Streaming Algorithms for Extent Problems in High Dimensions*

Pankaj K. Agarwal R. Sharathkumar

Department of Computer Science
Duke University
{pankaj, sharath}@cs.duke.edu

Abstract

We develop (single-pass) streaming algorithms for maintaining extent measures of a stream S of n points in $\mathbb{R}^d$. We focus on designing streaming algorithms whose working space is polynomial in d (poly(d)) and sub-linear in n. For the problems of computing diameter, width and minimum enclosing ball of S, we obtain lower bounds on the worst-case approximation ratio of any streaming algorithm that uses poly(d) space. On the positive side, we introduce the notion of blurred ball cover and use it for answering approximate farthest-point queries and maintaining approximate minimum enclosing ball and diameter of S. We describe a streaming algorithm for maintaining a blurred ball cover whose working space is linear in d and independent of n.

1 Introduction

In many applications, data arrives rapidly as a stream of points and there is limited space to store the input. Algorithms in the *streaming model* have to work with one or few passes over the data, using small space. Motivated by a wide range of applications, including networking, databases and geographic information systems, there is extensive work on streaming algorithms. Often, it is impossible to compute exact solutions given the space constraints. Hence most of the work has focused on developing approximation algorithms and several novel techniques have been developed. See books [5, 27] for a comprehensive review. In this paper, we design streaming algorithms for several geometric problems in high dimensions.

Problem statement. For a point $x \in \mathbb{R}^d$ and a value $r > 0$, let $\mathbb{B}(x, r)$ denote a ball of radius r centered at x. For a ball B, let $c(B)$, $r(B)$ denote its center and radius respectively, and let $(1+\varepsilon)B$ denote the *ε-expansion* of B, i.e., $\mathbb{B}(c(B), (1+\varepsilon)r(B))$.

Let S be a (finite) set of points in $\mathbb{R}^d$. We use MEB(S) to denote the smallest ball that contains S. For a parameter $\alpha > 1$, a ball B is called an α-approximation of the minimum enclosing ball of S, or α-MEB(S) for brevity, if $S \subset B$ and $r(B) \leq \alpha \cdot r(\text{MEB}(S))$. A set $C \subseteq S$ is called an α-coreset(S) if $S \subset \alpha \cdot \text{MEB}(C)$. Let diam($S$) denote any farthest pair (diameteral pair) of points in S, and any pair of points $r, s \in S$ such that, for every pair $p, q \in S$, $||pq|| \leq \alpha \cdot ||rs||$, is referred to as α-diam(S) (α-diameteral pair). For any slab J, which is a region bounded by a pair (h_1, h_2) of parallel $(d-1)$-dimensional hyperplanes, let d(J) be the minimum distance between h_1 and h_2. Let width(S) be any slab J such that $S \subset J$ and d(J) is minimized. For $\alpha > 1$, α-width(S) is a slab J' such that $S \subset J'$ and $\text{d}(\text{width}(S)) \leq \text{d}(J') \leq \alpha \cdot \text{d}(\text{width}(S))$. Finally, for a point $x \in \mathbb{R}^d$, an α-farthest-neighbor of x, α-FN(x) is a point $q \in S$ such that, for every $p \in S$, $||xp|| \leq \alpha \cdot ||xq||$.

In this paper, we study the problems of maintaining α-MEB(S), α-coreset(S), α-width(S), and α-diam(S), and of answering α-FN(x) queries, in the streaming model. We assume that the input points arrive one by one, and the algorithm updates the information quickly using little space. Our goal is to design streaming algorithms whose working space and update time are polynomial in d and sub-linear in the size of S.

Related work. There is extensive literature on fast algorithms for computing various extent measures, such as diameter, width, and smallest enclosing ball of a point set in low dimensions [3, 14, 15, 17, 26, 28]. Agarwal *et al.* [1] have developed approximation algorithms for computing many extent measures of a set of n points in $\mathbb{R}^d$ using coresets whose running time is $O(n + 1/\varepsilon^{O(d)})$; see also [2, 4, 11, 12]. Although these algorithms work well in small dimensions, the exponential dependency on d makes them unsuitable for higher dimensions. This has led to work on develop-

*Work on this paper is supported by NSF under grants CNS-05-40347, CCF-06-35000, CCF-09-40671 and DEB-04-25465, by ARO grants W911NF-07-1-0376 and W911NF-08-1-0452, by an NIH grant 1P50-GM-08183-01, by a DOE grant OEG-P200A070505, and by a grant from the U.S.–Israel Binational Science Foundation.

ing polynomial-time approximation algorithms (in n, d and $1/\varepsilon$) for different extent measures and other related problems. For example, Bădoiu and Clarkson [9] develop an elegant coreset-based algorithm that given a point set $S \in \mathbb{R}^d$ computes a $(1+\varepsilon)$-MEB(S) in time $O(nd/\varepsilon + 1/\varepsilon^5)$. See [10, 21, 23, 24] for other similar results. Clarkson [16] presented a general framework that gives coreset-based algorithms for a number of geometric problems in high dimensions; see also [19]. Goel *et al.* [20] describe fast approximation algorithms for several proximity problems. For instance, they describe a data structure that answers $\sqrt{2}$-FN(x) queries in $O(d)$ time after spending $O(nd)$ time in preprocessing.

There is also some work on streaming algorithms for extent problems in high dimensions. Chan and Zarrabi-Zadeh [29] gave a streaming algorithm, which maintains a $(3/2)$-MEB(S), for a stream $S \in \mathbb{R}^d$, using $O(d)$ space. They also proved that any deterministic algorithm that stores only a single ball in its working space cannot maintain a better than α-MEB(S), for $\alpha \leq \frac{1+\sqrt{2}}{2}$. Indyk [22] proposed a streaming algorithm for maintaining α-diam(S) for $\alpha > \sqrt{2}$, using $O(dn^{1/(\alpha^2-1)}\log n)$ space. Recently Clarkson and Woodruff [18] described streaming algorithms for several problems in numerical algebra.

Our results. We prove upper and lower bounds on the size of data structures for maintaining various extent measures under the streaming model. First, we present in Section 2, a data structure in the streaming model called the *ε-blurred ball cover*, for $\varepsilon > 0$, that maintains a subset $K \subseteq S$ of $O((1/\varepsilon^3)\log(1/\varepsilon))$ points. Roughly speaking K is the union of a set of $\{K_1, \ldots, K_u\}$, $u = O((1/\varepsilon^2)\log(1/\varepsilon))$, of subsets of S each of size $O(1/\varepsilon)$ so that S lies in the union of $(1+\varepsilon)$MEB(K_i). The data structure can be updated in $O((d/\varepsilon^2)\log(1/\varepsilon))$ amortized time. We then show in Section 3 that K can be used to:

(i) compute a α-FN(x), for any query $x \in \mathbb{R}^d$ and for $\alpha = \sqrt{2} + \varepsilon$;

(ii) maintain a α-diam(S) for $\alpha = \sqrt{2} + \varepsilon$;

(iii) maintain a α-coreset(S), for $\alpha = \sqrt{2} + \varepsilon$;

(iv) maintain a α-MEB(S) for $\alpha = \frac{1+\sqrt{3}}{2} + \varepsilon$.

Next, we show in Section 4 that any randomized streaming algorithm that maintains α-diam(S), α-width(S), α-MEB(S), or α-coreset(S) with probability at least 2/3, requires $\Omega(\min(|S|, \exp(d^{1/3})))$ space for certain values of α. In particular, $\alpha < \sqrt{2}(1 - 2/d^{1/3})$ for α-diam(S) and α-coreset(S), $\alpha \leq d^{1/3}/8$ for α-width(S), and $\alpha < \frac{1+\sqrt{2}}{2}(1 - 2/d^{1/3})$ for α-MEB(S). All lower bounds are obtained using known lower bounds on the one-round communication complexity of indexing [25]. Communication complexity has been widely used to prove lower bounds on the size of various data structures including in the streaming model [6, 7, 18].

Our algorithms for maintaining α-diam(S) and α-coreset(S) are close to optimal. Note the contrast between the lower and upper bounds for MEB — a coreset based algorithm has a lower bound of $\sqrt{2}$, but we are able to circumvent this bound by, roughly speaking, maintaining a set of $O((1/\varepsilon^2)\log(1/\varepsilon))$ balls and returning a ball that contains these balls. We can regard the centers of these balls as "weighted" points and thus we can get better results by maintaining a "weighted weak coreset" — it is not a subset of input points. Although it is known that weak ε-nets are more powerful than ε-nets [13], we are unaware of similar results for coresets of MEB and other extent measures.

2 Blurred Ball Cover

This section defines the notion of blurred ball cover and describes an algorithm for maintaining such a cover in the streaming model. For a parameter $0 < \varepsilon \leq 1$, an ε-blurred ball cover of a set S of n points in $\mathbb{R}^d$, denoted by $\mathcal{K} = \mathcal{K}(S)$, is a sequence $\langle K_1, K_2, \ldots, K_u \rangle$, where each $K_i \subseteq S$ is a subset of $O(1/\varepsilon)$ points that satisfies the following three properties; let $B_i = \text{MEB}(K_i)$ and $K = \bigcup_{i \leq u} K_i$.

(P1) For all $1 \leq i < u$, $r(B_{i+1}) \geq (1 + \varepsilon^2/8)r(B_i)$;

(P2) For all $i < j$, $K_i \subset (1+\varepsilon)B_j$;

(P3) For every $p \in S$, $\exists i \leq u$ such that $p \in (1+\varepsilon)B_i$.

Algorithm 1 describes a simple procedure called Update($\mathcal{K}, A$), that given $\mathcal{K} := \mathcal{K}(S)$ and a set of points $A \subset \mathbb{R}^d$, computes $\mathcal{K}(S \cup A)$. If we update $\mathcal{K}$ as each new point arrives, then A consists of a single point. However, as we will see below, it will be more efficient for some of our applications to update A in a batched mode. That is, newly arrived points are stored in a buffer A and when its size exceeds certain threshold, Update procedure is called to update $\mathcal{K}$. Update relies on a procedure Approx-MEB(Z, ε) that takes a set Z of points and a parameter $0 < \varepsilon \leq 1$ and returns a set $G \subseteq Z$ of $O(1/\varepsilon)$ points and $B = \text{MEB}(G)$ such that $Z \subset (1+\varepsilon)B$. Bădoiu and Clarkson [9] have described such a procedure with $O(d|Z|/\varepsilon + 1/\varepsilon^5)$ running time. For efficiency reasons, we explicitly maintain $B_i = \text{MEB}(K_i)$ for every $K_i \in \mathcal{K}$.

Update procedure. If there is a point in A that does not lie in the union of the ε-expansions of B_i's, then the Update procedure invokes Approx-MEB on $K \cup A$

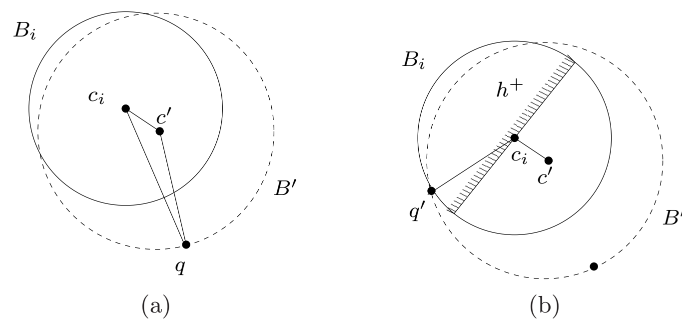

Figure 1: Proof of Lemma 2.2. (a) $||c_i c'|| \leq 2\varepsilon r_i/3$, (b)$||c_i c'|| > 2\varepsilon r_i/3$.

with approximation ratio $\varepsilon/3$. Let $K^* \subseteq K$ be the point set and $B^* = \text{MEB}(K^*)$ be the ball returned by Approx-MEB. The Update procedure adds K^* to $\mathcal{K}$ and then deletes all K_i's for which $r(K_i) \leq \varepsilon r(K^*)/4$.

Algorithm 1 Update($\mathcal{K}, A$)

> **if** $\exists p \in A$, $\forall i \leq u$, $p \notin (1+\varepsilon)B_i$ **then**
> $K^*, B^* :=$ Approx-MEB $((K \cup A), \varepsilon/3)$
> $\mathcal{K}_D := \{K_i \mid r(B_i) \leq \varepsilon r(B^*)/4\}$
> $\mathcal{K} := (\mathcal{K} \setminus \mathcal{K}_D) \circ \langle K^* \rangle$
> **end if**

Correctness. The proof of correctness relies on the following well-known observation; see e.g. [10].

Lemma 2.1. *Let P be a set of points in $\mathbb{R}^d$ and let $B = \text{MEB}(P)$. Then any closed halfspace that contains $c(B)$ also contains at least one point of P that is on ∂B.*

Lemma 2.2. *For any $i < u$, $r(B_{i+1}) \geq (1+\varepsilon^2/8)r(B_i)$.*

Proof. Update procedure adds K_{i+1} to $\mathcal{K}$ only if there is a point $q \in A$ such that $q \notin (1+\varepsilon)B_i$. Let $B' = \text{MEB}(K \cup \{q\})$; $r(B_{i+1}) \geq r(B')$. Let $r' = r(B')$, $c' = c(B')$, $c_i = c(B_i)$, and $r_i = r(B_i)$. We prove that $r' > (1+\varepsilon^2/8)r_i$. If $||c_i c'|| \leq 2\varepsilon r_i/3$ (Figure 1(a)), then

$$\begin{aligned} r' &\geq ||c'q|| \\ &\geq ||c_i q|| - ||c' c_i|| \\ &\geq (1+\varepsilon)r_i - 2\varepsilon r_i/3 \\ &\geq (1+\varepsilon^2/8)r_i. \end{aligned}$$

On the other hand, if $||c' c_i|| > 2\varepsilon r_i/3$ (Figure 1(b)), then let h be the hyperplane passing through c_i with the direction $c_i c'$ as its normal, and let h^+ be the halfspace, bounded by h, that does not contain c'. By Lemma 2.1, there is a point $q' \in K_i \cap h^+$ that is at a distance r_i from c_i. Therefore

$$\begin{aligned} r' &\geq ||q'c'|| \\ &\geq (||c_i c'||^2 + ||q' c_i||^2)^{1/2} \\ &\geq ((2\varepsilon r_i/3)^2 + r_i^2)^{1/2} \\ &\geq (1+\varepsilon^2/8)r_i. \end{aligned}$$

□

We now prove that (P1)–(P3) are maintained by Update. If for every $p \in A$, there is an $i \leq u$ such that $p \in (1+\varepsilon)B_i$, then the set $\mathcal{K}$ does not change, and (P1)–(P3) continue to hold. Hence, assume that there is a $p \in A$ such that $p \notin B_i$ for all $i \leq u$. By Lemma 2.2, property (P1) continues to hold after each update. Note that if $p \notin (1+\varepsilon)B_i$ for all $1 \leq i \leq u$, then Update computes a $(1+\varepsilon/3)$-MEB$(K \cup A)$. Since K^* is the only subset added to $\mathcal{K}$, $K \subseteq (1+\varepsilon)B^*$ and K^* is added at the end of the sequence, $K_i \subseteq (1+\varepsilon)B^*$. Thus property (P2) holds after each Update. Note that a prefix $\mathcal{K}_D$ is deleted from $\mathcal{K}$, so (P3) may be violated. However, the next two lemmas prove that (P3) is also satisfied.

Lemma 2.3. *For $i < j$, $c(B_i) \in (1+\varepsilon)B_j$.*

Proof. Suppose, on the contrary, that $c(B_i) \notin (1+\varepsilon)B_j$. Let γ be a halfspace that contains $c(B_i)$ but $\gamma \cap (1+\varepsilon)B_j = \emptyset$. By Lemma 2.1, γ contains a point $p \in K_i$, but $p \notin (1+\varepsilon)B_j$, which contradicts the fact $K_i \subset (1+\varepsilon)B_j$ (by (P2)). Hence, $c(B_i) \in (1+\varepsilon)B_j$. □

Lemma 2.4. *For all $K_i \in \mathcal{K}_D$, $(1+\varepsilon)B_i \subseteq (1+\varepsilon)B^*$.*

Proof. Let $c^* = c(B^*)$, $r^* = r(B^*)$, $c_i = c(B_i)$, and $r_i = r(B_i)$. Let $K_i \in \mathcal{K}_D$, then $r_i \leq \varepsilon r^*/4$. Since $K_i \subset (1+\varepsilon/3)B^*$, the proof of Lemma 2.3 implies $c(B_i) \in (1+\varepsilon/3)B^*$. For any point $x \in (1+\varepsilon)B_i$,

$$\begin{aligned} ||xc^*|| &\leq ||c^* c_i|| + ||c_i x|| \\ &\leq ||c^* c_i|| + (1+\varepsilon)r_i \\ &\leq (1+\varepsilon/3)r^* + \varepsilon(1+\varepsilon)r^*/4 \\ &\leq (1+\varepsilon)r^*. \end{aligned}$$

Therefore $q \in (1+\varepsilon)B^*$ and thus $(1+\varepsilon)B_i \subseteq (1+\varepsilon)B^*$. $\square$

Lemma 2.4 immediately implies that for any $K_i \in \mathcal{K}_D$, if there is a point $q \in (1+\varepsilon)B_i$ then $q \in (1+\varepsilon)B^*$. Hence, for every $q \in S$, there is an i such that $q \in (1+\varepsilon)B_i$, implying (P3).

Size and update time. Let $r_i = r(B_i)$. Update ensures that $r_u/r_1 \le 4/\varepsilon$. By (P1), $r_{i+1} \ge (1+\varepsilon^2/8)r_i$. Therefore $u \le \lceil \log_{1+\varepsilon^2/8}(4/\varepsilon) \rceil = O((1/\varepsilon^2)\log(1/\varepsilon))$. Hence $|K| \le \sum |K_i| = O((1/\varepsilon^3)\log(1/\varepsilon))$.

Note that Update spends $O(u.|A|) = O((d|A|/\varepsilon^2)\log(1/\varepsilon))$ time to test the condition in line 1. The time spent in computing K^* and B^* in line 2 is $O((|A|+|K|)d/\varepsilon + (1/\varepsilon^5))$ or

$$O((d|A|/\varepsilon) + (d/\varepsilon^4)\log(1/\varepsilon) + (1/\varepsilon^5)).$$

If we update $\mathcal{K}$ after the arrival of each new point, then the update time is $O(d/\varepsilon^5)$. However, if we batch the updates and invoke Update only after $O(1/\varepsilon^3)$ points have arrived, the total time spent will be $O((d/\varepsilon^5)\log(1/\varepsilon))$ which is the time taken by line 1 of the update procedure. The amortized time for Update will then be $O((d/\varepsilon^2)\log(1/\varepsilon))$. We thus conclude the following.

Theorem 2.1. *For any $0 < \varepsilon \le 1$, an ε-blurred ball cover of size $O((d/\varepsilon^3)\log(1/\varepsilon))$ of a stream of points can be maintained in $O((d/\varepsilon^2)\log(1/\varepsilon))$ amortized time.*

3 Applications of Blurred Ball Cover

We now describe streaming algorithms for answering $(\sqrt{2}+\varepsilon)$-FN(x) queries and for maintaining $(\sqrt{2}+\varepsilon)$-diam(S), $(\sqrt{2}+\varepsilon)$-MEB(S) and $(\frac{1+\sqrt{3}}{2}+\varepsilon)$-MEB$(S)$ using the blurred ball cover. We use the following simple inequality repeatedly. For any $x, y \in \mathbb{R}$,

$$(3.1) \qquad (x+y) \le \sqrt{2}(x^2+y^2)^{1/2}.$$

Farthest-neighbor queries. We maintain an ε-blurred ball cover $\mathcal{K}$ and update it in the batched mode. Let A be the set of newly arrived points in S that have not been processed yet. Set $Q = K \cup A$; $|Q| = O((1/\varepsilon^3)\log(1/\varepsilon))$. For any query point $x \in \mathbb{R}^d$, we return the point of Q that is farthest from x. We claim that for any $x \in \mathbb{R}^d$,

$$\max_{p \in S} ||xp|| \le (\sqrt{2}+\varepsilon) \max_{q \in Q} ||qx||.$$

Let $p' = \arg\max_{p \in S} ||px||$ and $q' = \arg\max_{q \in Q} ||qx||$. If $p' \in A$ then $p' = q'$ and the claim is obviously true. If $p' \in S \setminus A$, i.e., p' has been processed by Update, then, by (P3), there is an $i \le u$ such that $p' \in (1+\varepsilon)B_i$. Assuming $c_i = c(B_i)$ and $r_i = r(B_i)$,

$$(3.2) \quad ||xp'|| \le ||xc_i|| + ||c_i p'|| \le ||xc_i|| + (1+\varepsilon)r_i.$$

Let h be the hyperplane passing through c_i and normal to xc_i and let h^+ be the halfspace, bounded by h, that does not contain x. By Lemma 2.1, there is a point $z \in K_i$ such that $||zc_i|| = r_i$. Since $\angle xc_i z$ is obtuse,

$$\begin{aligned} ||xz|| &\ge (||xc_i||^2 + ||c_i z||^2)^{1/2} \\ (3.3) \qquad &\ge (||xc_i||^2 + r_i^2)^{1/2}. \end{aligned}$$

By combining (3.2) and (3.3) and using (3.1),

$$\begin{aligned} ||xp'|| &\le \sqrt{2}(||xc_i||^2 + r_i^2)^{1/2} + \varepsilon r_i \\ &\le (\sqrt{2}+\varepsilon)||xz|| \\ &\le (\sqrt{2}+\varepsilon)||xq'||. \end{aligned}$$

We conclude the following.

Theorem 3.1. *For a stream S of points in $\mathbb{R}^d$ and a parameter $0 < \varepsilon \le 1$, there is a data structure of size $O((d/\varepsilon^3)\log(1/\varepsilon))$ that answers $(\sqrt{2}+\varepsilon)$-FN(x) in time $O((d/\varepsilon^3)\log(1/\varepsilon))$. The amortized update time is $O((d/\varepsilon^2)\log(1/\varepsilon))$.*

Diameter. For a point set S, we maintain a α-diam(S), for $\alpha = \sqrt{2}+\varepsilon$, using the above data structure for answering α-FN(x) queries. Let p, q be the current α-diameteral pair maintained by the algorithm. When a new point z arrives, we compute the $w = \alpha$-FN(z). If $||wz|| > ||pq||$, we replace (p, q) with (w, z). We conclude the following.

Theorem 3.2. *For a stream S of n points in $\mathbb{R}^d$, there is a data structure of size $O((d/\varepsilon^3)\log(1/\varepsilon))$ that maintains a $(\sqrt{2}+\varepsilon)$-diam(S). The amortized update time of this structure is $O((d/\varepsilon^3)\log(1/\varepsilon))$.*

Coreset for minimum enclosing ball. For a stream S of points, we maintain an ε-blurred ball cover $\mathcal{K}$ and update it in the batched mode. Let A be the set of points not processed by Update and let $Q = K \cup A$. Let $B = \text{MEB}(Q)$, $c = c(B)$ and $r = r(B)$. We argue that for every $K_i \in \mathcal{K}$, $(1+\varepsilon)B_i \subseteq (\sqrt{2}+\varepsilon)B$.

Lemma 3.1. *For all $i \le u$, $(1+\varepsilon)B_i \subseteq (\sqrt{2}+\varepsilon)B$.*

Proof. For any ball B_i in the blurred ball cover, let $c_i = c(B_i)$ and $r_i = r(B_i)$. Observe that $K_i \subseteq B$. Let $t^* = \arg\max_{t \in (1+\varepsilon)B_i} ||tc||$. $||t^*c|| = ||cc_i|| + (1+\varepsilon)r_i$.

By Lemma 2.1, there is a point $p \in K_i$ such that $||pc|| \geq \sqrt{||cc_i||^2 + r_i^2}$. But $||pc|| \leq r$. Hence,

$$\frac{||t^*c||}{||pc||} \leq \frac{||cc_i|| + (1+\varepsilon)r_i}{\sqrt{||cc_i||^2 + r_i^2}} \leq \sqrt{2} + \varepsilon,$$

or $||t^*c|| \leq (\sqrt{2}+\varepsilon)||pc|| \leq (\sqrt{2}+\varepsilon)r$. Thus $(1+\varepsilon)B_i \subseteq (\sqrt{2}+\varepsilon)B$. □

Lemma 3.1 in conjunction with property (P3) of blurred ball cover implies that $S \subset (\sqrt{2}+\varepsilon)B$. Thus Q is a $(\sqrt{2}+\varepsilon)$-coreset(S). We conclude the following.

THEOREM 3.3. *For a stream S of points in $\mathbb{R}^d$ and a parameter $0 < \varepsilon \leq 1$, there is a data structure of size $O((d/\varepsilon^3)\log(1/\varepsilon))$ that maintains a $(\sqrt{2}+\varepsilon)$-coreset(S) of minimum enclosing ball. The amortized update time of this structure is $O((d/\varepsilon^2)\log(1/\varepsilon))$.*

Minimum enclosing ball. For a stream S of points, we maintain a $(\varepsilon/9)$-blurred ball cover $\mathcal{K}$ of S. $\mathcal{K}$ is updated whenever a new point arrives. Let $\mathcal{B} = \{B_1, \ldots, B_u\}$. Let $B^* = \text{MEB}(\mathcal{B})$. We return $(1+\varepsilon/3)B^*$ which can be computed in time $O(1/\varepsilon^5)$ [23]. Hence the total update time is $O(1/\varepsilon^5)$. Let $\hat{r} = r(\text{MEB}(S))$.

LEMMA 3.2. $r(B^*) \leq \left(\frac{1+\sqrt{3}}{2} + \varepsilon/3\right)\hat{r}$.

Proof. Let $\mathcal{K}' = \langle K_i \mid B_i \cap \partial B^* \neq \emptyset \rangle$ and $\mathcal{B}' = \{B_i \mid K_i \in \mathcal{K}'\}$. Note that any subsequence of $\mathcal{K}$, and hence $\mathcal{K}'$, satisfies properties (P1) and (P2) of blurred ball cover. Let B_t be the largest ball in $\mathcal{B}'$. Let $c_t = c(B_t)$, $r_t = r(B_t)$, $c^* = c(B^*)$, $r^* = r(B^*)$ and let $k = r^*/r_t$. Observe that $||c_tc^*|| = r^* - r_t = (k-1)r_t$. Let h be a hyperplane passing through c^* with a normal parallel to c_tc^*. Let h^+ be the halfspace bounded by h and that does not contain c_t. By Lemma 2.1, there is a point $b_i \in \partial B^* \cap B_i$, for some i, such that $||b_ic^*|| = r^* = kr_t$. Since $\angle c_tc^*b_i$ is obtuse, we have

$$||c_tb_i|| \geq \sqrt{(k-1)^2r_t^2 + k^2r_t^2}. \tag{3.4}$$

Since the proof of Lemma 3.1 requires only (P2) of blurred ball cover, it holds for any subsequence of $\mathcal{K}$. In particular, it holds for $\mathcal{K}'$. Hence, for all $B_i \in \mathcal{B}'$, $B_i \subseteq (\sqrt{2}+\varepsilon/9)B_t$, implying that

$$||c_tb_i|| \leq (\sqrt{2}+\varepsilon/9)r_t. \tag{3.5}$$

Combining (3.4) and (3.5), we have

$$(k-1)^2r_t^2 + k^2r_t^2 \leq (\sqrt{2}+\varepsilon/9)^2r_t^2 \leq 2(1+\varepsilon/9)^2r_t^2.$$

Solving this quadratic equation, we can deduce that, $k \leq \frac{1+\sqrt{3}}{2} + \varepsilon/3$.

Thus

$$r^* \leq kr_t \leq k\hat{r} \leq \left(\frac{1+\sqrt{3}}{2} + \varepsilon/3\right)\hat{r}.$$

□

By (P3),

$$S \subset \bigcup_{i \leq u}(1+\varepsilon/9)B_i \subseteq (1+\varepsilon/9)B^*.$$

By Lemma 3.2, we have

$$r^* \leq \left(\frac{1+\sqrt{3}}{2} + \varepsilon/3\right)\hat{r}.$$

Thus

$$\begin{aligned}(1+\varepsilon/3)r^* &\leq \left(\frac{1+\sqrt{3}}{2} + \varepsilon/3\right)(1+\varepsilon/3)\hat{r} \\ &\leq \left(\frac{1+\sqrt{3}}{2} + \varepsilon\right)\hat{r},\end{aligned}$$

implying that $(1+\varepsilon/3)B^*$ is indeed a $\left(\frac{1+\sqrt{3}}{2} + \varepsilon\right)$-MEB($S$). We conclude the following.

THEOREM 3.4. *Given a stream S of points in $\mathbb{R}^d$, there is a data structure of size $O((d/\varepsilon^3)\log(1/\varepsilon))$ that maintains a $(\frac{\sqrt{3}+1}{2}+\varepsilon)$-MEB($S$). The (worst-case) update time of this structure is $O(d/\varepsilon^5)$.*

Remark. A slightly better bound on r^*, the radius of the ball returned by the above algorithm can be obtained using a more involved argument, While, from the lower bounds on α-MEB(S) in Section 4 it follows that $r^* > \frac{1+\sqrt{2}}{2}$, the following example shows that one cannot hope to prove $r^* = \frac{1+\sqrt{2}}{2} + \varepsilon$. Let the input be a stream $S = \langle p_1, p_2, p_3, p_4, \ldots \rangle$, where $p_1 = (\frac{1}{\sqrt{2}}, \frac{1}{2\sqrt{2}}, \frac{\sqrt{3}}{2\sqrt{2}})$, $p_2 = (\frac{1}{\sqrt{2}}, \frac{1}{2\sqrt{2}}, -\frac{\sqrt{3}}{2\sqrt{2}})$, $p_3 = (\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}, 0)$, $p_4 = (-1, 0, 0)$, and $p_5, p_6, \ldots$, are points on a unit sphere in $\mathbb{R}^3$. When the points in S arrive in this order, after the arrival of p_4, the blurred ball cover will consist of $\mathcal{K} = \{\{p_1, p_2\}, \{p_1, p_2, p_3\}, \{p_1, p_2, p_3, p_4\}\}$. The three balls $\{B_1, B_2, B_3\}$ of $\mathcal{K}$ are described by

$$\begin{aligned}B_1 &= \mathbb{B}\left(\left(\frac{1}{\sqrt{2}}, \frac{1}{2\sqrt{2}}, 0\right), \frac{\sqrt{3}}{2\sqrt{2}}\right), \\ B_2 &= \mathbb{B}\left(\left(\frac{1}{\sqrt{2}}, 0, 0\right), \frac{1}{\sqrt{2}}\right), \\ B_3 &= \mathbb{B}\left((0,0,0), 1\right).\end{aligned}$$

Let $B' = \text{MEB}(B_1 \cup B_2)$ with $r(B') = (1+\sqrt{2})/2$ and $c(B') = ((1/\sqrt{2}) - (1/2), 0, 0)$. Let q be the farthest point of B_3 from $c(B')$. $||qc(B')|| = \sqrt{3/2} > \frac{1+\sqrt{2}}{2}$. We evaluate B^* to be θ-MEB(S) where $\theta \geq \frac{1+\sqrt{2}}{2} + 10^{-5}$.

4 Lower Bounds

In this section we show that any randomized streaming algorithm that maintains α-diam(S), α-width(S), α-MEB(S), or α-coreset(S) with probability at least 2/3, requires $\Omega(\min(|S|, \exp(d^{1/3})))$ space for certain values of α. In particular, $\alpha < \sqrt{2}(1 - 2/d^{1/3})$ for α-diam(S) and α-coreset(S), $\alpha \leq d^{1/3}/8$ for α-width(S), and $\alpha < \frac{1+\sqrt{2}}{2}(1 - 2/d^{1/3})$ for α-MEB(S). Let $\mathbb{S}^{d-1}$ denote $(d-1)$-dimensional unit sphere centered at origin. We show how to sample points from $\mathbb{S}^{d-1}$ which will be crucial for proving our lower bounds.

Sampling in $\mathbb{S}^{d-1}$. Let $u \in \mathbb{S}^{d-1}$ and $\varepsilon \in (0,1)$. Let H be the hyperplane $\langle x, u\rangle = \varepsilon$, i.e., H is the hyperplane that is at distance ε from the origin and normal to the vector u. H divides $\mathbb{S}^{d-1}$ into two spherical regions. We refer to the smaller spherical region as a *spherical cap* and denote it by $C(u, \varepsilon)$. We define its measure to be

$$\mu(u, \varepsilon) = \frac{\text{SA}(C(u,\varepsilon))}{\text{SA}(\mathbb{S}^{d-1})},$$

where $\text{SA}(X)$ is the surface area of X. It is well known [8] that

$$\mu(u, \varepsilon) \leq \exp(-d\varepsilon^2/2). \tag{4.6}$$

Lemma 4.1. *There is a centrally symmetric point set $K \subseteq \mathbb{S}^{d-1}$ of size $\Omega(\exp(d^{1/3}))$ such that for any pair of distinct points $p, q \in K$ if $p \neq -q$, then*

$$\sqrt{2}(1 - 2/d^{1/3}) \leq ||pq|| \leq \sqrt{2}(1 + 2/d^{1/3}).$$

Proof. The set K is constructed incrementally. We maintain the invariant that if $p, q \in K$ and $q \neq p$ then $q \notin C(p, 2/d^{1/3})$. We also maintain a centrally symmetric region $F = \mathbb{S}^{d-1} \setminus \bigcup_{q \in K} C(q, 2/d^{1/3})$. Initially $K = \emptyset$ and $F = \mathbb{S}^{d-1}$. If $F \neq \emptyset$, we choose an antipodal pair of points $p, -p$ from F. By construction, $p, -p \in C(q, 2/d^{1/3})$ for any point $q \in K$. Moreover, by symmetry, no point in K lies in $C(p, 2/d^{1/3}) \cup C(-p, 2/d^{1/3})$. We add p and $-p$ to K. We set $F = F \setminus \{C(p, 2/d^{1/3}) \cup C(-p, 2/d^{1/3})\}$ to ensure that the invariant holds after adding p and p'. The algorithm terminates when $F = \emptyset$.

By (4.6), $\mu(p, 2/d^{1/3}) + \mu(-p, 2/d^{1/3}) \leq \exp(-d^{1/3})$, therefore the above algorithm performs $\Omega(\exp(d^{1/3}))$ steps before it stops. Hence $|K| = \Omega(\exp(d^{1/3}))$. Let $s, t \in K$ be such that $s \neq t, -t$. By construction, $s \notin C(t, 2/d^{1/3}) \cup C(-t, 2/d^{1/3})$. Let t' (resp. t'') be a point on the boundary of $C(s, 2/d^{1/3})$ (resp. $C(-s, 2/d^{1/3})$). Then $||st'|| \leq ||st|| \leq ||st''||$. A simple trignometric calculation shows that (see Figure 2(a))

$$\begin{aligned} ||st|| &\leq ||st''|| \\ &\leq \left((1 + 2/d^{1/3})^2 + (1 - 4/d^{2/3})\right)^{1/2} \\ &\leq \sqrt{2}(1 + 2/d^{1/3})^{1/2} \\ &\leq \sqrt{2}(1 + 2/d^{1/3}). \end{aligned}$$

Similarly one can show

$$||st|| \geq ||st'|| \geq \sqrt{2}(1 - 2/d^{1/3}).$$

□

Diameter. The problem Index is defined as follows. Let Alice and Bob be two players. Suppose Alice has a binary string $\sigma \in \{0,1\}^k$ and Bob has an index $i \in [1:k]$. For any $j \in [1:k]$, let σ_j be the bit that appears in the jth position of σ. Alice sends a message to Bob after which Bob needs to determine σ_i with probability at least 2/3. It is well-known that the length of any message sent by Alice that helps Bob determine σ_i with probability at least 2/3 is $\Omega(k)$ [25].

We choose the smallest d so that $k \leq \exp(d^{1/3})$. Hence $k = \Theta(\exp(d^{1/3}))$. Let $\mathbb{A}$ be a streaming algorithm that maintains α-diam(S) for $\alpha \leq \sqrt{2}(1 - 2/d^{1/3})$ with probability at least 2/3. We choose $2\exp(d^{1/3})$ points on $\mathbb{S}^{d-1}$ using Lemma 4.1. Let L be the subset of at least k of these points that lie on the hemisphere $x_1 \geq 0$. We sort L in lexicographic order, which induces a map $\varphi : [1:k] \to L$. The map φ is independent of the input string σ and the index i. Hence we can assume that both Alice and Bob know φ without communicating any bits.

For any $\sigma \in \{0,1\}^k$, let $\varphi(\sigma) = \{\varphi(x) \mid \sigma_x = 1\}$. Alice adds $\varphi(\sigma)$ as input to $\mathbb{A}$ in an arbitrary order. Then, Alice communicates the working space of $\mathbb{A}$ to Bob. In order to determine the σ_i, Bob adds as input, $-\varphi(i)$ to $\mathbb{A}$. Let $p, q \in S$ be the pair of points returned by $\mathbb{A}$ at the end of the protocol. If $p = -q$, Bob determines $\sigma_i = 1$. Else, Bob determines $\sigma_i = 0$. The following lemma shows the correctness of this protocol.

Lemma 4.2. *For any $\sigma \in \{0,1\}^k$ and $i \in [1:k]$. Let $S = \varphi(\sigma) \cup \{-\varphi(\sigma_i)\}$ and $\alpha \leq \sqrt{2}(1 - 2/d^{1/3})$. Every α-diam(S) pair is an antipodal pair if and only if $\sigma_i = 1$.*

Proof. Since all points in $\varphi(\sigma)$ lie in one hemisphere, there is an antipodal pair of points in S if and only if $\varphi(i) \in \varphi(\sigma)$, i.e. $\sigma_i = 1$. If $\sigma_i = 0$, then $\varphi(i) \notin \varphi(\sigma)$

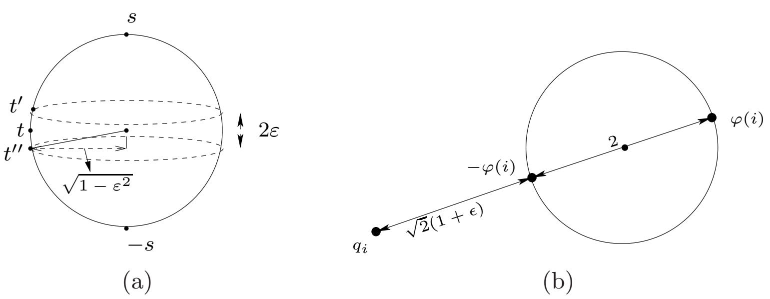

Figure 2: (a) Proof of Lemma 1, (b) Lower bound for α-MEB(S); here $\varepsilon = 2/d^{1/3}$.

and $\mathbb{A}$ does not return an antipodal pair of points. On the other hand, suppose $\sigma_i = 1$ and suppose there is an α-diametral pair $p, q \in S$ such that $p \neq -q$. Since $\{\varphi(\sigma_i), -\varphi(\sigma_i)\} \subseteq S$, diam$(S) \geq 2$. From Lemma 4.1, it follows that $||pq|| \leq \sqrt{2}(1 + 2/d^{1/3})$, i.e., diam$(S) > \alpha \cdot ||pq||$ for $\alpha \leq \sqrt{2}(1 - 2/d^{1/3})$. Thus p, q is not an α-diametral pair implying that α-diam(S) is an antipodal pair of points. □

Since $\mathbb{A}$ returns an α-diam(S) with probablity at least 2/3, Bob determines σ_i correctly with the same probability. Since the communication complexity of any randomized algorithm for the indexing problem is $\Omega(k) = \Omega(\exp(d^{1/3}))$, it follows that the work space of $\mathbb{A}$ is $\Omega(\exp(d^{1/3}))$. We thus conclude the following.

THEOREM 4.1. *Any streaming algorithm that maintains an* α-diam(S) *of a set* S *of* n *points in* $\mathbb{R}^d$*, for* $\alpha < \sqrt{2}(1 - 2/d^{1/3})$*, with probability at least* 2/3 *requires* $\Omega(\min\{n, \exp(d^{1/3})\})$ *bits of storage.*

Minimum enclosing ball. Let $\mathbb{A}$ be an algorithm that maintains an α-MEB(S), for $\alpha \leq \frac{1+\sqrt{2}}{2}(1-2/d^{1/3})$ with probability at least 2/3. The map φ is defined as above. Alice passes points in $\varphi(\sigma)$ as input to $\mathbb{A}$ in an arbitrary order and communicates the working space of $\mathbb{A}$ to Bob. For index i, let q_i be the point that is in the direction $-\varphi(i)$ and at distance $\sqrt{2}(1 + 2/d^{1/3})$ (resp. $2 + \sqrt{2}(1 + 2/d^{1/3})$) from $-\varphi(i)$(resp. $\varphi(i)$); see Figure 2(b). Bob adds q_i as input to $\mathbb{A}$. If $\mathbb{A}$ returns a ball of radius at least $(1 + \frac{1}{\sqrt{2}})$, then Bob declares $\sigma_i = 1$. Otherwise, Bob declares $\sigma_i = 0$.

Let $S = \{\varphi(\sigma) \cup \{q_i\}\}$. Note that $||\varphi(i)q_i|| = 2 + \sqrt{2}(1 + 2/d^{1/3})$. Hence if $\sigma_i = 1$, then $\varphi(i), q_i \in S$ and hence the radius of any ball containing S must be at least

$$||\varphi(i)q_i||/2 = (2 + \sqrt{2}(1 + 2/d^{1/3}))/2 > 1 + \frac{1}{\sqrt{2}}.$$

On the other hand, if $\sigma_i = 0$, then $\varphi(i) \notin S$. By Lemma 4.1, all points in $\varphi(\sigma)$ are within distance $\sqrt{2}(1+2/d^{1/3})$ from $-\varphi(i)$. Since, $||q_i(-\varphi(i))|| = \sqrt{2}(1 + 2/d^{1/3})$, the radius of MEB$(S)$ is at most $\sqrt{2}(1+2/d^{1/3})$. In this case, the radius of any α-MEB(S) is at most $\alpha \cdot r(\text{MEB}(S) \leq (1 + \frac{1}{\sqrt{2}})$, for $\alpha < \frac{1+\sqrt{2}}{2}(1 - 2/d^{1/3})$. Since $\mathbb{A}$ outputs a correct α-MEB(S) with probability at least 2/3, Bob can determine σ_i with the same probability. We can thus conclude that the workspace of $\mathbb{A}$ is $\Omega(\exp(d^{1/3}))$ and obtain the following.

THEOREM 4.2. *Any streaming algorithm that maintains an* α-MEB(S) *of a set* S *of* n *points in* $\mathbb{R}^d$*, for* $\alpha < \frac{1+\sqrt{2}}{2}(1 - 2/d^{1/3})$*, with probability at least* 2/3 *requires* $\Omega(\min\{n, \exp(d^{1/3})\})$ *bits of storage.*

Coreset for minimum enclosing ball. The reduction to α-MEB(S) can be modified to obtain lower bounds on α-coreset. Let $\mathbb{A}$ be an algorithm that maintains a α-coreset(S) for $\alpha \leq \sqrt{2}(1 - 2/d^{1/3})$ with probability 2/3. Let φ be a map as defined previously. Alice passes points in $\varphi(\sigma)$ as input to $\mathbb{A}$ in an arbitrary order and communicates the working space of $\mathbb{A}$ to Bob. For index i, q_i is the point as defined above; see Figure 2(b). Let C be α-coreset maintained by $\mathbb{A}$. For index i, Bob adds q_i as input to $\mathbb{A}$ and checks whether $\varphi(i) \in C$. If $\varphi(i) \in C$, Bob reports $\sigma_i = 1$. Otherwise, Bob reports $\sigma_i = 0$.

The correctness of this reduction follows from the following lemma.

LEMMA 4.3. *If* C *is an* α-coreset$(\varphi(\sigma) \cup \{q_i\})$*, for* $\alpha < \sqrt{2}(1 - 2/d^{1/3})$ *then,*

(i) $q_i \in C$*, and*

(ii) $\varphi(i) \in C$ *if and only if* $\sigma_i = 1$.

Proof. Let $S = \varphi(\sigma) \cup \{q_i\}$. First, we prove (i). Suppose, on the contrary, that $q_i \notin C$ and hence $C \subseteq \varphi(\sigma)$. Let $\mathbb{B}^d$ be the unit ball centered at origin. Let β be MEB(C). Since $C \subset \mathbb{S}^{d-1}$, $r(\beta) \leq 1$ and

$c(\beta) \in \mathbb{B}^d$. Observe that, for any point $t \in \mathbb{B}^d$, $||tq_i|| \geq \sqrt{2}(1 + 2/d^{1/3})$. Hence,

$$\begin{aligned} ||c(\beta)q_i|| &\geq \sqrt{2}(1 + 2/d^{1/3}) \\ &\geq \sqrt{2}(1 + 2/d^{1/3})r(\beta) \\ &> \alpha \cdot r(\beta), \end{aligned}$$

leading to a contradiction that C is an α-coreset(S).

Now, we prove (ii). If $\sigma_i = 0$, then $\varphi(i) \notin S$ and $C \subseteq S$ implies $\varphi(i) \notin C$. On the other hand, suppose $\sigma_i = 1$ but $\varphi(i) \notin C$. Let $\beta = \mathrm{MEB}(C)$. Observe that all points in C are within a distance $\sqrt{2}(1+2/d^{1/3})$ from $-\varphi(i)$. Hence $r(\beta) \leq \sqrt{2}(1 + 2/d^{1/3})$. It follows from (i) that the center $c(\beta) \in \beta^*$, where $\beta^* = \mathbb{B}(q_i, \sqrt{2}(1 + 2/d^{1/3}))$. For any point $p \in \beta^*$, $||p\varphi(i)|| \geq 2$, therefore

$$\begin{aligned} ||c(\beta)\varphi(i)|| &\geq 2 \\ &\geq \sqrt{2}(1 - 2/d^{1/3})\sqrt{2}(1 + 2/d^{1/3}) \\ &\geq \alpha \cdot r(\beta), \end{aligned}$$

for $\alpha < \sqrt{2}(1 - 2/d^{1/3})$, leading to a contradiction that C is an α-coreset(S). □

Hence, we conclude the following.

Theorem 4.3. *Any streaming algorithm that maintains an* α-coreset(S) *of a set* S *of* n *points in* $\mathbb{R}^d$*, for* $\alpha < \sqrt{2}(1-2/d^{1/3})$*, with probability at least* $2/3$ *requires* $\Omega(\min\{n, \exp(d^{1/3})\})$ *bits of storage.*

Width. Let B be a ball centered at origin with $r(B) = 1/2$. For any vector $u \in \mathbb{S}^{d-1}$, let h_u be a hyperplane passing through the origin and normal to u, i.e., $\langle x, u\rangle = 0$. Let $N_u \subset h_u$ be a set of d points such that $\mathrm{conv}(N_u \cup \{-u, u\})$ contains B; see Figure 3. For any $\sigma \in \{0,1\}^k$, let $-\varphi(\sigma) = \{-p \mid p \in \varphi(\sigma)\}$.

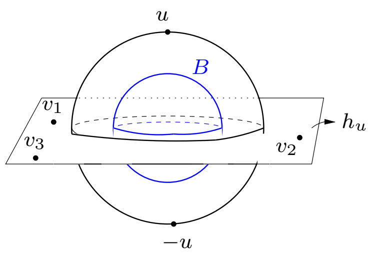

Figure 3: $N_u = \{v_1, v_2, v_3\}$, $B \subset \mathrm{conv}(N_u \cup \{-u, u\})$.

The Index problem can be reduced to α-width(S) as follows. Let $\mathbb{A}$ be a streaming algorithm that maintains α-width(S), for $\alpha < d^{1/3}/8$, with probability at least $2/3$. Alice passes points in $\{\varphi(\sigma) \cup -\varphi(\sigma)\}$ to $\mathbb{A}$ in an arbitrary order and then communicates the working space of $\mathbb{A}$ to Bob. For index i, Bob adds $N_{\varphi(i)}$ to $\mathbb{A}$. If $\mathbb{A}$ reports a slab J such that $\mathrm{d}(J)$ is at least 1, then Bob reports $\sigma_i = 1$ and otherwise Bob reports $\sigma_i = 0$.

The correctness of the reduction follows from the following observation. Let $S = \{\varphi(\sigma) \cup -\varphi(\sigma) \cup N_{\varphi(i)}\}$. If $\sigma_i = 1$, then $\{N_{\varphi(i)} \cup \{-\varphi(i), \varphi(i)\}\} \subseteq S$. By construction of $N_{\varphi(i)}$, $B \subset \mathrm{conv}(N_{\varphi(i)} \cup \{\varphi(i), -\varphi(i)\})$. Since $r(B) = 1/2$, $\mathrm{d}(J) \geq \mathrm{d}(\mathrm{width}(S)) \geq 1$. On the other hand, if $\sigma_i = 0$, then $(-\varphi(i), \varphi(i)) \notin S$. Let W be a slab bounded by the two hyperplanes $\langle x, \varphi(i)\rangle = 2/d^{1/3}$ and $\langle x, \varphi(i)\rangle = -2/d^{1/3}$. By Lemma 4.1 and the fact that $N_{\varphi(i)} \subset h_{\varphi(i)} \subset W$, we have $S \subset W$ and $\mathrm{d}(\mathrm{width}(S)) \leq 4/d^{1/3}$. For $\alpha < d^{1/3}/8$,

$$\mathrm{d}(J) \leq \alpha \cdot \mathrm{d}(\mathrm{width}(S)) \leq \alpha \cdot 4/d^{1/3} \leq 1/2.$$

Hence, we conclude the following.

Theorem 4.4. *Any streaming algorithm that maintains an* α-width(S) *of a set of* n *points in* $\mathbb{R}^d$*, for* $\alpha < d^{1/3}/8$*, with probability at least* $2/3$*, requires* $\Omega(\min\{n, \exp(d^{1/3})\})$ *bits of storage.*

5 Conclusion

In this paper, we design streaming algorithms for maintaining several extent measures on high-dimensional point sets. Our approach is based on the notion of blurred ball cover that, for a stream S of points, maintains a subset $K \subseteq S$. K can be interpreted as the union of a set of $\{K_1, \ldots, K_u\}$, $u = O((1/\varepsilon^2)\log(1/\varepsilon))$, of subsets of S each of size $O(1/\varepsilon)$ so that S lies in the union of $(1+\varepsilon)\mathrm{MEB}(K_i)$. We provide a simple streaming algorithm for maintaing blurred ball cover. We then use blurred ball cover to provide streaming algorithms for various extent measures. We conclude by stating related problems.

- Is there a fully dynamic data structure, whose size is linear in n and $\mathrm{poly}(d, 1/\varepsilon)$, that maintains a $(1+\varepsilon)$-MEB(S)?
- Can the concept of blurred ball cover be generalized for maintaining approximate smallest encosing convex shape of a high dimensional point set, e.g., minimum enclosing ellipsoids, under the streaming model?

References

[1] P. K. Agarwal, S. Har-Peled, and K. R. Varadarajan, Approximating extent measures of points, *J. ACM*, 51 (2004), 606–635.

[2] P. K. Agarwal, S. Har-Peled, and K. R. Varadarajan, Geometric approximation via coresets, in: *Combinatorial and Computational Geometry* (J. Goodman, J. Pach, and E.Welzl, eds.), Cambridge University Press, 2005, pp. 1–30.

[3] P. K. Agarwal, J. Matoušek, and S. Suri, Farthest neighbors, maximum spanning trees and related problems in higher dimensions, *Comput. Geom. Theory Appl.*, 1 (1992), 189–201.

[4] P. K. Agarwal and H. Yu, A space-optimal data-stream algorithm for coresets in the plane, *Proc. 23rd Annual Sympos. Comput. Geom.*, 2007, pp. 1–10.

[5] C. Aggarwal, *Data Streams: Models and Algorithms*, Springer, 2007.

[6] N. Alon, Y. Matias, and M. Szegedy, The space complexity of approximating the frequency moments, *J. Comput. Syst. Sci.*, 58 (1999), 137–147.

[7] A. Andoni, D. Croitoru, and M. Patrascu, Hardness of nearest neighbor under L-infinity, *Proc. 49th Annual IEEE Sympos. Foundations of Comp. Sc.*, 2008, pp. 424–433.

[8] K. Ball, An elementary introduction to modern convex geometry, *in Flavors of Geometry*, 31 (1997), 1–58.

[9] M. Bădoiu and K. L. Clarkson, Optimal core-sets for balls, *Comput. Geom. Theory Appl.*, 40 (2008), 14–22.

[10] M. Bădoiu, S. Har-Peled, and P. Indyk, Approximate clustering via core-sets, *Proc. 34th Annual ACM Sympos. Theory Comput.*, 2002, pp. 250–257.

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

[12] T. M. Chan, Faster core-set constructions and data-stream algorithms in fixed dimensions, *Comput. Geom. Theory Appl.*, 35 (2006), 20 – 35.

[13] B. Chazelle, H. Edelsbrunner, M. Grigni, L. Guibas, M.Sharir, and E. Welzl, Improved bounds on weak ε-nets for convex sets, *Discrete Comput. Geom.*, 13 (1995), 1–15.

[14] B. Chazelle and J. Matoušek, On linear-time deterministic algorithms for optimization problems in fixed dimension, *J. Algorithms*, 21 (1996), 579 – 597.

[15] K. L. Clarkson, Las Vegas algorithms for linear and integer programming when the dimension is small, *J. ACM*, 42 (1995), 488–499.

[16] K. L. Clarkson, Coresets, sparse greedy approximation, and the Frank-Wolfe algorithm, *Proc. 19th Annual ACM-SIAM Sympos. Discrete Algorithms*, 2008, pp. 922–931.

[17] K. L. Clarkson and P. W. Shor, Applications of random sampling in computational geometry, II., *Discrete Comput. Geom.*, 2 (1987), 195–222.

[18] K. L. Clarkson and D. P. Woodruff, Numerical linear algebra in the streaming model, *Proc. 41st Annual ACM Sympos. Theory of Comput.*, 2009, pp. 205–214.

[19] B. Gärtner and M. Jaggi, Coresets for polytope distance, *Proc. 25th Annual Sympos. Comput. Geom.*, 2009, pp. 33–42.

[20] A. Goel, P. Indyk, and K. Varadarajan, Reductions among high dimensional proximity problems, *Proc. 12th Annual ACM-SIAM Sympos. Discrete Algorithms*, 2001, pp. 769–778.

[21] S. Har-Peled and K. R. Varadarajan, Projective clustering in high dimensions using core-sets, *In Proc. 18th Annual Sympos. Comput. Geom*, 2003, pp. 312–318.

[22] P. Indyk, Better algorithms for high-dimensional proximity problems via asymmetric embeddings, *Proc. 14th Annual ACM-SIAM Sympos. Discrete Algorithms*, 2003, pp. 539–545.

[23] P. Kumar, J. S. B. Mitchell, and E. A. Yildirim, Approximate minimum enclosing balls in high dimensions using core-sets, *J. Exp. Algorithmics*, 8 (2003), 1.1.

[24] P. Kumar and E. A. Yildirim, Minimum-Volume Enclosing Ellipsoids and Core Sets, *J. Optimization Theory and Appl.*, 126 (2005), 1–21.

[25] E. Kushilevitz and N. Nisan, *Communication Complexity*, Cambridge University Press, 1997.

[26] J. Matoušek, M. Sharir, and E. Welzl, A subexponential bound for linear programming, *Algorithmica*, 16 (1996), 498 – 516.

[27] S. Muthukrishnan, *Data Streams: Algorithms and Applications*, Now Publishers Inc., 2005.

[28] E. A. Ramos, An Optimal Deterministic Algorithm for Computing the Diameter of a three-dimensional point set, *Discrete Comput. Geom.*, 26 (2001), 233–244.

[29] H. Zarrabi-Zadeh and T. Chan, A simple streaming algorithm for minimum enclosing balls, *Proc. 18th Annual Canadian Conf. Comput. Geom.*, 2006, pp. 139–142.