points $p, q \in P$ such that $\|pq\| \ge \mathrm{diam}(P)/\sqrt{3}$ in linear time (by computing an axis parallel bounding box of P, and picking two points of P on the two parallel faces of the box furtherest away from each other). Next, let h be the bisector plane passing through the middle of the segment pq and perpendicular to it. Let $P_p = P \cap h_p, P_q = P \cap h_q$, where h_p, and h_q are the two half spaces defined by h containing p and q, respectively.

Let C_p, C_q be $\varepsilon/2$-coresets of size $O(1/\varepsilon^2)$ of P_p, P_q, respectively. By Lemma 3.1, we can compute them in $O(n+1/\varepsilon^2)$ time. Now, in the optimal solution $(\mathcal{S}, \mathcal{S}')$, either one of the pairs of $U = (C_p \times \{q\}) \cup (C_q \times \{p\})$ belong to the same slab, or alternatively $C_p \subseteq \mathcal{S}, C_q \subseteq \mathcal{S}'$.

For the first case, for each pair (a, b) of U (note, that such a pair has length at least $\mathrm{diam}(P)/(2\sqrt{3})$) use the algorithm of Lemma 3.7 to find a 2-slab cover of P with the pair (a, b) on the center of one of the slabs of width at most $(1+\varepsilon)r$. This takes $O\big(n(\log^3 n/\varepsilon^8 + 1/\varepsilon^{12})\big)$ time overall.

For the later case, where $C_p \subseteq \mathcal{S}, C_q \subseteq \mathcal{S}'$, we compute the minimum width slab $\S_p$ containing C_p and the minimum width slab $\S_q$ containing C_q. This takes $O(1/\varepsilon^4)$ time, by [HT88], as $|C_p|, |C_q| = O(1/\varepsilon^2)$. Expanding each slab by a factor of $(1+\varepsilon)$ results in a set of two slabs that covers P, as C_p, C_q are ε-coresets of P_p, P_q, respectively.

This had generated $O(1/\varepsilon^2)$ possible candidate solutions. We return the solution with the minimum width computed. ∎

3.3 A Strongly Polynomial Algorithm

The algorithm of Lemma 3.8 provides immediately a weakly polynomial algorithm that works by performing a binary search for the optimal width on the range $[0, \mathrm{diam}(P)]$. The resulting algorithm would execute the algorithm of Lemma 3.8 $O(\log(\mathrm{diam}(P)/(\mathrm{width}_{opt}(P, 2)\varepsilon)))$ times. In practice, this might be quite acceptable, although it is only weakly polynomial.

In this section, we present a *strongly* polynomial algorithm for approximating the 2-slab width. We observe, that it is enough to compute a c-approximation w to the 2-slab cover to P, where $c > 1$ is a constant. Indeed, once we have such an approximation, one can compute a better approximation using a binary search over the range $[w/c, w]$.

Lemma 3.9 *Let p, q be a pair of points of P. One can compute in $O\big(n \log^4 n\big)$ a 2-slab cover $(\mathcal{S}, \mathcal{S}')$ of P of width w, such that p, q lie on the center of $\mathcal{S}$, and w is constant factor approximation to the optimal 2-slab cover under this condition.*

Proof: Let $(\mathcal{S}_{opt}, \mathcal{S}'_{opt})$ be the optimal 2-slab cover of P, such that p, q lie on the center of $\mathcal{S}_{opt}$, such that $\mathcal{S}_{opt}$ is of minimal width (it might be that the wider slab of the two is $\mathcal{S}'_{opt}$, as such $\mathcal{S}_{opt}$ is not necessarily uniquely defined without this min-width requirement). Let w_{opt} denote the width of $\mathcal{S}_{opt}$.

Let ℓ be the line passing through p and q, and P' be the projection of P in the direction of ℓ into the two dimensional plane perpendicular to ℓ. In this plane, ℓ becomes a point, which we assume is the origin o, and $w_{opt}/2$ is no more than the distance of o to one of the lines of L, where L is the set of lines spanned by all the pairs of points of P'.

Let $\alpha_1, \ldots, \alpha_m$ be the sorted distance of the lines of L from the origin. Clearly, $\mathrm{width}_{opt}/2$ is one of the numbers in this sequence, Indeed, by the minimality requirement over $\mathcal{S}_{opt}$, it must have two points of P on the same plane of its boundary (otherwise, $\mathcal{S}_{opt}$ can be further minimized by slightly rotating it around pq), and the projection of those two points induce a line ℓ in L which is in distance $\mathrm{width}_{opt}/2$ from o.

Note, that we have an algorithm to solve the (approximate) decision problem, to decide if $w_{opt} \leq r$ or $w_{opt} \geq (1 + \varepsilon/3)r$, by using the algorithm of Lemma 3.4. Thus, we can use parametric search techniques (or more specifically, distance selection techniques) to find w, which is a constant factor approximation to w_{opt}.

Let $C(r)$ be the disk or radius r centered at the origin. We maintain for each point of P' its two tangents to $C(r)$. Clearly, as r increases, the locations of those tangents changes, but two tangents become identical (and change their ordering along the boundary of $C(r)$) only when $C(r)$ hits a line of L. Storing those tangents in a balanced binary tree we can handle a single such event in $O(\log n)$ time. Thus, we only need to jump-start the radius to the right range, and then we can perform "sweeping" to compute the range of critical distances in this range. Once we have those critical values, we can use our decision procedure to compute the optimal solution.

Specifically, randomly select $O(n \log n)$ lines of L (by randomly picking $O(n \log n)$ pairs of points of P'). For each such line, compute its distance from the origin, and use binary search together with the decision procedure, to figure out the range $I = [l, r]$ in which the optimal solution lies. By careful implementation, this requires $O(n \log^4 n)$ time (the cost of the $O(\log n)$ calls to the procedure of Lemma 3.4 is $O(n \log^4 n)$ when used with $\varepsilon = 0.1$).

Note, that the expected number of events in the range I (i.e., number of lines of L with distance between l and r to the origin) is $O(n^2/(n \log n)) = O(n/ \log n)$, and $O(n)$ with high probability. In particular, by performing sweeping, we can compute all those values in $O(n \log n)$ expected time. Finally, taking those values, we can find w (approximately) using $O(\log n)$ calls to the decision procedure of Lemma 3.4. ■

Lemma 3.10 *Given a set P of n points in $\mathbb{R}^3$, one can compute a cover of P by two slabs of width $O(\mathrm{width}_{opt}(P, 2))$ in $O\big(n \log^4 n\big)$ time.*

Proof: As in the proof of Lemma 3.8, we compute two points $p, q \in P$ which are $\mathrm{diam}(P) / \sqrt{3}$ furthest away from each other. And we compute an 0.1-coresets C_p, C_q of the sets P_p, P_q, where P_p and P_q are as in the proof of Lemma 3.8. Note that $|C_p|, |C_q| = O(1)$.

Let $(\S_p, \S_q)$ be the 2-slab cover resulting from computing the optimal slab cover of C_p, C_q, respectively, and expanding it by a constant factor. Either $(\S_p, \S_q)$ forms a constant factor approximation to the 2-slab width of P, or alternatively, at least one of the pair of points of $U = (C_p \times \{q\}) \cup (C_q \times \{p\})$ belong to the same slab of the optimal solution.

Thus, for each such pair $(a, b) \in U$, we would like to compute a constant factor approximate minimum 2-slab cover $(\overline{\mathcal{S}}, \overline{\mathcal{S}'})$ such that a, b is contained in $\overline{\mathcal{S}}$.

We claim that there exists a 2-slab cover $(\mathcal{S}, \mathcal{S}')$ of P such that a, b lie on the center of $\mathcal{S}$, and $\mathrm{width}(\mathcal{S}, \mathcal{S}') = O\big(\mathrm{width}\big(\overline{\mathcal{S}}, \overline{\mathcal{S}'}\big)\big)$. The claim follows, by observing that (a, b) is a long pair, and arguing as in the proof of Lemma 3.7.

Thus, we compute an approximate 2-slab cover of P, having a, b on one of its centers, in $O(n \log^4 n)$ using the algorithm of Lemma 3.9. Let w denote the width of $\mathcal{S}$.

It follows that the minimum such cover computed (over all pairs in U) would be a constant factor approximation to the optimal solution.

The running time bound follows, as we deploy the algorithm of Lemma 3.9 a constant number of times. ■

Now, once we have a constant factor approximation, an $(1+\varepsilon)$-approximation can be easily performed by doing a binary search, and using the decision procedure (Lemma 3.8). This would require $O(\log 1/\varepsilon)$ calls to the decision procedure. We conclude:

Theorem 3.11 *Given a set P of n points in $\mathbb{R}^3$, one can compute a cover of P by two slabs of width $\leq (1+\varepsilon)\mathrm{width}_{opt}(P,2)$ in $O\big(n \log^4 n + n(\log^3 n/\varepsilon^8 + 1/\varepsilon^{12}) \log 1/\varepsilon\big)$ time.*

4 Conclusions

In this paper, we showed that coresets do no exists for the problem of 2-slabs in three dimensions. The author find this fact to be quite bewildering, considering the fact that such coresets exists for balls and cylinders. This implies that solving the k-slab problem efficiently in low dimensions (i.e., in near linear time) would require developing new techniques and algorithms. We took a small and tentative step in this direction, providing a near linear time algorithm for approximating the min-width 2-slab cover of a point set in $\mathbb{R}^3$.

The main open question for further research, is to develop an efficient approximation algorithms for the k-slab problem in three and higher dimensions, for $k > 2$. Currently, the author is unaware of any efficient constant factor approximation algorithm for this problem.

Finally, there seems to be a connection between solving the problem of k-slabs in $\mathbb{R}^d$ in the presence of outliers, and solving the problem of $k+1$ slabs in $\mathbb{R}^{d+1}$. (Intuitively and imprecisely, thats what our 2-slab algorithm in $\mathbb{R}^3$ is doing: It reduces the problem into the problem of covering a point set in $\mathbb{R}^2$ with a single slab, while making sure that the points that are not covered can be ignored.) Understanding this connection might be a key in understanding why the k-slab problem seems to be harder than one might expect.

Acknowledgments

The author wishes to thank Vladlen Koltun, Magda Procopiuc, Edgar Ramos, and Kasturi Varadarajan for useful discussions on the problems studied in this paper. The author would also thank the anonymous referees for their insightful 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.

[AAS97] P. K. Agarwal, B. Aronov, and M. Sharir. Computing envelopes in four dimensions with applications. *SIAM J. Comput.*, 26:1714–1732, 1997.

[AHV04] P. K. Agarwal, S. Har-Peled, and K. R. Varadarajan. Approximating extent measures of points. *J. Assoc. Comput. Mach.*, 51(4):606–635, 2004.

[AP02] P. K. Agarwal and C. M. Procopiuc. Exact and approximation algorithms for clustering. *Algorithmica*, 33(2):201–226, 2002.

[APV02] P. K. Agarwal, C. M. Procopiuc, and K. R. Varadarajan. Approximation algorithms for k-line center. In *Proc. 10th Annu. European Sympos. Algorithms*, pages 54–63, 2002.

[AS96] P. K. Agarwal and M. Sharir. Efficient randomized algorithms for some geometric optimization problems. *Discrete Comput. Geom.*, 16:317–337, 1996.

[AS98] P. K. Agarwal and M. Sharir. Efficient algorithms for geometric optimization. *ACM Comput. Surv.*, 30:412–458, 1998.

[AST94] P. K. Agarwal, M. Sharir, and S. Toledo. Applications of parametric searching in geometric optimization. *J. Algorithms*, 17:292–318, 1994.

[BC03] M. Bădoiu and K. L. Clarkson. Optimal coresets for balls. `http://cm.bell-labs.com/who/clarkson/coresets2.pdf`, 2003.

[BE97] M. Bern and D. Eppstein. Approximation algorithms for geometric problems. In D. S. Hochbaum, editor, *Approximationg algorithms for NP-Hard problems*, pages 296–345. PWS Publishing Company, 1997.

[BHI02] M. Bădoiu, S. Har-Peled, and P. Indyk. Approximate clustering via coresets. In *Proc. 34th Annu. ACM Sympos. Theory Comput.*, pages 250–257, 2002.

[Cha02] T. M. Chan. Approximating the diameter, width, smallest enclosing cylinder and minimum-width annulus. *Internat. J. Comput. Geom. Appl.*, 12(2):67–85, 2002.

[EFNN89] H. Ebara, N. Fukuyama, H. Nakano, and Y. Nakanishi. Roundness algorithms using the Voronoi diagrams. In *Proc. 1rd Canad. Conf. Comput. Geom.*, page 41, 1989.

[EGS86] H. Edelsbrunner, L. J. Guibas, and J. Stolfi. Optimal point location in a monotone subdivision. *SIAM J. Comput.*, 15(2):317–340, 1986.

[GLRS98] J. García-Lopez, P. Ramos, and J. Snoeyink. Fitting a set of points by a circle. *Discrete Comput. Geom.*, 20(3):389–402, 1998.

[Gon85] T. Gonzalez. Clustering to minimize the maximum intercluster distance. *Theoret. Comput. Sci.*, 38:293–306, 1985.

[Har04] S. Har-Peled. Clustering motion. *Discrete Comput. Geom.*, 31(4):545–565, 2004.

[HT88] M. E. Houle and G. T. Toussaint. Computing the width of a set. *IEEE Trans. Pattern Anal. Mach. Intell.*, PAMI-10(5):761–765, 1988.

[HV02] S. Har-Peled and K. R. Varadarajan. Projective clustering in high dimensions using coresets. In *Proc. 18th Annu. ACM Sympos. Comput. Geom.*, pages 312–318, 2002.

[HV04] S. Har-Peled and K. R. Varadarajan. High-dimensional shape fitting in linear time. *Discrete Comput. Geom.*, 32(2):269–288, 2004.

[HW04] S. Har-Peled and Y. Wang. Shape fitting with outliers. *SIAM J. Comput.*, 33(2):269–285, 2004.

[KMY03] P. Kumar, J. S. B. Mitchell, and E. A. Yildirim. Approximate minimum enclosing balls in high dimensions using coresets. *J. Exp. Algorithmics*, 8:1.1, 2003.

[LL91] V. B. Le and D. T. Lee. Out-of-roundness problem revisited. *IEEE Trans. Pattern Anal. Mach. Intell.*, PAMI-13(3):217–223, 1991.

[Meg90] N. Megiddo. On the complexity of some geometric problems in unbounded dimension. *J. Symb. Comput.*, 10:327–334, 1990.

[MSY97] K. Mehlhorn, T. C. Shermer, and C. K. Yap. A complete roundness classification procedure. In *Proc. 13th Annu. ACM Sympos. Comput. Geom.*, pages 129–138, 1997.

[MT82] N. Megiddo and A. Tamir. On the complexity of locating linear facilities in the plane. *Oper. Res. Lett.*, 1:194–197, 1982.

[PS85] F. P. Preparata and M. I. Shamos. *Computational Geometry: An Introduction.* Springer-Verlag, New York, NY, 1985.

[Riv79] T. J. Rivlin. Approximating by circles. *Computing*, 21:93–104, 1979.

[RLW91] U. Roy, C. R. Liu, and T. C. Woo. Review of dimensioning and tolerancing: Representation and processing. *Comput. Aided Design*, 23(7):466–483, 1991.

[RZ92] U. Roy and X. Zhang. Establishment of a pair of concentric circles with the minimum radial separation for assessing roundness error. *Comput. Aided Design*, 24(3):161–168, 1992.

[SJ99] M. Smid and R. Janardan. On the width and roundness of a set of points in the plane. *Internat. J. Comput. Geom. Appl.*, 9(1):97–108, 1999.

[SY95] T. C. Shermer and C. K. Yap. Probing for near centers and relative roundness. In *Proc. ASME Workshop on Tolerancing and Metrology*, 1995.

[YC97] C. K. Yap and E.-C. Chang. Issues in the metrology of geometric tolerancing. In J.-P. Laumond and M. H. Overmars, editors, *Robotics Motion and Manipulation*, pages 393–400. A. K. Peters, 1997.