Improved Range Searching Lower Bounds

Kasper Green Larsen*
MADALGO†
Department of Computer Science
Aarhus University, Aarhus, Denmark
larsen@cs.au.dk

Huy L. Nguyễn‡
Center for Computational Intractability
Department of Computer Science
Princeton University, Princeton, NJ, USA
hlnguyen@cs.princeton.edu

ABSTRACT

In this paper we present a number of improved lower bounds for range searching in the pointer machine and the group model. In the pointer machine, we prove lower bounds for the approximate simplex range reporting problem. In approximate simplex range reporting, points that lie within a distance of $\varepsilon \cdot \operatorname{diam}(s)$ from the border of a query simplex s, are free to be included or excluded from the output, where $\varepsilon \geq 0$ is an input parameter to the range searching problem. We prove our lower bounds by constructing a hard input set and query set, and then invoking Chazelle and Rosenberg's [CGTA'96] general theorem on the complexity of navigation in the pointer machine.

For the group model, we show that input sets and query sets that are hard for range reporting in the pointer machine (i.e. by Chazelle and Rosenberg's theorem), are also hard for dynamic range searching in the group model. This theorem allows us to reuse decades of research on range reporting lower bounds to immediately obtain a range of new group model lower bounds. Amongst others, this includes an improved lower bound for the fundamental problem of dynamic d-dimensional orthogonal range searching, stating that $t_q t_u = \Omega((\lg n/\lg\lg n)^{d-1})$. Here t_q denotes the query time and t_u the update time of the data structure. This is an improvement of a $\lg^{1-\delta} n$ factor over the recent lower bound of Larsen [FOCS'11], where $\delta > 0$ is a small constant depending on the dimension.

*Supported in part by MADALGO and in part by a Google Fellowship in Search and Information Retrieval. In addition, this research was done while the author was visiting Princeton University. This visit was supported by an EliteForsk travel scholarship from the Danish Ministry of Science and by a grant from Aarhus University's Research Foundation.

†Center for Massive Data Algorithmics, a Center of the Danish National Research Foundation.

‡Supported in part by CCF 0832797 and a Gordon Wu fellowship.

SCG'12, June 17–20, 2012, Chapel Hill, North Carolina, USA.

Categories and Subject Descriptors

F.2.2 [**Nonnumerical Algorithms and Problems**]: Geometrical problems and computations

General Terms

Theory

Keywords

Range searching, lower bounds, pointer machine, group model

1. INTRODUCTION

Range searching is one of the most fundamental topics in computational geometry. Here the goal is to represent an input set of geometric objects, i.e. maintain a data structure, such that given a query range, one can efficiently compute various statistics over the input objects intersecting the query. The input objects are typically points in d-dimensional space and the most common query ranges are axis-aligned rectangles, simplices, halfspaces and balls. The type of information computed when answering a query includes counting the number of input objects in the intersection (range counting), reporting the k objects in the intersection (range reporting), determining whether the intersection is empty (range emptiness) and computing the sum of a set of weights assigned the input objects in the intersection (weighted range counting and range searching in the semi-group and group model).

In this paper we present improved lower bounds for a number of range searching problems. In the pointer machine model, we present the first lower bounds for approximate simplex range reporting. Secondly, we present a new powerful theorem for proving lower bounds for dynamic range searching in the group model. This theorem allows us to reuse decades of research in pointer machine range reporting lower bounds to immediately obtain a range of improved group model lower bounds.

1.1 Range Reporting in the Pointer Machine

The pointer machine model was introduced in 1979 by Tarjan [18]. In this model, a range reporting data structure is represented by a directed graph. Each node of the graph may store either an input object or some auxiliary data. The nodes have constant out-degrees and one node is designated as the root. When answering a query, the data structure starts by reading the root node. The data structure then examines the contents of that node and if the node stores an input object, the data structure may choose to report that object if it intersects the query range. Following that, it either terminates or selects an edge leaving the root and retrieves the node pointed to by that edge. This process continues, where at each step, the data structure selects an edge leaving one of the previously seen nodes and retrieves the node pointed to by that edge. When the process terminates, we require that all input objects that intersects the query range have been reported, i.e. each reported input object must be stored in at least one of the explored nodes. Thus a data structure in the pointer machine model, is a data structure where all memory accesses are through pointers and random accesses are disallowed.

The space of a pointer machine data structure is defined as the number of nodes in the corresponding graph, and the query time is the number of nodes explored when answering a query.

While the pointer machine model is somewhat constrained compared to the popular word-RAM model, there are several motivations for studying the complexity of range searching in this model. First and foremost, we can prove polynomially high and often very tight lower bounds in this model. This stands in sharp contrast to the highest query time lower bound for any static data structure problem in the word-RAM (or cell probe model), which is a mere $\Omega(\lg n/\lg\lg n)$, see e.g. [17]. Additionally, most word-RAM range reporting upper bounds are really pointer-based, or can easily be implemented without random accesses with a small overhead, typically at most an $O(\lg n)$ multiplicative cost in the query time and/or space. Thus pointer machine lower bounds indeed shed much light on the complexity of range reporting problems.

Previous Results.

With only a few exceptions (see [1, 3]), lower bounds for range reporting in the pointer machine model have all been proved by appealing to a theorem initially due to Chazelle [5] and later refined by Chazelle and Rosenberg [10]. Since our lower bounds also rely on this theorem, we introduce it in the following: First, let P be a set of input objects to a range searching problem and let $\mathcal{R}$ be a set of query ranges. Then we say that $\mathcal{R}$ is (t,h)-*favorable* if

1. $|R \cap P| \geq t$ for all $R \in \mathcal{R}$.
2. $|R_1 \cap R_2 \cap \cdots \cap R_h \cap P| = O(1)$ for all sets of h different queries $R_1, \ldots, R_h \in \mathcal{R}$.

Letting k denote the output size of a query, Chazelle and Rosenberg proved that

Theorem 1 (Chazelle and Rosenberg [10]). *Let P be a set of n input objects to a range searching problem and $\mathcal{R}$ a set of m query ranges. If $\mathcal{R}$ is $(\Omega(t_q),h)$-favorable, then any pointer machine data structure for P with query time $t_q + O(k)$ must use space $\Omega(mt_q/h)$.*

This theorem completely reduces the task of proving pointer machine range reporting lower bounds to a geometric problem of constructing a hard input and query set. In the case of simplex range reporting, Chazelle and Rosenberg [10] used this theorem to prove that any pointer machine data structure with query time $t_q + O(k)$ for d-dimensional simplex range reporting must use space $\Omega((n/t_q)^{d-\varepsilon})$, where $\varepsilon > 0$ is an arbitrarily small constant. Thus to obtain a query time as inefficient as $O(\sqrt{n}+k)$ for simplex range reporting in $\mathbb{R}^{10}$, one needs space almost $\Omega(n^5)$. This is extremely prohibitive for most natural applications.

Because of the high space/time requirement for most exact range reporting problems, researchers also look at the approximate version of the question: for a query range s, if an input object is within distance $\varepsilon \cdot \text{diam}(s)$ from the boundary of s, the data structure has the freedom to arbitrarily include or exclude it from the output. There is an extensive literature in this area but the result most relevant to us, is the data structure by Arya, da Fonseca, and Mount [4] for approximate simplex range searching in the semi-group model. Their data structure answers queries in $O(\lg n+\lg(1/\varepsilon)+1/\varepsilon^{d-1-u})$ time and uses $O(n/\varepsilon^{\frac{u^2 d}{d-1}})$ space for any integer $u \geq 0$. While their data structure is for semi-group range searching, they briefly describe a data structure for reporting with the same space and query time, except for an additional linear term in the number of reported points. We are not aware of any previous lower bounds for approximate range reporting for any kind of ranges despite the long history of the problem.

Since favorable query sets play an important role in our group model results, we also mention a number of previous favorable query set constructions. All these results were used to prove pointer machine range reporting lower bounds by invoking Theorem 1. Chazelle [5] (or alternatively [6]), showed that one can construct a $(t,2)$-favorable set of $\Theta(n/t \cdot (\lg n/\lg t)^{d-1})$ queries for *orthogonal range reporting* in d-dimensional space. Henceforth, $t = \Omega(1)$ is an adjustable parameter. We note that the input to orthogonal range reporting is a set of n points and the queries are axis-aligned rectangles. For the "dual" problem of d-dimensional *rectangle stabbing* (the input is n axis-aligned rectangles and a query asks to report all rectangles containing a query point), Afshani, Arge and Larsen [2] showed that one can construct a $(t,2)$-favorable set of $n/t \cdot 2^{\Theta\left(\frac{\lg n}{t^{1/(d-1)}}\right)}$ queries. For *line range reporting* (the input consists of n two-dimensional points and a query asks to report all points on a query line), a classical construction of Erdös (see e.g. [14]) shows that one can construct a $(\Theta(n^{1/3}),2)$-favorable set of n query lines. Finally, for *convex polytope intersection reporting* in $\mathbb{R}^3$ (the input is an n-vertex convex polytope and a query asks to report all edges of the polytope intersecting a query plane), Chazelle and Liu [8] showed that one can construct a $(t,2)$-favorable set of $\Theta(n^2/t^3)$ query planes.

1.2 Range Searching in the Group Model

The group model was introduced in 1982 by Fredman [11]. In this model, each input object to a range searching problem is assigned a weight from a commutative group and the goal is to preprocess the input into a collection of precomputed group elements, such that one can efficiently compute the group sum of the weights assigned to all input objects intersecting a query range. In this paper, we focus on dynamic range searching in the group model. Here a data structure must also support updating the weights assigned to the input objects (while the set of input objects is fixed).

Since the group model was first introduced, there has been two slightly different definitions of dynamic data structures in the group model. These types of data structures have been named *weakly oblivious* and *oblivious* data structures, respectively. The lower bounds we prove apply to *oblivious* data structures:

Oblivious Data Structures.

An *oblivious* data structure in the group model (see [11] or [13]), is a dynamic data structure with no understanding of the particular group in question, i.e. it can only access and manipulate weights through black-box addition and subtraction. Thus from the data structure's point of view, each precomputed group element is just a linear combination over the weights assigned to the input objects. When answering a query, such a data structure adds and subtracts a subset of these linear combinations (another linear combination) to finally yield the linear combination summing exactly the weights assigned to the input objects intersecting the query range. When given an update request, the data structure simply re-evaluates every precomputed group element for which the weight of the updated input object occurs with non-zero coefficient in the corresponding linear combination.

We define the query time of an oblivious data structure as the number of precomputed group elements used when answering a query, and the update time is defined as the number of linear combinations that need to be re-evaluated when updating the weight of an input object. For a more formal definition, see Section 2.

We note that weakly oblivious data structures differ from oblivious data structures, in that they are allowed slightly more elaborate update procedures and therefore lower bounds proved for weakly oblivious data structures also apply to oblivious data structures. For a definition of weakly oblivious data structures, we refer the reader to [13].

Previous Results.

Proving high lower bounds in the group model has until very recently remained a great barrier. When Fredman [11] defined the model, he proved a lower bound of $\Omega(n \lg n)$ over a sequence of n updates and queries to the *partial sums* problem (which is the special case of 1-d orthogonal range searching where the points have coordinates $0, \dots, n-1$). This bound holds for oblivious data structures. Following that, Fredman and Saks [12] proved an $\Omega(n \lg n / \lg\lg n)$ lower bound for the same problem, which does however hold for weakly oblivious data structures.

Chazelle [6, 7] later proved lower bounds for *offline* range searching in the group model. He also proved a lower bound of $\Omega(n \lg n)$ for offline two-dimensional *halfspace range searching* [7]. The input to the offline halfspace range searching problem is a set of n query halfspaces and n input points, each assigned a weight from a commutative group, and the goal is to compute for every query halfspace, the group sum of the weights assigned to the points contained therein. In [6] he considered offline two-dimensional orthogonal range searching and proved a lower bound of $\Omega(n \lg\lg n)$. Both these lower bounds were established using a general theorem for proving lower bounds for offline range searching in the group model: Letting A denote the *incidence matrix* corresponding to the input set of points and queries (i.e. the matrix with one row for each query R_i and one column for each point p_j, such that entry $a_{i,j}$ is 1 if $p_j \in R_i$ and it is 0 otherwise), Chazelle showed that for any $1 \le k \le n$, if λ_k denotes the k'th largest eigenvalue of A, then the offline problem requires $\Omega(k \cdot \lg(\lambda_k))$ time [7]. Thus proving offline group model lower bounds was reduced to constructing input and query sets where the incidence matrix has large eigenvalues.

The next step forward was by Pătraşcu and Demaine [16], who improved the lower bound of Fredman and Saks to $\Omega(n \lg n)$, i.e. they matched the initial bound of Fredman, but also for weakly oblivious data structures. Following that, Pătraşcu [15] considered two-dimensional orthogonal range searching and proved an $\Omega(\lg n / \lg(s/n + \lg n))$ query time lower bound for the static case. Here s denotes the space usage of the data structure. He also showed how to extend this static lower bound to give a query time lower bound of $\Omega((\lg n / \lg(\lg n + t_u))^2)$ for dynamic data structures with update time t_u. This bound holds for weakly oblivious data structures.

None of the above bounds exceed $\Omega((\lg n / \lg\lg n)^2)$ per query, and this barrier was not overcome until Larsen's [13] recent results. Larsen moved past this barrier by introducing one additional restriction on oblivious data structures: Recall that an oblivious data structure stores group elements corresponding to linear combinations over the weights assigned to the input objects, and it answers queries by again computing linear combinations over the precomputed group elements. Larsen defined the *multiplicity* of an oblivious data structure as the largest absolute value of any coefficient in these linear combinations, and noted that all known upper bounds use only coefficients amongst $\{-1, 0, 1\}$, i.e. they have multiplicity 1. With this definition, he demonstrated a connection between oblivious data structures and the *discrepancy* of the corresponding range searching problem. The discrepancy, disc, of a range searching problem with query ranges $\mathcal{R}$, is the maximum over all sets P of n input objects, of the deviation of the best 2-coloring of the objects in P from an even coloring, i.e.

$$\mathrm{disc} = \max_{P:|P|=n} \min_{\chi:P\to\{-1,+1\}} \max_{R\in\mathcal{R}} \Big|\sum_{p\in R\cap P} \chi(p)\Big|.$$

More precisely, Larsen's main theorem states that $t_q t_u = \Omega(\mathrm{disc}^2 / \Delta^4 \lg n)$, where t_q is the worst case query time, t_u the worst case update time and Δ is the multiplicity of the data structure. Plugging in the vast amount of discrepancy lower bounds, this relation immediately yielded a whole range of polynomially high group model lower bounds for data structures with constant multiplicity. The two bounds most important to our work (stated here for constant multiplicity), is a lower bound of $t_q t_u = \Omega(n^{1/3} / \lg n)$ for line range searching and a bound of $t_q t_u = \Omega(\lg^{d-2+\mu(d)} n)$ for d-dimensional orthogonal range searching, where $\mu(d) > 0$ is a small but strictly positive function of d.

1.3 Our Results

In the following two paragraphs, we present our new results for range searching in the group model and approximate range reporting in the pointer machine.

Our Group Model Results.

In Section 2, we present an alternative to Larsen's connection to discrepancy theory. Our relation allows us to almost immediately translate range reporting lower bounds in the pointer machine to dynamic group model lower bounds for data structures with bounded multiplicity. More specifically, let P be a set of n input objects to a range searching problem and $\mathcal{R}$ a set of m query ranges over P. We say that $\mathcal{R}$ is *strongly* (t, h)*-favorable* if

1. $|R \cap P| = \Theta(t)$ for all $R \in \mathcal{R}$.

2. $|R_1 \cap R_2 \cap \cdots \cap R_h \cap P| = O(1)$ for all sets of h different queries $R_1, \ldots, R_h \in \mathcal{R}$.

3. $|\{R \in \mathcal{R} \mid p \in R\}| = O(mt/n)$ for all $p \in P$.

Thus a favorable query set is strongly favorable, if in addition, all query ranges contain roughly equally many input objects and all input objects are contained in roughly equally many query ranges. With this definition, we prove the following result

THEOREM 2. *Let P be a set of n input objects to a range searching problem and $\mathcal{R}$ a set of $m \leq n$ query ranges over P. If $\mathcal{R}$ is strongly $(t, 2)$-favorable, then any oblivious data structure for the range searching problem must have $t_q t_u = \Omega(m^2 t / n^2 \Delta^4)$ on the input set P. Here t_q denotes the worst case query time, t_u the worst case update time and Δ the multiplicity of the data structure. For $m = \Theta(n)$ and $\Delta = O(1)$, this bound simplifies to $t_q t_u = \Omega(t)$.*

Fortunately, all the favorable query sets described in Section 1.1 are also strongly favorable. By adjusting the parameter t (in Section 1.1) such that the number of queries in the favorable query sets is $m = \Theta(n)$ and $m \leq n$, we immediately obtain the following lower bounds (listed here for constant multiplicity):

For d-dimensional orthogonal range searching, we get a lower bound of

$$t_q t_u = \Omega((\lg n / \lg\lg n)^{d-1}).$$

Compared with the result of Larsen [13], this is an improvement of a $\lg^{1-\mu(d)} n / (\lg\lg n)^{d-1} = \Omega(\lg^{1-\delta} n)$ factor, where $\delta > 0$ is some small constant. For rectangle stabbing, we get a matching bound either by a reduction or from the construction of Afshani, Arge and Larsen [2]. This again improves over Larsen's result [13] by a $\lg^{1-\delta} n$ factor. For line range searching, we get a lower bound of $t_q t_u = \Omega(n^{1/3})$, which is an improvement of a $\lg n$ factor over [13]. Additionally, we get a lower bound for convex polytope intersection searching in $\mathbb{R}^3$ of $t_q t_u = \Omega(n^{1/3})$ using the result of Chazelle and Liu [8].

Our proof of Theorem 2 is based on carefully bounding the eigenvalues of the incidence matrix corresponding to P and $\mathcal{R}$. In fact, we prove the following stronger theorem during our establishment of Theorem 2:

THEOREM 3. *Let P be a set of n input objects to a range searching problem, $\mathcal{R}$ a set of m query ranges over P and A the corresponding incidence matrix. Then for every $3 \leq k \leq n$, any oblivious data structure for the range searching problem must have $t_q t_u = \Omega(\lambda_k k^2 / mn\Delta^4)$ on the input set P. Here λ_k denotes the k'th largest eigenvalue of $A^T A$, t_q the worst case query time of the data structure, t_u the worst case update time and Δ the multiplicity of the data structure.*

This theorem can be considered a complement to Chazelle's theorem [7] for establishing lower bounds on offline range searching in the group model, however our dependence on λ_k is exponentially better than the one of Chazelle. Using this theorem, we also obtain a lower bound of $t_q t_u = n^{\Omega(1/\lg\lg n)}$ for orthogonal range searching in non-constant dimension $d = \Omega(\lg n / \lg\lg n)$ (see Section 2 for a proof).

Our Pointer Machine Results.

In Section 3, we give the first lower bound for approximate range reporting in the pointer machine model. Specifically, we show that for any fixed constant $\delta > 0$, any data structure for approximate simplex range reporting that answers a simplex query s in $t_q + O(k)$ time, where k is the number of points either inside s or within a distance $\varepsilon \cdot \mathrm{diam}(s)$ from the boundary of s, needs space at least $n\varepsilon^{1+\delta-d}/t_q^{1+\delta}$. This bound matches the upper bound from [4] up to a factor $\varepsilon^{\delta d/(1+\delta)}$ when the space is $\Theta(n)$, but unfortunately there is still a rather large gap between the two bounds in the rest of the tradeoff.

2. THE GROUP MODEL

In this section, we present the proofs of Theorem 2 and Theorem 3, relating range reporting in the pointer machine and dynamic range searching in the group model. Before starting our proof, we present an equivalent but more formal definition of oblivious data structures (see [13] or [11]):

Oblivious Data Structures.

An oblivious data structure for a range searching problem, is a factorization of each incidence matrix A, corresponding to a set n input objects P and a set of m query ranges $\mathcal{R}$, into two matrices Q and D such that $Q \cdot D = A$.

The *data matrix* $D \in \mathbb{Z}^{s \times n}$ represents the precomputed group sums stored by the data structure on input P. Each of the s rows is interpreted as a linear combination over the weights assigned to the n input objects, and we think of the data structure as maintaining the corresponding group sums when given an assignment of weights to the input objects.

The *query matrix* $Q \in \mathbb{Z}^{m \times s}$ specifies the query algorithm. It has one row for each query R in $\mathcal{R}$, and we interpret this row as a linear combination over the precomputed group elements, denoting which precomputed elements to add and subtract when answering the query R on input set P.

The *worst case query time* of an oblivious data structure is defined as the maximum number of non-zero entries in a row of Q over all input sets P and query sets $\mathcal{R}$. The *worst case update time* is similarly defined as the maximum number of non-zero entries in a column of D over all P. The *space* of the data structure is the maximum number of columns in Q (equivalently number of rows in D) over all P and $\mathcal{R}$. Finally, we define the *multiplicity* as the largest absolute value of an entry in D and Q over all P and $\mathcal{R}$.

We refer the reader to [13] or [11] for more intuition on this definition.

2.1 Proofs of Theorem 2 and Theorem 3

In light of the above definition, proving lower bounds for oblivious data structures boils down to arguing when an incidence matrix cannot be factored into two sparse matrices Q and D. The key insight here is, that for sparse matrices Q and D with bounded coefficients, the product QD must have small singular values (i.e. $(QD)^T QD$ has small eigenvalues). Thus if A has large singular values, Q and D cannot be sparse if $QD = A$. This is precisely the intuition behind our proof of Theorem 3:

Proof of Theorem 3.

Let P, $\mathcal{R}$ and A be as in Theorem 3. Furthermore, let $QD = A$ be the factorization of A provided by an oblivious data structure, where Q is an $m \times s$ matrix such that each row has at most t_q non-zero entries and where D is an $s \times n$ matrix where each column has at most t_u non-zero entries. Finally, let Δ be the multiplicity of the oblivious data structure, i.e. any coefficient in Q and D is bounded in absolute value by Δ. Now let $U(D)\Sigma(D)V(D)^T$ be the singular value decomposition of D. Here $U(D)$ and $V(D)$ are unitary matrices and $\Sigma(D)$ is a diagonal matrix where the diagonal entries equals the singular values of D, i.e. if we let $\gamma_i(D^TD) \geq 0$ denote the i'th largest eigenvalue of the $n \times n$ positive semi-definite matrix D^TD, then the i'th diagonal entry of $\Sigma(D)$ is $\sigma_{i,i}(D) = \sqrt{\gamma_i(D^TD)}$. Similarly, let $U(Q)\Sigma(Q)V(Q)^T$ be the singular value decomposition of Q. Letting $\gamma_i(Q^TQ) \geq 0$ denote the i'th largest eigenvalue of Q^TQ, we have that the i'th diagonal entry of $\Sigma(Q)$ is $\sigma_{i,i}(Q) = \sqrt{\gamma_i(Q^TQ)}$. Letting $d_{i,j}$ denote entry (i,j) in D, it now follows from D^TD being square and real that

$$\sum_i \gamma_i(D^TD) = \operatorname{tr}(D^TD) = \sum_{i,j} d_{i,j}^2 \leq t_u\Delta^2 n,$$

where we used that the coefficients of D are bounded in absolute value by Δ. Similarly, we have $\sum_i \gamma_i(Q^TQ) \leq t_q\Delta^2 m$. Finally since $\gamma_i(D^TD)$ and $\gamma_i(Q^TQ)$ are non-negative for all i, we conclude that $\gamma_{\lfloor k/2 \rfloor}(D^TD) = O(t_u\Delta^2 n/k)$ and $\gamma_{\lceil k/2 \rceil - 1}(Q^TQ) = O(t_q\Delta^2 m/k)$.

Our last step is to bound from above the eigenvalues of $(QD)^TQD$. Letting $\gamma_k((QD)^TQD)$ denote the k'th largest eigenvalue of $(QD)^TQD$, we get from the Courant-Fischer characterization of eigenvalues that

$$\gamma_k((QD)^TQD) = \min_{S:\dim(S)\geq n-k+1} \max_{x\in S:\|x\|_2=1} \|QDx\|_2^2,$$

i.e. $\gamma_k((QD)^TQD)$ equals the minimum over all subspaces S of $\mathbb{R}^n$ of dimension at least $n-k+1$, of the maximum square of the stretch of a unit length vector x when multiplying with QD. We thus aim to find a subspace S of dimension at least $n-k+1$, such that every unit vector in S is scaled as little as possible when multiplied with QD. We choose the subspace S consisting of all vectors x, for which $\langle v_i(D), x\rangle = 0$ for $i = 1, \dots, \lfloor k/2 \rfloor$ and $\langle v_i(Q), Dx\rangle = 0$ for $i = 1, \dots, \lceil k/2 \rceil - 1$. Here $\langle \cdot, \cdot \rangle$ denotes the standard inner product, $v_i(D)$ denotes the i'th column vector of $V(D)$ and $v_i(Q)$ denotes the i'th column vector of $V(Q)$. Clearly $\dim(S) \geq n-k+1$. Now let $x \in S$ be a unit length vector and consider first the product $Dx = U(D)\Sigma(D)V(D)^Tx$. Since x is orthogonal to the first $\lfloor k/2 \rfloor$ row vectors in $V(D)^T$ and since $\sigma_{i,i}(D) = O(\sqrt{t_u n/k}\Delta)$ for $i \geq \lfloor k/2 \rfloor$, we get that $\|\Sigma(D)V(D)^Tx\|_2 = O(\sqrt{t_u n/k}\Delta)$. Since $U(D)$ is unitary, this implies $\|Dx\|_2 = O(\sqrt{t_u n/k}\Delta)$. Finally, since Dx is orthogonal to the first $\lceil k/2 \rceil - 1$ row vectors of $V(Q)^T$, we conclude $\|QDx\|_2^2 = O(t_q t_u \Delta^4 mn/k^2)$. But $A^TA = (QD)^TQD$ and thus it must hold that $t_q t_u = \Omega(\lambda_k k^2/mn\Delta^4)$. This completes the proof of Theorem 3.

Proof of Theorem 2.

Let P and $\mathcal{R}$ be as in Theorem 2, i.e. $\mathcal{R}$ is a strongly $(t,2)$-favorable set of queries. Furthermore, let A be the $m \times n$ incidence matrix corresponding to P and $\mathcal{R}$, where $m \leq n$. Our proof is based on lower bounding the eigenvalues of $M = A^TA$, and then applying Theorem 3. We lower bound these eigenvalues using the following theorem of Chazelle and Lvov:

Theorem 4 (Chazelle and Lvov [9]). *Let A be an $m \times n$ real matrix where $m \leq n$ and let $M = A^TA$. Then M has at least*

$$\frac{n}{16\operatorname{tr}(M^2)n/9\operatorname{tr}(M)^2 - 7/9}$$

eigenvalues that are greater than or equal to $\operatorname{tr}(M)/4n$.

To use Theorem 4, we bound $\operatorname{tr}(M)$ and $\operatorname{tr}(M^2)$. The first is easily seen to be $\operatorname{tr}(M) = \sum_{R\in\mathcal{R}} |R| = \Omega(mt)$ and the latter is bounded by

$$\begin{aligned}
\operatorname{tr}(M^2) &= \sum_{R_1\in\mathcal{R}}\sum_{R_2\in\mathcal{R}} |R_1 \cap R_2|^2 \\
&= \sum_{R\in\mathcal{R}} |R|^2 + \sum_{R_1\in\mathcal{R}}\sum_{R_2\in\mathcal{R}|R_1\neq R_2} |R_1\cap R_2|^2 \\
&= O(mt^2) + \\
&\quad \sum_{p\in P}\sum_{R_1\in\mathcal{R}|p\in R_1}\sum_{R_2\in\mathcal{R}|p\in R_2\wedge R_1\neq R_2} |R_1\cap R_2| \\
&= O(mt^2) + \sum_{p\in P}\sum_{R_1\in\mathcal{R}|p\in R_1}\sum_{R_2\in\mathcal{R}|p\in R_2\wedge R_1\neq R_2} O(1) \\
&= O(mt^2) + \sum_{p\in P} O((mt/n)^2) \\
&= O(mt^2).
\end{aligned}$$

Plugging these values into Theorem 4, we conclude that $M = A^TA$ has $\Omega(nm^2t^2/mt^2n) = \Omega(m)$ eigenvalues greater than $\Omega(mt/n)$. Finally invoking Theorem 3, we get that $t_q t_u = \Omega(mt/n \cdot m^2/mn\Delta^4) = \Omega(m^2t/n^2\Delta^4)$, which completes the proof of Theorem 2.

Implications.

As already mentioned in Section 1.2, we obtain a number of lower bounds from Theorem 2 by reusing favorable query sets constructed for proving pointer machine range reporting lower bounds. Thus in the following, we only mention our proof of the lower bound for orthogonal range reporting in non-constant dimension $d = \Omega(\lg n/\lg\lg n)$.

For non-constant dimensions $d = \Omega(\lg n/\lg\lg n)$, Chazelle and Lvov [9] showed one can construct a set of n points and n query rectangles such that the corresponding incidence matrix A satisfies $\operatorname{tr}(A^TA) = n^{1+\Omega(1/\lg\lg n)}$ and $\operatorname{tr}((A^TA)^2) = O(\operatorname{tr}(A^TA)^2/n)$. From Theorem 4, this means that A^TA has $\Omega(n)$ eigenvalues that are greater than $n^{\Omega(1/\lg\lg n)}$. The result follows immediately from Theorem 3.

3. APPROXIMATE SIMPLEX RANGE REPORTING

In this section we show our lower bounds for approximate simplex range reporting in constant dimensions. The idea of the proof is as follows. From the result by Chazelle and Rosenberg [10], we know that thin *slab* (everything contained between two parallel hyperplanes) queries are hard for range reporting. Therefore, we consider a set of simplex queries that approximate thin slabs and then appeal to Theorem 1 and the properties of thin slabs shown in [10]. First, before proving the lower bound, we formally define the notion of approximations for range reporting.

Definition 1. *Given a query range R, define*

$$R^+ = \{x \mid \operatorname{dist}(x,R) \leq \varepsilon\operatorname{diam}(R)\}$$

and $R^- = \{x \in R | \operatorname{dist}(x, \partial R) \geq \varepsilon \operatorname{diam}(R)\}$ where ∂R is the boundary of R.

For any given query range R, a data structure for approximate range reporting should report all points inside R^- and none of the point outside R^+.

We now define the approximate analog of *favorable* query sets.

Definition 2. *Let P be a set of points and $\mathcal{R}$ a set of queries. Then $\mathcal{R}$ is (t, h)-approximate favorable for P if*

1. *The set $\{R^- | R \in \mathcal{R}\}$ is (t, h)-favorable for P.*
2. *For any $R \in \mathcal{R}$, $|P \cap R^+| = O(|P \cap R^-|)$.*

We get the following corollary of Theorem 1.

Corollary 1. *Let P be a set of points and $\mathcal{R}$ a (t_q, h)-approximate favorable set of queries for P. If for any query $R \in \mathcal{R}$, a data structure can report all points in $P \cap R^-$ and none outside $P \cap R^+$ in time $O(t_q + |R^+ \cap P|)$, then the space used by the data structure is $\Omega(|\mathcal{R}| t_q / h)$.*

Proof. Since $|R^+ \cap P| = O(R^- \cap P)$, the data structure satisfies the condition of Theorem 1 with respect to the query set $\{R^- | R \in \mathcal{R}\}$ and the corollary follows immediately. $\square$

Now we construct our set $\mathcal{R}$ of (t_q, h)-approximate favorable queries for d-dimensional approximate simplex range reporting. First select a set of $m = \varepsilon^{1+\delta-d}/t_q^{1+\delta}$ uniform random points in $[0,1]^{d-1}$, for an arbitrary fixed constant $\delta > 0$. Then map each point $(x_1, \ldots, x_{d-1})$ to $\Theta(1/\varepsilon)$ points using the map

$$(x_1, \ldots, x_{d-1}) \rightarrow 10\varepsilon d^3 i(x_1 + 1, \ldots, x_{d-1} + 1, 2)$$

where i ranges over integers such that $1/2 \leq 20\varepsilon d^3 i \leq 3/4$. For each new point p, consider the slab $H_{p,2\varepsilon d^2} = \{x \in [0,1]^d \big| \, |\langle x, p\rangle - |p|^2| \leq 2\varepsilon d^2 |p|\}$. The intersection of each such slab and $[0,1]^{d-1} \times \mathbb{R}$ forms a parallelotope I_p. Let $I'_p = H_{p,4\varepsilon d^2} \cap [-2\varepsilon d^2, 1 + 2\varepsilon d^2]^{d-1} \times \mathbb{R}$. Let $u_{p,0}$ be a vertex of I'_p and $u_{p,1}, \ldots, u_{p,d}$ be its neighboring vertices of I'_p. Let R_p be the simplex whose vertices are $u_{p,0}$ and $u_{p0} + d(u_{p,i} - u_{p,0})$ for $i = 1, \ldots, d$. The query set $\mathcal{R}$ is the set of all R_p's constructed above. Also define the parallelotope $O_p = H_{p,10\varepsilon d^3} \cap [-2d, 2d]^{d-1} \times \mathbb{R}$. First, we show some properties of the constructed queries.

Lemma 1. *For any query R_p in the query set $\mathcal{R}$, $I_p \subset R_p^- \subset R_p \subset R_p^+ \subset O_p$.*

Proof. First notice that the diameter of I'_p is bounded by $2d$ so the diameter of R_p is bounded by $2d^2$.

Now we prove the first containment. Notice that every vertex of I'_p can be written as a convex combination of vertices of R_p so $I'_p \subset R_p$. The distance from any point in I_p to $\partial I'_p$ is at least $2\varepsilon d^2 \geq \varepsilon \operatorname{diam}(R_p)$ so $I_p \subset R_p^-$.

Finally we prove the last containment. Since $I'_p \subset H_{p,4\varepsilon d^2} \cap [-2\varepsilon d^2, 1 + 2\varepsilon d^2]^{d-1} \times \mathbb{R}$, by construction, $R_p \subset H_{p,8\varepsilon d^3} \cap [-3d/2, 3d/2]^{d-1} \times \mathbb{R}$. Thus, $R_p^+ \subset H_{p,10\varepsilon d^3} \cap [-2d, 2d]^{d-1} \times \mathbb{R} \subset O_p$. $\square$

We have the following lemma from [10], adapted to our parameters.

Lemma 2. (Chazelle and Rosenberg [10] (Lemma 3.1)) *The constructed query set $\mathcal{R}$ satisfies*

1. *$\mathcal{R}$ is a set of size $\Theta(m/\varepsilon) = \Theta(\varepsilon^{\delta-d}/t_q^{1+\delta})$.*
2. *I_p and O_p have intersection with $[0,1]^d$ of volume $\Theta(\varepsilon)$ for any $R_p \in \mathcal{R}$.*
3. *For any $k = \lg m$ different O_p's, there exists a subset $\{O_{p_1}, O_{p_2}, \ldots, O_{p_d}\}$ of size d such that $O_{p_1} \cap \cdots \cap O_{p_d} \cap [0,1]^d$ has volume $O(\varepsilon^d m (\lg m)^{d-2}) = O(\varepsilon^{1+\delta} (\lg(1/\varepsilon))^{d-2}/t_q^{1+\delta})$.*

Applying this lemma, we show the following.

Lemma 3. *Consider $t_q = \omega(\lg(1/\varepsilon))$. Choose t_q/ε points $P = \{p_1, \ldots, p_{t_q/\varepsilon}\}$ independently and uniformly in $[0,1]^d$. With probability $1 - o(1)$, the query set $\mathcal{R}$ is $(t_q, O(\lg(1/\varepsilon)))$-approximate favorable for the point set P.*

Proof. Since the volume of the intersection between I_p and $[0,1]^d$ is $\Theta(\varepsilon)$, each point p_i is included in I_p with probability $\Theta(\varepsilon)$. By the Chernoff bound, with probability $1 - \exp(\Omega(t_q))$, the number of points in I_p is within a factor 2 of its expectation of $\Theta(t_q)$. Similarly with probability $1 - \exp(\Omega(t_q))$, the number of points in O_p is within a factor 2 of its expectation of $\Theta(t_q)$. Therefore, if $t_q = \omega(\lg(1/\varepsilon))$, then by the union bound, with probability $1 - o(1)$, $|I_p \cap P| = \Theta(t_q)$ and $|O_p \cap P| = \Theta(t_q)$ for all queries Q_p.

Next we apply the following lemma from [10].

Lemma 4. (Chazelle and Rosenberg [10] (Lemma 3.4)) *With probability $1 - o(1)$, for all distinct $p_1, \ldots, p_k$ where $k = \lceil \lg m \rceil = O(\lg 1/\varepsilon)$, we have $|P \cap O_{p_1} \cap \ldots O_{p_k} \cap [0,1]^d| = O(1)$.*

By the above lemma and the fact that $R_{p_i} \subset O_{p_i}$, we have $|P \cap R_{p_1}^- \cap \ldots \cap R_{p_k}^- \cap [0,1]^d| = O(1)$. Thus, with probability $1 - o(1)$, $\mathcal{R}$ is $(t_q, O(\lg(1/\varepsilon)))$-approximate favorable. $\square$

Theorem 5. *Consider a function $t_q = \omega(\lg(1/\varepsilon))$. For any data structure that can answer any query $R \in \mathcal{R}$ on the point set P in time $O(t_q + |R^+ \cap P|)$, the space it uses is at least $n\varepsilon^{1+\delta-d}/t_q^{1+\delta}$ for any arbitrary constant $\delta > 0$.*

Proof. Consider the disjoint union of $n\varepsilon/t_q$ hypercubes and the corresponding queries constructed above. The union of the query sets is still $(t_q, \lg(1/\varepsilon))$-approximate favorable. Thus, by corollary 1, the space needed is at least $n\varepsilon/t_q \cdot (\varepsilon^{\delta-d}/t_q^{1+\delta} \cdot t_q / \lg(1/\varepsilon)) = n\varepsilon^{1+\delta-d}/(t_q^{1+\delta} \lg(1/\varepsilon))$ for any $\delta > 0$. $\square$

4. CONCLUSION

In this paper, we presented a new theorem for proving lower bounds on dynamic range searching in the group model. This theorem allows us to almost immediately translate lower bounds for range reporting in the pointer machine to lower bounds in the group model. However, as in the recent result of Larsen [13], the lower bounds obtained are conditioned on data structures having low multiplicity. Echoing Larsen, we believe that proving polynomially high lower bounds, that are independent of the multiplicity, remains one of the most important open problems in the group model. This seems to require a radically different approach. A more modest goal would be to develop a technique that allows for different types of query time and update time tradeoffs than the linear tradeoffs presented here and in [13], even under the assumption of constant multiplicity.

We also presented the first lower bounds for approximate range reporting in the pointer machine. Tightening the gap between the upper and lower bound for approximate simplex range reporting remains an interesting open problem. We believe both are subject to improvement.

5. REFERENCES

[1] P. Afshani. Improved pointer machine and I/O lower bounds for simplex range reporting and related problems. In *Proc. 28th ACM Symposium on Computational Geometry*, 2012.

[2] P. Afshani, L. Arge, and K. D. Larsen. Orthogonal range reporting: Query lower bounds, optimal structures in 3d, and higher dimensional improvements. In *Proc. 26th ACM Symposium on Computational Geometry*, pages 240–246, 2010.

[3] P. Afshani, L. Arge, and K. G. Larsen. Higher-dimensional orthogonal range reporting and rectangle stabbing in the pointer machine model. In *Proc. 28th ACM Symposium on Computational Geometry*, 2012.

[4] S. Arya, G. D. da Fonseca, and D. M. Mount. Tradeoffs in approximate range searching made simpler. In *Proceedings of the XXI Brazilian Symposium on Computer Graphics and Image Processing (SIBGRAPI)*, pages 237–244, 2008.

[5] B. Chazelle. Lower bounds for orthogonal range searching: I. the reporting case. *Journal of the ACM*, 37(2):200–212, 1990.

[6] B. Chazelle. Lower bounds for off-line range searching. In *Proc. 27th ACM Symposium on Theory of Computation*, pages 733–740, 1995.

[7] B. Chazelle. A spectral approach to lower bounds with applications to geometric searching. *SIAM Journal on Computing*, 27:545–556, 1998.

[8] B. Chazelle and D. Liu. Lower bounds for intersection searching and fractional cascading in higher dimension. In *Proc. 33rd ACM Symposium on Theory of Computation*, pages 322–329, 2001.

[9] B. Chazelle and A. Lvov. A trace bound for the hereditary discrepancy. In *Proc. 16th ACM Symposium on Computational Geometry*, pages 64–69, 2000.

[10] B. Chazelle and B. Rosenberg. Simplex range reporting on a pointer machine. *Computational Geometry: Theory and Applications*, 5:237–247, January 1996.

[11] M. L. Fredman. The complexity of maintaining an array and computing its partial sums. *Journal of the ACM*, 29:250–260, January 1982.

[12] M. L. Fredman and M. Saks. The cell probe complexity of dynamic data structures. In *Proc 21st ACM Symposium on Theory of Computation*, pages 345–354, 1989.

[13] K. G. Larsen. On range searching in the group model and combinatorial discrepancy. In *Proc. 52nd IEEE Symposium on Foundations of Computer Science*, 2011.

[14] J. Pach and P. K. Agarwal. *Combinatorial geometry*. Wiley-Interscience series in discrete mathematics and optimization. Wiley, 1995.

[15] M. Pătraşcu. Lower bounds for 2-dimensional range counting. In *Proc. 39th ACM Symposium on Theory of Computation*, pages 40–46, 2007.

[16] M. Pătraşcu and E. D. Demaine. Logarithmic lower bounds in the cell-probe model. *SIAM Journal on Computing*, 35:932–963, April 2006.

[17] M. Pătraşcu and M. Thorup. Higher lower bounds for near-neighbor and further rich problems. *SIAM Journal on Computing*, 39(2):730–741, 2010.

[18] R. E. Tarjan. A class of algorithms that require nonlinear time to maintain disjoint sets. *Journal of Computer and System Sciences*, 18:110–127, 1979.