No Coreset, No Cry: II

Michael Edwards[1] and Kasturi Varadarajan[1]

Department of Computer Science
The University of Iowa
Iowa City, IA 52242-1419
Email: [mcedward,kvaradar]@cs.uiowa.edu

Abstract. Let P be a set of n points in d-dimensional Euclidean space, where each of the points has integer coordinates from the range $[-\Delta, \Delta]$, for some $\Delta > 1$. Let $\varepsilon > 0$ be a given parameter. We show that there is subset Q of P, whose size is polynomial in $(\log \Delta/\varepsilon)$, such that for any k slabs that cover Q, their ε-expansion covers P. In this result, k and d are assumed to be constants. The set Q can also be computed efficiently, in time that is roughly n times the bound on the size of Q. Besides yielding approximation algorithms that are linear in n and polynomial in $\log \Delta$ for the k-slab cover problem, this result also yields small coresets and efficient algorithms for several other clustering problems.

1 Introduction

A slab in $\Re^d$ is specified by a hyperplane h and a real number $r \geq 0$: $\text{Slab}(h, r)$ is the set of points at distance at most r from h. The *width* of such a slab is $2r$. Note that such a slab can be viewed as the set of points enclosed between two parallel hyperplanes at distance $2r$ apart. For an $\varepsilon \geq 0$, the ε-expansion of $\text{Slab}(h, r)$ is $\text{Slab}(h, r(1 + \varepsilon))$; note that its width is $2r(1 + \varepsilon)$.

For an integer $k \geq 1$ and a parameter $0 < \varepsilon < 1$, a (k, ε) (multiplicative) coreset of a point set $P \subseteq \Re^d$ is a subset $Q \subset P$ such that given any k slabs that cover Q, the ε-expansion of the k slabs covers P. (A set of k slabs is said to cover a point set if the point set is contained in the union of the k slabs.)

A $(1, \varepsilon)$ coreset for any set of n points $P \subseteq \Re^d$ of size $O(1/\varepsilon^{(d-1)/2})$ exists, and can be computed in $O(n)$ time [3, 4, 1, 5]. (We are ignoring constants in the running time that depend on ε. Throughout this paper, d will be treated as a constant.) Such a coreset immediately implies a linear time algorithm for computing an approximately minimum width slab enclosing P. Moreover, a $(1, \varepsilon)$ coreset automatically implies a small $(1, \varepsilon)$ coreset of other kinds, obtained essentially by replacing 'slab' in the definition above by 'ball', 'cylinder', 'spherical shell', 'cylindrical shell', etc. [1]. One immediately obtains linear-time approximation algorithms for various extent measure problems, such as finding the minimum-width slab, cylinder, spherical shell, cylindrical shell, etc. enclosing a point set. Furthermore, a small $(1, \varepsilon)$ coreset also yields small coresets corresponding to such extent measure problems for points with algebraic motion [1]. This pleasant state of affairs continues to persist if we want to handle a few outliers [8].

It is therefore natural to ask if small (k, ε) coresets exist, for $k \geq 2$. We are asking, informally, if the pleasant state of affairs for the one cluster case also holds for the multiple cluster case. Answering this question in the negative, Har-Peled [7] gave an example of a point set $P \subset \Re^2$ for which any $(2, 1/2)$ coreset has size at least $|P| - 2$. In other words, the coreset needs to contain nearly all the points. This is unfortunate, since a small coreset would yield small coresets for several clustering problems. (Nevertheless, small coresets exist for the k balls case [7], and small coresets of a weaker type exist for the k cylinders case [2]. These are the exceptions.)

Har-Peled's construction, when embedded on an integer grid, uses coordinates that are exponentially large in the number of input points. In this paper, we ask whether small coresets exist if the coordinates are reasonably small. The main result of this paper is the following theorem, which answers the question in the affirmative.

Theorem 1. *Let P be any set of n points in $\Re^d$, with the co-ordinates of each point in P being integers in the range $[-\Delta, \Delta]$, where $\Delta \geq 2$. For any integer $k \geq 1$, and $0 < \varepsilon < 1$, there is a (k, ε) coreset of P with at most $(\log \Delta/\varepsilon)^{f(d,k)}$ points, where $f(d, k)$ is a function of only d and k. Such a coreset can be constructed in $n(\log \Delta/\varepsilon)^{f(d,k)}$ time.*

We remark that k and d are treated as constants in the big-O notation.

Evidently, the theorem implies an algorithm whose running time is linear in n and polynomial in $\log \Delta$ (ignoring 'constants' that depend on ε, d, and k) for computing k slabs of width $(1 + \varepsilon)r^*$ that cover P, where r^* is the smallest number such that k slabs of width r^* cover P. (That is, r^* is the width of the optimal k-slab cover of P.) Such an algorithm is obtained by computing a (k, ε) coreset of P, computing an optimal k-slab cover for the coreset, and taking their ε-expansion. (An algorithm that is more efficient in terms of the hidden constants can be obtained, if really needed, by working through the proof of Theorem 1.)

The theorem also holds if we replace 'slab' in the definition of a (k, ε) coreset by an 'ℓ-cylinder', where an ℓ-cylinder is the set of points within a certain distance from an ℓ-dimensional flat (affine subspace of dimension ℓ). The proof readily carries over to this case. Consequently, we also obtain efficient algorithms for approximating the k-ℓ-cylinder cover of the point set P.

Other consequences for clustering follow from Theorem 1 using the machinery developed in Agarwal et al. [1]. We give two illustrative examples. An annulus in $\Re^2$ is the set of points between two concentric circles, and its width is the difference between the two radii. An ε-expansion of an annulus is defined accordingly. Let P be a set of points in $\Re^2$ with integer coordinates in the range $[-\Delta, \Delta]$, and let $k \geq 1$ be an integer and $0 < \varepsilon < 1$ be a parameter. We can compute, in linear time, a subset $Q \subset P$ of $(\log \Delta/\varepsilon)^{g(k)}$ points such that for any k annuli that cover Q, their ε-expansion covers P. Here, g is only a function of k.

The second example concerns moving points [6]. Let $P = \{p_1, \ldots, p_n\}$ be a set of points moving linearly in $\Re^2$, where the position of point p_i at time t is given by $p_i[t] = a_i + b_i t$, where $a_i, b_i \in \Re^2$ have integer coordinates in the range

$[-\Delta, \Delta]$. Let $P[t] = \{p_1[t], \ldots, p_n[t]\}$ denote the point set at time t. Let $k \geq 1$ be an integer and $0 < \varepsilon < 1$ be a parameter. We can compute, in linear time, a subset $Q \subset P$ of size $(\log \Delta/\varepsilon)^{g(k)}$ such that for any time t, and any k balls that cover $Q[t]$, their ε-expansion covers $P[t]$. These examples by no means exhaust the consequences. For instance, we can replace linear motion by quadratic motion and balls by slabs in the second example.

In summary, small coresets do exist for the multiple cluster case, provided we are willing to expand our definition of 'small' in a reasonable way.

Technique. The proof of Theorem 1 builds (k, ε) coresets from $(k-1, \varepsilon)$ coresets. The idea is to add to the coreset a subset Q' composed of $(k-1, \varepsilon)$ coresets of a small number of appropriately chosen subsets of P. The subset Q' will have the property that for any set of k slabs, the points of Q' contained in the k'th slab tell us approximately which subset of P is contained in the k'th slab. We are then left with the problem of adding a $(k-1, \varepsilon)$ coreset for the remainder of P. The cases where Q' fails to give us such meaningful information are precisely those where the k'th slab plays no essential role – the ε-expansion of the first $k-1$ slabs covers P. Crucial to the entire construction is an idea from [3], which says, in a technical sense that is convenient to us, that in order to know the shape of a cluster, it is sufficient to know its d principal dimensions.

Is bounded spread enough? We modify the construction of Har-Peled [7] to show that merely assuming bounded spread is not enough to obtain a coreset of the type obtained in Theorem 1. The point set is $P = \{p_1, \ldots, p_n\}$ in $\Re^3$, where $p_i = (1/2^{n-i}, 1/2^{i-1}, i-1)$. The spread of this point set, that is, the ratio of the maximum to minimum interpoint distance, is clearly $O(n)$. We claim that any $(2, 1/2)$ coreset for this point set must include each p_i, for $1 \leq i \leq n$, and must consequently have all the points. Suppose, to the contrary, that there is such a coreset without p_i. Then the slab $\mathrm{Slab}(h_1, 1/2^{n-(i-1)})$, where h_1 is the hyperplane $x = 0$, covers the points $p_1, \ldots, p_{i-1}$, and the slab $\mathrm{Slab}(h_2, 1/2^i)$, where h_2 is the hyperplane $y = 0$, covers the points $p_{i+1}, \ldots, p_n$. Therefore the two slabs cover the coreset points. But evidently a 1/2-expansion of these two slabs does not cover p_i, a contradiction.

In Section 2, we establish some geometrical facts needed in Section 3, where we prove Theorem 1. We omit from this preliminary version the proofs for the consequences of Theorem 1 claimed above. These consequences follow, with some care, via the arguments used for the one cluster case in [1].

2 Preliminaries

For any subset $V = \{v_1, \ldots, v_\ell\}$ of points in $\Re^d$, let

$$\mathrm{Aff}(V) = \{a_1 v_1 + \cdots + a_\ell v_\ell \mid a_1 + \cdots a_\ell = 1\}$$

be the *affine subspace* or *flat* spanned by them. If $\mathrm{Aff}(V)$ has dimension t, then it is called a t-flat.

Let $\mathrm{proj}(q, F)$ denote the closest point on flat F to a point q, and let $\mathrm{dist}(q, F)$ denote the diatance between q and $\mathrm{proj}(q, F)$.

For any subset $V = \{v_1, \dots, v_\ell\}$ of points in $\Re^d$, let

$$\text{conv}(V) = \{a_1 v_1 + \cdots + a_\ell v_\ell \mid a_1, \dots, a_\ell \geq 0, a_1 + \cdots a_\ell = 1\}$$

be the *convex hull* of V.

Let $\mathcal{D}$ denote the points in $\Re^d$ with integer co-ordinates in the range $[-\Delta, \Delta]$. The following proposition is well known.

Proposition 1 *There exists a constant $c_d > 0$, depending only on the dimension d, such that for any subset $V \subseteq \mathcal{D}$ and point $q \in \mathcal{D}$, $dist(q, Aff(V))$ is either 0 or a number in the range $[c_d/\Delta^d, 4d\Delta]$.*

Lemma 1. *There exists a constant c'_d, depending only on the dimension, for which the following is true. Let $v_0, \dots, v_t$ be any set of points, where $t \leq d$. For $1 \leq i \leq t$, let u_i denote the vector $v_i - proj(v_i, Aff(\{v_0, \dots, v_{i-1}\}))$, and suppose that $||u_i|| > 0$. Suppose that for every $i \geq 1$ and $j \geq i$, we have $dist(v_j, Aff(\{v_0, \dots, v_{i-1}\})) \leq 2||u_i||$. Then the t-simplex $conv(\{v_0, \dots, v_t\})$ contains a translate of the hyper-rectangle*

$$\{c'_d(a_1 u_1 + a_2 u_2 + \cdots + a_t u_t) | 0 \leq a_i \leq 1\}.$$

Proof. This is the central technical lemma that underlies the algorithm of Barequet and Har-Peled [3] for computing an approximate bounding box of a point set. For expository purposes, we sketch a proof. We may assume without loss of generality that v_0 is the origin, and $u_1, \dots, u_t$ are multiples of the first t unit vectors in the standard basis for $\Re^d$. Scale the first t axes so that $u_1, \dots, u_t$ map to unit vectors. The conditions of the lemma ensure that the images $v'_0, \dots, v'_t$ of $v_0, \dots, v_t$ lie in the " cube "

$$C = \{(x_1, \dots, x_d) | -2 \leq x_i \leq 2 \text{ for } i \leq t, x_i = 0 \text{ for } i > t\},$$

and the (t-dimensional) volume of $\text{conv}(\{v'_0, \dots, v'_t\})$ is at least $1/t!$, which is at least $\frac{1}{4^d d!}$ of the volume of C. It follows (see Lemma 3.5 of [3]) that there exists $c'_d > 0$, depending only on d, such that a translate of $c'_d C$ is contained in $\text{conv}(\{v'_0, \dots, v'_t\})$. Scaling back gives the required hyper-rectangle.

It is worth stating that under the conditions of Lemma 1, the set $\{v_0, \dots, v_t\}$ is contained in the hyperrectangle

$$v_0 + \{(a_1 u_1 + a_2 u_2 + \cdots + a_t u_t) | -2 \leq a_i \leq 2\}.$$

3 The Coreset Construction

In this section, we describe our algorithm for constructing a (k, ε) coreset for any given subset of $\mathcal{D}$, for $k \geq 2$. Our construction is inductive and will assume an algorithm for constructing a $(k-1, \varepsilon)$ coreset for any given subset of $\mathcal{D}$. As the base case, we know that a $(1, \varepsilon)$ coreset of size $O(1/\varepsilon^{d-1})$ for any subset $P' \subset \mathcal{D}$

can be constructed in $O(|P'|+1/\varepsilon^{d-1})$ time [1]. Let λ denote the smallest integer that is at least $\log_2 \frac{4d\Delta}{c_d/\Delta^d}$, where $c_d > 0$ is the constant in Proposition 1. Note that $\lambda = O(\log \Delta)$.

Let $P \subset \mathcal{D}$ be the point set for which we wish to construct a (k, ε) coreset. Our algorithm can we viewed as having $d+1$ levels. At level t, we do some work corresponding to each instantiation of the variables $v_0, \ldots, v_t$. Let Q denote the final coreset that the algorithm returns; Q is initialized to be the empty set.

We construct a $(k-1, \varepsilon)$ coreset K of the point set P and add K to Q. Each point in K is a choice for the variable v_0. For each choice of v_0 from K, we proceed to Level 0 with the point set $P[v_0] = P$.

Level 0: Suppose we have entered this level with $\{v_0\}$ and $P[v_0]$. We partition $P[v_0]$ into $\lambda+1$ buckets. The 0'th bucket $B_0[v_0]$ contains just v_0 and for $1 \le i \le \lambda$, the i'th bucket $B_i[v_0]$ contains all points $p \in P[v_0]$ such that $c_d 2^{i-1}/\Delta^d \le \mathrm{dist}(p, \mathrm{Aff}(\{v_0\})) < c_d 2^i/\Delta^d$. (Note that $\mathrm{Aff}(\{v_0\})$ simply consists of the point v_0.) By Proposition 1, we do indeed have a partition of $P[v_0]$. For each $1 \le i \le \lambda$, we construct a $(k-1, \varepsilon)$ coreset $K_i[v_0]$ of $B_i[v_0]$ and add $K_i[v_0]$ to Q.

Each point in $\bigcup_{i=1}^{\lambda} K_i[v_0]$ is a choice for v_1. If v_1 is chosen from $K_j[v_0]$, we enter Level 1 with $\{v_0, v_1\}$ and the corresponding set $P[v_0, v_1] = \bigcup_{i=0}^{j} B_i[v_0]$. Note that for any $p \in P[v_0, v_1]$, we have $\mathrm{dist}(p, \mathrm{Aff}(v_0)) \le 2\mathrm{dist}(v_1, \mathrm{Aff}(v_0))$.

Level 1: Suppose we have entered this level with $\{v_0, v_1\}$ and $P[v_0, v_1]$. We partition $P[v_0, v_1]$ into $\lambda+1$ buckets. The 0'th bucket $B_0[v_0, v_1]$ contains all the points of $P[v_0, v_1]$ that lie on $\mathrm{Aff}(\{v_0, v_1\})$. (Note that $\mathrm{Aff}(\{v_0, v_1\})$ is simply the line through v_0 and v_1.) For $1 \le i \le \lambda$, the i'th bucket $B_i[v_0, v_1]$ contains all points $p \in P[v_0, v_1]$ such that $c_d 2^{i-1}/\Delta^d \le \mathrm{dist}(p, \mathrm{Aff}(\{v_0, v_1\})) < c_d 2^i/\Delta^d$. By Proposition 1, we do indeed have a partition of $P[v_0, v_1]$.

Let $u_1 = v_1 - \mathrm{proj}(v_1, \mathrm{Aff}(\{v_0\}))$. Cover the "rectangle"

$$R[v_0, v_1] = v_0 + \{a_1 u_1 | -2 \le a_1 \le 2\}$$

by $O(1/\varepsilon)$ copies of translates of the scaled down rectangle

$$R'[v_0, v_1] = \{\frac{\varepsilon}{2} c'_d a_1 u_1 | 0 \le a_1 \le 1\}.$$

Here, $c'_d > 0$ is the constant in Lemma 1. Note that the bigger rectangle $R[v_0, v_1]$ lies on $\mathrm{Aff}(\{v_0, v_1\})$ and contains $B_0[v_0, v_1]$. For each of the $O(1/\varepsilon)$ copies of $R'[v_0, v_1]$, we compute a $(k-1, \varepsilon)$ coreset of the points of $B_0[v_0, v_1]$ contained in that copy, and add all these coreset points to Q.

For each $1 \le i \le \lambda$, we construct a $(k-1, \varepsilon)$ coreset $K_i[v_0, v_1]$ of $B_i[v_0, v_1]$ and add $K_i[v_0, v_1]$ to Q. Each point in $\bigcup_{i=1}^{\lambda} K_i[v_0, v_1]$ is a choice for v_2. If v_2 is chosen from $K_j[v_0, v_1]$, we enter Level 2 with $\{v_0, v_1, v_2\}$ and the corresponding set $P[v_0, v_1, v_2] = \bigcup_{i=0}^{j} B_i[v_0, v_1]$. Note that for any $p \in P[v_0, v_1, v_2]$, we have $\mathrm{dist}(p, \mathrm{Aff}(v_0, v_1)) \le 2\mathrm{dist}(v_2, \mathrm{Aff}(v_0, v_1))$.

Level t ($2 \le t < d$): Suppose we have entered this level with $\{v_0, \ldots, v_t\}$ and $P[v_0, \ldots, v_t]$. We partition $P[v_0, \ldots, v_t]$ into $\lambda+1$ buckets. The 0'th bucket $B_0[v_0, \ldots, v_t]$ contains all the points of $P[v_0, \ldots, v_t]$ that lie on $\mathrm{Aff}(\{v_0, \ldots, v_t\})$. For $1 \le i \le \lambda$, the i'th bucket $B_i[v_0, \ldots, v_t]$ contains all points $p \in P[v_0, \ldots, v_t]$

such that $c_d 2^{i-1}/\Delta^d \leq \mathrm{dist}(p, \mathrm{Aff}(\{v_0, \ldots, v_t\})) < c_d 2^i/\Delta^d$. By Proposition 1, we do indeed have a partition of $P[v_0, \ldots, v_t]$.

For $1 \leq i \leq t$, let u_i denote the vector $v_i - \mathrm{proj}(v_i, \mathrm{Aff}(\{v_0, \ldots, v_{i-1}\}))$. Cover the rectangle

$$R[v_0, \ldots, v_t] = v_0 + \{a_1 u_1 + a_2 u_2 + \cdots + a_t u_t | -2 \leq a_i \leq 2\}$$

by $O(1/\varepsilon^t)$ copies of translates of the scaled down rectangle

$$R'[v_0, \ldots, v_t] = \{\frac{\varepsilon}{2} c'_d (a_1 u_1 + a_2 u_2 + \cdots + a_t u_t) | 0 \leq a_i \leq 1\}.$$

Note that the bigger rectangle $R[v_0, \ldots, v_t]$ lies on $\mathrm{Aff}(\{v_0, \ldots, v_t\})$ and contains $B_0[v_0, \ldots, v_t]$. For each of the $O(1/\varepsilon^t)$ copies of $R'[v_0, \ldots, v_t]$, we compute a $(k-1, \varepsilon)$ coreset of the points of $B_0[v_0, \ldots, v_t]$ contained in that copy, and add all these coreset points to Q.

For each $1 \leq i \leq \lambda$, we construct a $(k-1, \varepsilon)$ coreset $K_i[v_0, \ldots, v_t]$ of $B_i[v_0, \ldots, v_t]$ and add $K_i[v_0, \ldots, v_t]$ to Q. Each point in $\bigcup_{i=1}^{\lambda} K_i[v_0, \ldots, v_t]$ is a choice for v_{t+1}. If v_{t+1} is chosen from $K_j[v_0, \ldots, v_t]$, we enter Level $t+1$ with $\{v_0, \ldots, v_t, v_{t+1}\}$ and the corresponding set $P[v_0, \ldots, v_t, v_{t+1}] = \bigcup_{i=0}^{j} B_i[v_0, \ldots, v_t]$. Note that for any $p \in P[v_0, \ldots, v_{t+1}]$, we have

$$\mathrm{dist}(p, \mathrm{Aff}(v_0, \ldots, v_t)) \leq 2\mathrm{dist}(v_{t+1}, \mathrm{Aff}(v_0, \ldots, v_t)).$$

Level d: Suppose we have entered this level with $\{v_0, \ldots, v_d\}$ and $P[v_0, \ldots, v_d]$. For $1 \leq i \leq d$, let u_i denote the vector $v_i - \mathrm{proj}(v_i, \mathrm{Aff}(\{v_0, \ldots, v_{i-1}\}))$. Cover the rectangle

$$R[v_0, \ldots, v_d] = v_0 + \{a_1 u_1 + a_2 u_2 + \cdots + a_d u_d | -2 \leq a_i \leq 2\}$$

by $O(1/\varepsilon^d)$ copies of translates of the scaled down rectangle

$$R'[v_0, \ldots, v_d] = \{\frac{\varepsilon}{2} c'_d (a_1 u_1 + a_2 u_2 + \cdots + a_d u_d) | 0 \leq a_i \leq 1\}.$$

Note that the bigger rectangle contains $P[v_0, \ldots, v_d]$. For each of the $O(1/\varepsilon^d)$ copies, we compute a $(k-1, \varepsilon)$ coreset of the points of $P[v_0, \ldots, v_d]$ contained in that copy, and add all these coreset points to Q.

This completes the description of the algorithm for computing Q.

Running Time and Size

Let $S(k)$ be an upper bound on the size of a (k, ε) coreset of any subset of points from $\mathcal{D}$ computed by our algorithm. We derive a bound for $S(k)$, for $k \geq 2$, using a bound for $S(k-1)$, noting that $S(1) = O(1/\varepsilon^{d-1})$.

There are $S(k-1)$ choices for v_0. For a choice of v_0, there are $O(\log \Delta) S(k-1)$ choices of v_1. For a given choice of $v_0, \ldots, v_t$ $(1 \leq t \leq d-1)$, there are $O(\log \Delta) S(k-1)$ choices of v_{t+1}. Thus for $0 \leq t \leq d$, we may bound the number of choices $v_0, \ldots, v_t$ by $O(\log^d \Delta (S(k-1))^{d+1})$. For each choice of $v_0, \ldots, v_t$, we compute $(k-1, \varepsilon)$ coresets $O(\log \Delta + 1/\varepsilon^d)$ times. We therefore have

$$S(k) \leq O((\frac{\log \Delta}{\varepsilon})^{d+1})(S(k-1))^{d+2}.$$

The bound in Theorem 1 on the size of Q follows from this.

A similar analysis bounds the running time.

Proof of Coreset Property

Let $S_1, \ldots, S_k$ be any k slabs that cover Q. We argue that an ε-expansion of the slabs covers P. Suppose the last slab S_k contains no point from $K \subset Q$. Then since K is a $(k-1, \varepsilon)$ coreset for P, and the first $k-1$ slabs $S_1, \ldots, S_{k-1}$ cover K, their ε-expansion covers P and we are done. Let us therefore assume that there is some $v_0 \in K$ that is contained in S_k. We now need to argue that an ε-expansion of $S_1, \ldots, S_k$ covers $P[v_0] = P$.

Stage 0: Let $j \geq 1$ be the largest integer such that S_k contains some point from $K_j[v_0]$. If no such j exists, let $j = 0$. The sets $K_i[v_0]$, $j+1 \leq i \leq \lambda$, are contained in the first $k-1$ slabs $S_1, \ldots, S_{k-1}$. Thus an ε-expansion of these slabs covers $B_i[v_0]$, $j+1 \leq i \leq \lambda$. If $j = 0$, we are done, since $B_0[v_0] = \{v_0\}$ is contained in S_k, and all points in $P[v_0] = \bigcup_{i=0}^{\lambda} B_i[v_0]$ are covered by an ε-expansion of the slabs.

So let us assume that $j \geq 1$. Let $v_1 \in K_j[v_0]$ be a point contained in S_k. We now need to argue that an ε-expansion of $S_1, \ldots, S_k$ covers $P[v_0, v_1] = \bigcup_{i=0}^{j} B_i[v_0]$.

Stage 1: First consider the point set $B_0[v_0, v_1]$ that lies on $\mathrm{Aff}(\{v_0, v_1\})$. Let us consider the points of $B_0[v_0, v_1]$ contained in one of the $O(1/\varepsilon)$ copies ρ of $R'[v_0, v_1]$. Since a $(k-1, \varepsilon)$ coreset of these points has been added to Q, these points will be covered by an ε-expansion of the first $k-1$ slabs if the slab S_k does not intersect ρ. So let us assume that S_k does intersect ρ. Since S_k contains v_0, v_1, by Lemma 1, it contains a rectangle that is a translate of a scaling of $R'[v_0, v_1]$ by a factor of $2/\varepsilon$. So this copy ρ of $R'[v_0, v_1]$ is contained in a slab 'parallel' to S_k (the hyperplane defining the two slabs are parallel) but whose width is $\varepsilon/2$ of the width of S_k. Since S_k intersects ρ, we may conclude that an ε-expansion of S_k covers ρ.

We have just argued that the point set $B_0[v_0, v_1]$ is covered by an ε-expansion of the k slabs, since each point in $B_0[v_0, v_1]$ is contained in one of the copies of $R'[v_0, v_1]$.

Let $j \geq 1$ be the largest integer such that S_k contains some point from $K_j[v_0, v_1]$. If no such j exists, set $j = 0$. The sets $K_i[v_0, v_1]$, $j+1 \leq i \leq \lambda$, are contained in the first $k-1$ slabs $S_1, \ldots, S_{k-1}$. Thus an ε-expansion of these slabs covers $B_i[v_0, v_1]$, $j+1 \leq i \leq \lambda$.

If $j = 0$, we are done, since all the points in $P[v_0, v_1] = \bigcup_{i=0}^{\lambda} B_i[v_0, v_1]$ are covered by an ε-expansion of the k slabs.

So let us assume that $j \geq 1$. Let $v_2 \in K_j[v_0, v_1]$ be a point contained in S_k. We now need to argue that an ε-expansion of $S_1, \ldots, S_k$ covers $P[v_0, v_1, v_2] = \bigcup_{i=0}^{j} B_i[v_0, v_1]$.

Stage t ($2 \leq t < d$): We enter this stage to argue that an ε-expansion of the k slabs contains $P[v_0, \ldots, v_t]$, for some choice of $v_0, \ldots, v_t$ that are contained in S_k.

First consider the point set $B_0[v_0, \ldots, v_t]$ that lies on $\mathrm{Aff}(\{v_0, \ldots, v_t\})$. Let us consider the points of $B_0[v_0, \ldots, v_t]$ contained in one of the $O(1/\varepsilon^t)$ copies ρ of $R'[v_0, \ldots, v_t]$. Since a $(k-1, \varepsilon)$ coreset of these points has been added to Q, these points will be covered by an ε-expansion of the first $k-1$ slabs if the slab S_k does not intersect ρ. So let us assume that S_k does intersect ρ. Since S_k contains $v_0, \ldots, v_t$, by Lemma 1, it contains a rectangle that is a translate of a scaling of $R'[v_0, \ldots, v_t]$ by a factor of $2/\varepsilon$. So this copy ρ of $R'[v_0, \ldots, v_t]$ is contained in a slab 'parallel' to S_k but whose width is $\varepsilon/2$ of the width of S_k. Since S_k intersects ρ, we may conclude that an ε-expansion of S_k covers ρ.

We have just argued that the point set $B_0[v_0, \ldots, v_t]$ is covered by an ε-expansion of the k slabs, since each point in $B_0[v_0, \ldots, v_t]$ is contained in one of the copies of $R'[v_0, \ldots, v_t]$.

Let $j \geq 1$ be the largest integer such that S_k contains some point from $K_j[v_0, \ldots, v_t]$. If no such j exists, set $j = 0$. The sets $K_i[v_0, \ldots, v_t]$, $j+1 \leq i \leq \lambda$, are contained in the first $k-1$ slabs $S_1, \ldots, S_{k-1}$. Thus an ε-expansion of these slabs covers $B_i[v_0, \ldots, v_t]$, $j+1 \leq i \leq \lambda$.

If $j = 0$, we are done, since all the points in $P[v_0, \ldots, v_t] = \bigcup_{i=0}^{\lambda} B_i[v_0, \ldots, v_t]$ are covered by an ε-expansion of the k slabs.

So let us assume that $j \geq 1$. Let $v_{t+1} \in K_j[v_0, \ldots, v_t]$ be a point contained in S_k. We now need to argue that an ε-expansion of $S_1, \ldots, S_k$ covers $P[v_0, \ldots, v_t, v_{t+1}] = \bigcup_{i=0}^{j} B_i[v_0, \ldots, v_t]$.

Stage d: We enter this stage to argue that an ε-expansion of the k slabs contains $P[v_0, \ldots, v_d]$, for some choice of $v_0, \ldots, v_d$ that are contained in S_k. This argument is identical to the argument given above for $B_0[v_0, \ldots, v_t]$. In fact, $P[v_0, \ldots, v_d]$ may be thought of as $B_0[v_0, \ldots, v_d]$.

We have completed the proof of Theorem 1.

Acknowledgements

We thank Piotr Indyk for suggesting the problem that is addressed in this paper, and Sariel Har-Peled for raising it again.

References

1. P. K. Agarwal, S. Har-Peled, and K. R. Varadarajan, Approximating extent measures of points, *J. Assoc. Comput. Mach.*, 51 (2004), 606–635.
2. P. K. Agarwal, C. M. Procopiuc, and K. R. Varadarajan, Approximation algorithms for k-line center, *Proc. 10th Annu. European Sympos. Algorithms*, 2002, pp. 54–63.

3. G. Barequet and S. Har-Peled, Efficiently approximating the minimum-volume bounding box of a point set in three dimensions, *J. Algorithms*, 38 (2001), 91–109.
4. T. M. Chan, Approximating the diameter, width, smallest enclosing cylinder and minimum-width annulus, *Internat. J. Comput. Geom. Appl.*, 12 (2002), 67–85.
5. T. M. Chan, Faster core-set constructions and data stream algorithms in fixed dimensions, *Proc. 20th Annu. ACM Sympos. Comput. Geom.*, 2004, pp. 152–159.
6. S. Har-Peled, Clustering motion, *Discrete Comput. Geom.*, 31 (2004), 545–565.
7. S. Har-Peled, No coreset, no cry, *Proc. 24th Conf. Found. Soft. Tech. Theoret. Comput. Sci.*, 2004.
8. S. Har-Peled and Y. Wang, Shape fitting with outliers, *SIAM J. Comput.*, 33 (2004), 269–285.