

ON RANGE SEARCHING WITH SEMIALGEBRAIC SETS. II*

PANKAJ K. AGARWAL[†], JIŘÍ MATOUŠEK[‡], AND MICHA SHARIR[§]

Abstract. Let P be a set of n points in \mathbb{R}^d . We present a linear-size data structure for answering range queries on P with constant-complexity semialgebraic sets as ranges, in time close to $O(n^{1-1/d})$. It essentially matches the performance of similar structures for simplex range searching, and, for $d \geq 5$, significantly improves earlier solutions by the first two authors obtained in 1994. This almost settles a long-standing open problem in range searching. The data structure is based on a partitioning technique of Guth and Katz [*On the Erdős distinct distances problem in the plane*, arXiv:1011.4105, 2010], which shows that for a parameter r , $1 < r \leq n$, there exists a d -variate polynomial f of degree $O(r^{1/d})$ such that each connected component of $\mathbb{R}^d \setminus Z(f)$ contains at most n/r points of P , where $Z(f)$ is the zero set of f . We present an efficient randomized algorithm for computing such a polynomial partition, which is of independent interest and is likely to have additional applications.

Key words. range searching, ham-sandwich cuts, polynomial partitions, cylindrical algebraic decomposition

AMS subject classifications. 52C45, 68P05, 68Q01, 68U05

DOI. 10.1137/120890855

1. Introduction.

Range searching. Let P be a set of n points in \mathbb{R}^d , where d is a small constant. Let Γ be a family of geometric “regions,” called *ranges*, in \mathbb{R}^d , each of which can be described algebraically by some fixed number of real parameters (a more precise definition is given below). For example, Γ can be the set of all axis-parallel boxes, balls, simplices, or cylinders, or the set of all intersections of pairs of ellipsoids. In the Γ -*range searching* problem, we want to preprocess P into a data structure so that the number of points of P lying in a query range $\gamma \in \Gamma$ can be counted efficiently. Similar to many previous papers, we actually consider a more general setting, the so-called *semigroup model*, where we are given a weight function on the points in P and we ask for the cumulative weight of the points in $P \cap \gamma$. The weights are assumed to belong to a semigroup, i.e., subtractions are not allowed. We assume that the semigroup operation can be executed in constant time.

In this paper we consider the case in which Γ is a set of constant-complexity semialgebraic sets. We recall that a *semialgebraic set* is a subset of \mathbb{R}^d obtained

*Received by the editors September 10, 2012; accepted for publication (in revised form) July 15, 2013; published electronically November 5, 2013. Part of this work was done while the first and third authors were visiting ETH Zürich. A preliminary version of the paper appeared in *Proceedings of the 53rd Annual IEEE Symposium on Foundations of Computer Science*, 2012.

<http://www.siam.org/journals/sicomp/42-6/89085.html>

[†]Department of Computer Science, Duke University, Durham, NC 27708-0129 (pankaj@cs.duke.edu). This author’s research was supported by the NSF under grants IIS-07-13498, CCF-09-40671, CCF-10-12254, and CCF-11-61359, by ARO grants W911NF-07-1-0376 and W911NF-08-1-0452, and by an ARL award W9132V-11-C-0003.

[‡]Department of Applied Mathematics, Charles University, 118 00 Praha 1, Czech Republic, and Institute of Theoretical Computer Science, ETH Zurich, 8092 Zurich, Switzerland (matousek@kam.mff.cuni.cz). This author’s research was supported by the ERC Advanced grant 267165.

[§]School of Computer Science, Tel Aviv University, Tel Aviv 69978, Israel, and Courant Institute of Mathematical Sciences, New York University, New York, NY 10012 (michas@post.tau.ac.il). This author’s research was supported by NSF grant CCF-08-30272, by grant 338/09 from the Israel Science Fund, by the Israeli Centers for Research Excellence (I-CORE) program (Center no. 4/11), and by the Hermann Minkowski–MINERVA Center for Geometry at Tel Aviv University.

from a finite number of sets of the form $\{x \in \mathbb{R}^d \mid g(x) \geq 0\}$, where g is a d -variate polynomial with integer coefficients,¹ by Boolean operations (unions, intersections, and complementations). Specifically, let $\Gamma_{d,\Delta,s}$ denote the family of all semialgebraic sets in \mathbb{R}^d defined by at most s polynomial inequalities of degree at most Δ each. If d, Δ, s are all regarded as constants, we refer to the sets in $\Gamma_{d,\Delta,s}$ as *constant-complexity semialgebraic sets* (such sets are sometimes also called *Tarski cells*). By *semialgebraic range searching* we mean $\Gamma_{d,\Delta,s}$ -range searching for some parameters d, Δ, s ; in most applications the actual collection Γ of ranges is only a restricted subset of some $\Gamma_{d,\Delta,s}$. Besides being interesting in its own right, semialgebraic range searching also arises in several geometric searching problems, such as searching for a point nearest to a query geometric object, counting the number of input objects intersecting a query object, and many others.

This paper focuses on the *low storage* version of range searching with constant-complexity semialgebraic sets—the data structure is allowed to use only linear or near-linear storage, and the goal is to make the query time as small as possible. At the other end of the spectrum we have the *fast query* version, where we want queries to be answered in polylogarithmic time using as little storage as possible. This variant is discussed briefly in section 8.

As is typical in computational geometry, we will use the *real RAM* model of computation, where we can compute exactly with arbitrary real numbers and each arithmetic operation is executed in constant time.

Previous work. Motivated by a wide range of applications, several variants of range searching have been studied in computational geometry and database systems at least since the 1980s. See [1, 22] for comprehensive surveys of this topic. The early work focused on the so-called *orthogonal range searching*, where ranges are axis-parallel boxes. After three decades of extensive work on this particular case, some basic questions still remain open. However, geometry plays little role in the known data structures for orthogonal range searching.

The most basic and most studied truly geometric instance of range searching is with *halfspaces*, or, more generally, *simplices*, as ranges. Studies in the early 1990s have essentially determined the optimal trade-off between the worst-case query time and the storage (and preprocessing time) required by any data structure for simplex range searching.² Lower bounds for this trade-off have been given by Chazelle [7] under the semigroup model of computation, where subtraction of the point weights is not allowed. It is possible that, say, the counting version of the simplex range searching problem, where we ask just for the number of points in the query simplex, might admit better solutions using subtractions, but no such solutions are known. Moreover, there are recent lower-bound results when subtractions are also allowed; see [18] and references therein.

The data structures proposed for simplex range searching over the last twenty years [20, 21] match the known lower bounds within polylogarithmic factors. The state-of-the-art upper bounds are by (i) Chan [6], who, building on many earlier

¹The usual definition of a semialgebraic set requires these polynomials to have integer coefficients. However, for our purposes, since we are going to assume the real RAM model of computation, we can actually allow for arbitrary real coefficients without affecting the asymptotic overhead.

²This applies when d is assumed to be *fixed* and the implicit constants in the asymptotic notation may depend on d . This is the setting in all the previous papers, including the present one. Of course, in practical applications, this assumption may be unrealistic unless the dimension is really small. However, the known lower bounds imply that if the dimension is large, no efficient solutions to simplex range searching exist, at least in the worst-case setting.

results, provides a linear-size data structure with $O(n \log n)$ expected preprocessing time and $O(n^{1-1/d})$ query time, and (ii) Matoušek [21], who provides a data structure with $O(n^d)$ storage, $O((\log n)^{d+1})$ query time, and $O(n^d(\log n)^\varepsilon)$ preprocessing time.³ A trade-off between space and query time can be obtained by combining these two data structures [21].

Yao and Yao [31] were perhaps the first to consider range searching in which ranges were delimited by graphs of polynomial functions. Agarwal and Matoušek [2] have introduced a systematic study of semialgebraic range searching. Building on the techniques developed for simplex range searching, they presented a linear-size data structure with $O(n^{1-1/b+\varepsilon})$ query time, where $b = \max(d, 2d - 4)$. For $d \leq 4$, this almost matches the performance for the simplex range searching, but for $d \geq 5$ there is a gap in the exponents of the corresponding bounds. Also see [27] for related recent developments.

The bottleneck in the performance of the just mentioned range-searching data structure of [2] is a combinatorial geometry problem, known as the *decomposition of arrangements into constant-complexity cells*. Here, we are given a set Σ of t algebraic surfaces in \mathbb{R}^d (i.e., zero sets of d -variate polynomials), with degrees bounded by a constant Δ_0 , and we want to decompose each cell of the arrangement $\mathcal{A}(\Sigma)$ (see section 4 for details) into subcells that are constant-complexity semialgebraic sets, i.e., belong to $\Gamma_{d,\Delta,s}$ for some constants Δ (bound on degrees) and s (number of defining polynomials), which may depend on d and Δ_0 , but not on t . The crucial quantity is the total number of the resulting subcells over all cells of $\mathcal{A}(\Sigma)$; namely, if one can construct such a decomposition with $O(t^b)$ subcells, with some constant b , for every t and Σ , then the method of [2] yields query time $O(n^{1-1/b+\varepsilon})$ (with linear storage). The only known general-purpose technique for producing such a decomposition is the so-called *vertical decomposition* [8, 26], which decomposes $\mathcal{A}(\Sigma)$ into roughly t^{2d-4} constant-complexity subcells, for $d \geq 4$ [17, 26].

An alternative approach, based on *linearization*, was also proposed in [2]. It maps the semialgebraic ranges in \mathbb{R}^d to simplices in some higher-dimensional space and uses simplex range searching there. However, its performance depends on the specific form of the polynomials defining the ranges. In some special cases (e.g., when ranges are balls in \mathbb{R}^d), linearization yields better query time than the decomposition-based technique mentioned above, but for general constant-complexity semialgebraic ranges, linearization has worse performance.

Our results. In a recent breakthrough, Guth and Katz [11] have presented a new space decomposition technique, called polynomial partitioning. For a set $P \subset \mathbb{R}^d$ of n points and a real parameter r , $1 < r \leq n$, an *r -partitioning polynomial* for P is a nonzero d -variate polynomial f such that each connected component of $\mathbb{R}^d \setminus Z(f)$ contains at most n/r points of P , where $Z(f) := \{x \in \mathbb{R}^d \mid f(x) = 0\}$ denotes the zero set of f . The decomposition of \mathbb{R}^d into $Z(f)$ and the connected components of $\mathbb{R}^d \setminus Z(f)$ is called a *polynomial partition* (induced by f). Guth and Katz show that an r -partitioning polynomial of degree $O(r^{1/d})$ always exists, but their argument does not lead to an efficient algorithm for constructing such a polynomial, mainly because it relies on ham-sandwich cuts in high-dimensional spaces, for which no efficient construction is known. Our first result is an efficient randomized algorithm for computing an r -partitioning polynomial.

³Here and in what follows, ε denotes an arbitrarily small positive constant. The implicit constants in the asymptotic notation may depend on it, generally tending to infinity as ε decreases to 0.

THEOREM 1.1. *Given a set P of n points in \mathbb{R}^d , for some fixed d , and a parameter $r \leq n$, an r -partitioning polynomial for P of degree $O(r^{1/d})$ can be computed in randomized expected time $O(nr + r^3)$.*

Next, we use this algorithm to bypass the arrangement-decomposition problem mentioned above. Namely, based on polynomial partitions, we construct *partition trees* [1, 22] that answer range queries with constant-complexity semialgebraic sets in near-optimal time, using linear storage. An essential ingredient in the performance analysis of these partition trees is a recent combinatorial result of Barone and Basu [3], originally conjectured by the second author, which deals with the complexity of certain kinds of arrangements of zero sets of polynomials (see Theorem 4.2). While there have already been several combinatorial applications of the Guth–Katz technique (the most impressive being the original one in [11], which solves the famous Erdős’s distinct distances problem, and some of the others presented in [13, 14, 28, 33]), ours seems to be the first *algorithmic* application.

We establish two range-searching results, both based on polynomial partitions. For the first result, we need to introduce the notion of D -general position, for an integer $D \geq 1$. We say that a set $P \subset \mathbb{R}^d$ is in D -general position if no k points of P are contained in the zero set of a nonzero d -variate polynomial of degree at most D , where $k := \binom{D+d}{d}$. This is the number one expects for a “generic” point set.⁴

THEOREM 1.2. *Let d, Δ, s , and $\varepsilon > 0$ be constants. Let $P \subset \mathbb{R}^d$ be an n -point set in D_0 -general position, where D_0 is a suitable constant depending on d, Δ , and ε . Then the $\Gamma_{d, \Delta, s}$ -range searching problem for P can be solved with $O(n)$ storage, $O(n \log n)$ expected preprocessing time, and $O(n^{1-1/d+\varepsilon})$ query time.*

We note that both here and in the next theorem, while the preprocessing algorithm is randomized, the queries are answered deterministically, and the query time bound is worst-case.

Of course, we would like to handle arbitrary point sets, not only those in D_0 -general position. This can be achieved by an infinitesimal perturbation of the points of P . A general technique known as “simulation of simplicity” (in the version considered by Yap [32]) ensures that the perturbed set P' is in D_0 -general position. If a point $p \in P$ lies in the interior of a query range γ , then so does the corresponding perturbed point $p' \in P'$, and similarly for p in the interior of $\mathbb{R}^d \setminus \gamma$. However, for p on the boundary of γ , we cannot be sure if p' ends up inside or outside γ .

Let us say that a *boundary-fuzzy* solution to the $\Gamma_{d, \Delta, s}$ -range searching problem is a data structure that, given a query $\gamma \in \Gamma_{d, \Delta, s}$, returns an answer in which all points of P in the interior of γ are counted and none in the interior of $\mathbb{R}^d \setminus \gamma$ is counted, while each point $p \in P$ on the boundary of γ may or may not be counted. In some applications, we can think of the points of P being imprecise anyway (e.g., their coordinates come from some imprecise measurement), and then boundary-fuzzy range searching may be adequate.

COROLLARY 1.3. *Let d, Δ, s , and $\varepsilon > 0$ be constants. Then for every n -point set in \mathbb{R}^d , there is a boundary-fuzzy $\Gamma_{d, \Delta, s}$ -range searching data structure with $O(n)$ storage, $O(n \log n)$ expected preprocessing time, and $O(n^{1-1/d+\varepsilon})$ query time.*

⁴Indeed, d -variate polynomials of degree at most D have at most $k - 1$ distinct nonconstant monomials. The Veronese map (see, e.g., [11]) maps \mathbb{R}^d to \mathbb{R}^{k-1} , and hyperplanes in \mathbb{R}^{k-1} correspond bijectively to k -variate polynomials of degree at most D . It follows that any set of $k - 1$ points in \mathbb{R}^d is contained in the zero set of a d -variate polynomial of degree at most D , corresponding to the hyperplane in \mathbb{R}^{k-1} passing through the Veronese images of these points. Similarly, k points in general position are not expected to have this property, because one does not expect their images to lie in a common hyperplane. See [9, 10] for more details.

Actually, previous results on range searching that use simulation of simplicity to avoid degenerate cases also solve only the boundary-fuzzy variant (see, e.g., [20, 21]). However, the previous techniques, even if presented only for point sets in general position, can usually be adapted to handle degenerate cases as well, perhaps with some effort, which is nevertheless routine. For our technique, degeneracy appears to be a more substantial problem because it is possible that a large subset of P (maybe even all of P) is contained in the zero set of the partitioning polynomial f , and the recursive divide-and-conquer mechanism yielded by the partition of f does not apply to this subset.

Partially in response to this issue, we present a different data structure that, at a somewhat higher preprocessing cost, not only gets rid of the boundary-fuzziness condition but also has a slightly improved query time (in terms of n). The main idea is that we build an auxiliary recursive data structure to handle the potentially large subset of points that lie in the zero set of the partitioning polynomial.

THEOREM 1.4. *Let d, Δ, s , and $\varepsilon > 0$ be constants. Then the $\Gamma_{d, \Delta, s}$ -range searching problem for an arbitrary n -point set in \mathbb{R}^d can be solved with $O(n)$ storage, $O(n^{1+\varepsilon})$ expected preprocessing time, and $O(n^{1-1/d} \log^B n)$ query time, where B is a constant depending on d, Δ, s , and ε .*

We remark that the dependence of B on Δ, s , and ε is reasonable, but its dependence on d is superexponential.

Our algorithms work for the semigroup model described earlier. Assuming that a semigroup operation can be executed in constant time, the query time remains the same as for the counting query. A reporting query—report the points of P lying in a query range—also fits in the semigroup model, except one cannot assume that a semigroup operation in this case takes constant time. The time taken by a reporting query is proportional to the cost of a counting query plus the number of reported points.

Roadmap of the paper. Our algorithm is based on the polynomial partitioning technique by Guth and Katz, and we begin by briefly reviewing it in section 2. Next, in section 3, we describe the randomized algorithm for constructing such a partitioning polynomial. Section 4 presents an algorithm for computing the cells of a polynomial partition that are crossed by a semialgebraic range, and discusses several related topics. Section 5 presents our first data structure, which is as in Theorem 1.2. Section 6 describes the method for handling points lying on the zero set of the partitioning polynomial, and section 7 presents our second data structure. We conclude in section 8 by mentioning a few open problems.

2. Polynomial partitions. In this section we briefly review the Guth–Katz technique for later use. We begin by stating their result.

THEOREM 2.1 (Guth–Katz [11]). *Given a set P of n points in \mathbb{R}^d and a parameter $r \leq n$, there exists an r -partitioning polynomial for P of degree at most $O(r^{1/d})$ (for d fixed).*

The degree in the theorem is asymptotically optimal in the worst-case because the number of connected components of $\mathbb{R}^d \setminus Z(f)$ is $O((\deg f)^d)$ for every polynomial f (see, e.g., Warren [30, Theorem 2]).

Sketch of proof. The Guth–Katz proof uses the *polynomial ham-sandwich* theorem of Stone and Tukey [29], which we state here in a version for finite point sets: *If A_1, \dots, A_k are finite sets in \mathbb{R}^d and D is an integer satisfying $\binom{D+d}{d} - 1 \geq k$, then there exists a nonzero polynomial f of degree at most D that simultaneously bisects*

all the sets A_i . Here “ f bisects A_i ” means that $f > 0$ in at most $\lfloor |A_i|/2 \rfloor$ points of A_i and $f < 0$ in at most $\lfloor |A_i|/2 \rfloor$ points of A_i ; f might vanish at any number of the points of A_i , possibly even at all of them.

Guth and Katz inductively construct collections $\mathcal{P}_0, \mathcal{P}_1, \dots, \mathcal{P}_m$ of subsets of P . For $j = 0, 1, \dots, m$, \mathcal{P}_j consists of at most 2^j pairwise-disjoint subsets of P , each of size at most $n/2^j$; the union of these sets does not have to contain all points of P .

Initially, we have $\mathcal{P}_0 = \{P\}$. The algorithm stops as soon as each subset in \mathcal{P}_m has at most n/r points. This implies that $m \leq \lceil \log_2 r \rceil$. Having constructed \mathcal{P}_{j-1} , we use the polynomial ham-sandwich theorem to construct a polynomial f_j that bisects each set of \mathcal{P}_{j-1} , with $\deg f_j = O(2^{j/d})$ (this is indeed an asymptotic upper bound for the smallest D satisfying $\binom{D+d}{d} - 1 \geq 2^{j-1}$, assuming d to be a constant). For every subset $Q \in \mathcal{P}_{j-1}$, let $Q^+ = \{q \in Q \mid f_j(q) > 0\}$ and $Q^- = \{q \in Q \mid f_j(q) < 0\}$. We set $\mathcal{P}_j := \{Q^+, Q^- \mid Q \in \mathcal{P}_{j-1}\}$; empty subsets are not included in \mathcal{P}_j .

The desired r -partitioning polynomial for P is then the product $f := f_1 f_2 \cdots f_m$. We have

$$\deg f = \sum_{j=1}^m \deg f_j = \sum_{j=1}^m O(2^{j/d}) = O(r^{1/d}).$$

By construction, the points of P lying in a single connected component of $\mathbb{R}^d \setminus Z(f)$ belong to a single member of \mathcal{P}_m , which implies that each connected component contains at most n/r points of P . \square

Sketch of proof of the Stone–Tukey polynomial ham-sandwich theorem.

We begin by observing that $\binom{D+d}{d} - 1$ is the number of all nonconstant monomials of degree at most D in d variables. Thus, we fix a collection \mathcal{M} of $k \leq \binom{D+d}{d} - 1$ such monomials. Let $\Phi: \mathbb{R}^d \rightarrow \mathbb{R}^k$ be the corresponding Veronese map, which maps a point $x = (x_1, \dots, x_d) \in \mathbb{R}^d$ to the k -tuple of the values at (x_1, \dots, x_d) of the monomials from \mathcal{M} . For example, for $d = 2$, $D = 3$, and $k = 8 \leq \binom{3+2}{2} - 1$, we may use $\Phi(x_1, x_2) = (x_1, x_2, x_1^2, x_1x_2, x_2^2, x_1^3, x_1^2x_2, x_1x_2^2) \in \mathbb{R}^8$, where \mathcal{M} is the set of the eight monomials appearing as components of Φ .

Let $B_i := \Phi(A_i) \subset \mathbb{R}^k$ be the image of the given A_i under this Veronese map, for $i = 1, \dots, k$. By the standard ham-sandwich theorem (see, e.g., [23]), there exists a hyperplane h in \mathbb{R}^k that simultaneously bisects all the B_i 's, in the sense that each open halfspace bounded by h contains at most half of the points of each of the sets B_i . In a more algebraic language, there is a nonzero k -variate linear polynomial, which we also call h , that bisects all the B_i 's, in the sense of being positive on at most half of the points of each B_i , and being negative on at most half of the points of each B_i . Then $f := h \circ \Phi$ is the desired d -variate polynomial of degree at most D bisecting all the A_i 's. \square

3. Constructing a partitioning polynomial. In this section we present an efficient randomized algorithm that, given a point set P and a parameter $r < n$, constructs an r -partitioning polynomial. The main difficulty in converting the above proof of the Guth–Katz partitioning theorem into an efficient algorithm is the use of the ham-sandwich theorem in the possibly high-dimensional space \mathbb{R}^k . A straightforward algorithm for computing ham-sandwich cuts in \mathbb{R}^k inspects all possible ways of splitting the input point sets by a hyperplane, and has running time about n^k . Compared to this easy upper bound, the best known ham-sandwich algorithms can save a factor of about n [19], but this is insignificant in higher dimensions. A recent result of Knauer, Tiwari, and Werner [16] shows that a certain incremental variant of

computing a ham-sandwich cut is $W[1]$ -hard (where the parameter is the dimension), and thus perhaps one should not expect much better exact algorithms.

We observe that the exact bisection of each A_i is not needed in the Guth–Katx construction—it is sufficient to replace the Stone–Tukey polynomial ham-sandwich theorem by a weaker result, as described below.

Constructing a well-dissecting polynomial. We say that a polynomial f is *well-dissecting* for a point set A if $f > 0$ on at most $\frac{7}{8}|A|$ points of A and $f < 0$ on at most $\frac{7}{8}|A|$ points of A . Given point sets A_1, \dots, A_k in \mathbb{R}^d with n points in total, we present a Las-Vegas algorithm for constructing a polynomial f of degree $O(k^{1/d})$ that is well-dissecting for at least $\lceil k/2 \rceil$ of the A_i 's.

As in the above proof of the Stone–Tukey polynomial ham-sandwich theorem, let D be the smallest integer satisfying $\binom{D+d}{d} - 1 \geq k$. We fix a collection \mathcal{M} of k distinct nonconstant monomials of degree at most D , and let Φ be the corresponding Veronese map. For each $i = 1, 2, \dots, k$, we pick a point $a_i \in A_i$ uniformly at random and compute $b_i := \Phi(a_i)$. Let h be a hyperplane in \mathbb{R}^k passing through b_1, \dots, b_k , which can be found by solving a system of linear equations, in $O(k^3)$ time.

If the points b_1, \dots, b_k are not affinely independent, then h is not determined uniquely (this is a technical nuisance, which the reader may want to ignore on first reading). In order to handle this case, we prepare in advance, before picking the a_i 's, *auxiliary* affinely independent points q_1, \dots, q_k in \mathbb{R}^k , which are in general position with respect to $\Phi(A_1), \dots, \Phi(A_k)$; here we mean the “ordinary” general position, i.e., no unnecessary affine dependences, that involve some of the q_i 's and the other points, arise. The points q_i can be chosen at random, say, uniformly in the unit cube; with high probability, they have the desired general position property. (If we do not want to assume the capability of choosing a random real number, we can pick the q_i 's uniformly at random from a sufficiently large discrete set). If the dimension of the affine hull of b_1, \dots, b_k is $k' < k - 1$, we choose the hyperplane h through b_1, \dots, b_k and $q_1, \dots, q_{k-k'+1}$. If h is not unique, i.e., $q_1, \dots, q_{k-k'+1}$ are not affinely independent with respect to b_1, \dots, b_k , which we can detect while solving the linear system, we restart the algorithm by choosing q_1, \dots, q_k anew and then picking new a_1, \dots, a_k . In this way, after a constant expected number of iterations, we obtain the uniquely determined hyperplane h through b_1, \dots, b_k and $q_1, \dots, q_{k-k'+1}$ as above, and we let $f = h \circ \Phi$ denote the corresponding d -variate polynomial. We refer to these steps as one *trial* of the algorithm. For each A_i , we check whether f is well-dissecting for A_i . If f is well-dissecting for only fewer than $k/2$ sets, then we discard f and perform another trial.

We now analyze the expected running time of the algorithm. The intuition is that f is expected to well-dissect a significant fraction, say at least half, of the sets A_i . This intuition is reflected in the next lemma. Let X_i be the indicator variable of the event: A_i is **not** well-dissected by f .

LEMMA 3.1. *For every $i = 1, 2, \dots, k$, $\mathbb{E}[X_i] \leq 1/4$.*

Proof. Let us fix i and the choices of a_j (and thus of $b_j = \Phi(a_j)$) for all $j \neq i$. Let k_0 be the dimension of F_0 , the affine hull of $\{b_j \mid j \neq i\}$. Then the resulting hyperplane h passes through the $(k - 2)$ -flat F spanned by F_0 and q_1, \dots, q_{k-k_0-2} , irrespective of which point of A_i is chosen. If a_i , the point chosen from A_i , is such that $b_i = \Phi(a_i)$ lies on F_0 , then h also passes through q_{k-k_0-1} .

Put $B_i := \Phi(A_i)$, and let us project the configuration orthogonally to a 2-dimensional plane π orthogonal to F . Then F appears as a point $F^* \in \pi$, and B_i projects to a (multi)set B_i^* in π . The random hyperplane h projects to a random line

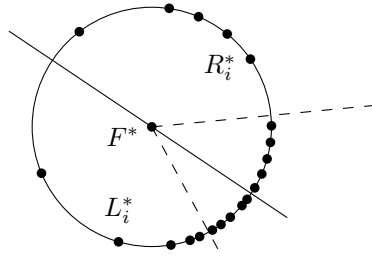


FIG. 1. Illustration to the proof of Lemma 3.1.

h^* in π , whose choice can be interpreted as follows: pick $b_i^* \in B_i^*$ uniformly at random; if $b_i^* \neq F^*$, then h^* is the unique line through b_i^* and F^* ; otherwise, when $b_i^* = F^*$, h^* is the unique line through F^* and $q_{k-k_0-1}^*$; by construction, $q_{k-k_0-1}^* \neq F^*$. The indicator variable X_i is 1 if and only if the resulting h^* has more than $\frac{7}{8}|B_i^*|$ points of B_i^* , counted with multiplicity, (strictly) on one side.

The special role of $q_{k-k_0-1}^*$ can be eliminated if we first move the points of B_i^* coinciding with F^* to the point $q_{k-k_0-1}^*$, and then slightly perturb the points so as to ensure that all points of B_i^* are distinct and lie at distinct directions from F^* ; it is easy to see that these transformations cannot decrease the probability of $X_i = 1$. Finally, we note that the side of h^* containing a point $b^* \in B_i^*$ only depends on the direction of the vector $\overrightarrow{F^*b^*}$, so we can also assume the points of B_i^* to lie on the unit circle around F^* .

Using (a simple instance of) the standard planar ham-sandwich theorem, we partition B_i^* into two subsets L_i^* and R_i^* of equal size by a line through the center F^* . Then we bisect L_i^* by a ray from F^* , and we do the same for R_i^* . It is easily checked (see Figure 1) that there always exist two of the resulting quarters, one of L_i^* and one of R_i^* (the ones whose union forms an angle $\leq \pi$ between the two bisecting rays), such that every line connecting F^* with a point in either quarter contains at least $\frac{1}{4}|B_i^*|$ points of B_i^* on each side. Referring to these quarters as “good,” we now take one of the bisecting rays, say that of L_i^* , and rotate it about F^* away from the good quarter of L_i^* . Each of the first $\frac{1}{8}|B_i^*|$ points that the ray encounters has the property that the line supporting the ray has at least $\frac{1}{8}|B_i^*|$ points of B_i^* on each side. This implies that, for at least half of the points in each of the two remaining quarters, the line connecting F^* to such a point has at least $\frac{1}{8}|B_i^*|$ points of B_i^* on each side. Hence at most $\frac{1}{4}|B_i^*|$ points of B_i^* can lead to a cut that is not well-dissecting for B_i^* .

We conclude that, still conditioned on the choices of $a_j, j \neq i$, the event $X_i = 1$ has probability at most $1/4$. Since this holds for every choice of the $a_j, j \neq i$, the unconditional probability of $X_i = 1$ is also at most $1/4$, and thus $\mathbb{E}[X_i] \leq 1/4$ as claimed. \square

Hence, the expected number of sets A_i that are not well-dissected by f is

$$\mathbb{E} \left[\sum_{i=1}^k X_i \right] = \sum_{i=1}^k \mathbb{E}[X_i] \leq k/4.$$

By Markov’s inequality, with probability at least $1/2$, at least half of the A_i ’s are well-dissected by f . We thus obtain a polynomial that is well-dissecting for at least half of the A_i ’s after an expected constant number of trials.

It remains to estimate the running time of each trial. The points b_1, \dots, b_k can be chosen in $O(n)$ time. Computing h involves solving a $k \times k$ linear system, which can be done in $O(k^3)$ time using Gaussian elimination. Note that we do *not* actually compute the entire sets $\Phi(A_i)$. No computation is needed for passing from h to f —we just reinterpret the coefficients. To check which of A_1, \dots, A_k are well-dissected by f , we evaluate f at each point of $A = \bigcup_i A_i$. First we evaluate each of the k monomials in \mathcal{M} at each point of A . If we proceed incrementally, from lower degrees to higher ones, this can be done with $O(1)$ operations per monomial and point of A , in $O(nk)$ time in total. Then, in additional $O(nk)$ time, we compute the values of $f(q)$, for all $q \in A$, from the values of the monomials. Putting everything together we obtain the following lemma.

LEMMA 3.2. *Given point sets A_1, \dots, A_k in \mathbb{R}^d (for fixed d) with n points in total, a polynomial f of degree $O(k^{1/d})$ that is well-dissecting for at least $\lceil k/2 \rceil$ of the A_i 's can be constructed in $O(nk + k^3)$ randomized expected time.*

Constructing a partitioning polynomial. We now describe the algorithm for computing an r -partitioning polynomial f . We essentially imitate the Guth–Katz construction, with Lemma 3.2 replacing the polynomial ham-sandwich theorem, but with an additional twist.

The algorithm works in phases. At the end of the j th phase, for $j \geq 1$, we have a family f_1, \dots, f_j of j polynomials and a family \mathcal{P}_j of at most 2^j pairwise-disjoint subsets of P , each of size at most $(7/8)^j n$. Similar to the Guth–Katz construction, \mathcal{P}_j is not necessarily a partition of P , since the points of $P \cap Z(f_1 f_2 \cdots f_j)$ do not belong to $\bigcup \mathcal{P}_j$. Initially, $\mathcal{P}_0 = \{P\}$. The algorithm stops when each set in \mathcal{P}_j has at most n/r points. In the j th phase, the algorithm constructs f_j and \mathcal{P}_j from f_1, \dots, f_{j-1} and \mathcal{P}_{j-1} , as follows.

At the beginning of the j th phase, let $\mathcal{L}_j = \{Q \in \mathcal{P}_{j-1} \mid |Q| > (7/8)^j n\}$ be the family of the “large” sets in \mathcal{P}_{j-1} , and set $\kappa_j = |\mathcal{L}_j| \leq (8/7)^j$. We also initialize the collection \mathcal{P}_j to $\mathcal{P}_{j-1} \setminus \mathcal{L}_j$, the family of “small” sets in \mathcal{P}_{j-1} . Then we perform at most $\lceil \log_2 \kappa_j \rceil$ dissecting steps, as follows: After s steps, we have a family g_1, \dots, g_s of polynomials, the current set \mathcal{P}_j , and a subfamily $\mathcal{L}_j^{(s)} \subseteq \mathcal{L}_j$ of size at most $\kappa_j / 2^s$, consisting of the members of \mathcal{L}_j that were not well-dissected by any of g_1, \dots, g_s . If $\mathcal{L}_j^{(s)} \neq \emptyset$, we choose, using Lemma 3.2, a polynomial g_{s+1} of degree at most $c(\kappa_j / 2^s)^{1/d}$ (with a suitable constant c that depends only on d) that well-dissects at least half of the members of $\mathcal{L}_j^{(s)}$. For each $Q \in \mathcal{L}_j^{(s)}$, let $Q^+ = \{q \in Q \mid g_{s+1}(q) > 0\}$ and $Q^- = \{q \in Q \mid g_{s+1}(q) < 0\}$. If Q is well-dissected, i.e., $|Q^+|, |Q^-| \leq \frac{7}{8}|Q|$, then we add Q^+, Q^- to \mathcal{P}_j , and otherwise, we add Q to $\mathcal{L}_j^{(s+1)}$. Note that in the former case the points $q \in Q$ satisfying $g_{s+1}(q) = 0$ are “lost” and do not participate in the subsequent dissections. By Lemma 3.2, $|\mathcal{L}_j^{(s+1)}| \leq |\mathcal{L}_j^{(s)}|/2 \leq \kappa_j / 2^{s+1}$.

The j th phase is completed when $\mathcal{L}_j^{(s)} = \emptyset$, in which case we set⁵ $f_j := \prod_{\ell=1}^s g_\ell$. By construction, each point set in \mathcal{P}_j has at most $(7/8)^j n$ points, and the points of P not belonging to any set of \mathcal{P}_j lie in $Z(f_1 \cdots f_j)$. Furthermore,

$$\deg f_j \leq \sum_{s \geq 0} c(\kappa_j / 2^s)^{1/d} = O(\kappa_j^{1/d}),$$

⁵Note that f_j is not necessarily well-dissecting, because it does not control the sizes of subsets with positive or with negative signs.

where again the constant of proportionality depends only on d . Since every set in \mathcal{P}_{j-1} is split into at most two sets before being added to \mathcal{P}_j , $|\mathcal{P}_j| \leq 2|\mathcal{P}_{j-1}| \leq 2^j$.

If \mathcal{P}_j contains subsets with more than n/r points, we begin the $(j+1)$ st phase with the current \mathcal{P}_j ; otherwise, the algorithm stops and returns $f := f_1 f_2 \cdots f_j$. This completes the description of the algorithm.

Clearly, m , the number of phases of the algorithm, is at most $\lceil \log_{8/7} r \rceil$. Following the same argument as in [11], and as briefly sketched in section 2, it can be shown that all points lying in a single connected component of $\mathbb{R}^d \setminus Z(f)$ belong to a single member of \mathcal{P}_m , and thus each connected component contains at most n/r points of P . Since the degree of f_j is $O(\kappa_j^{1/d})$, $\kappa_j \leq (8/7)^j$, and $m \leq \lceil \log_{8/7} r \rceil$, we conclude that

$$\deg f = O\left(\sum_{j=1}^m \kappa_j^{1/d}\right) = O\left(\sum_{j=1}^m (8/7)^{j/d}\right) = O(r^{1/d}).$$

As for the expected running time of the algorithm, the s th step of the j th phase takes $O(n\kappa_j/2^s + (\kappa_j/2^s)^3)$ expected time, so the j th phase takes a total of $O(n\kappa_j + \kappa_j^3)$ expected time. Substituting $\kappa_j \leq (8/7)^j$ in the above bound and summing over all j , the overall expected running time of the algorithm is $O(nr + r^3)$. This completes the proof of Theorem 1.1.

Remark 1. Theorem 1.1 is employed for the preprocessing in our range-searching algorithms in Theorems 1.2 and 1.4. In Theorem 1.2 we take r to be a large constant, and the expected running time in Theorem 1.1 is $O(n)$. However, in Theorem 1.4, we require r to be a small fractional power of n , say $r = n^{0.001}$. It is a challenging open problem to improve the expected running time in Theorem 1.1 to $O(n \text{ polylog}(n))$ when r is such a small fractional power of n . The bottleneck in the current algorithm is the subproblem of evaluating a given d -variate polynomial f of degree $D = O(r^{1/d})$ at n given points; everything else can be performed in $O(n \text{ polylog}(r) + r^{O(1)})$ expected time. Finding the signs of f at those points would actually suffice, but this probably does not make the problem any simpler.

This problem of *multievaluation* of multivariate real polynomials has been considered in the literature, and there is a nontrivial improvement over the straightforward $O(nr)$ algorithm, due to Nüsken and Ziegler [24]. Concretely, in the bivariate case ($d = 2$), their algorithm can evaluate a bivariate polynomial of degree $D \leq \sqrt{n}$ at n given points using $O(nD^{0.667})$ arithmetic operations. It is based on fast matrix multiplication, and even under the most optimistic possible assumption on the speed of matrix multiplication, it cannot get below $nD^{1/2}$. Although this is significantly faster than our naive $O(nr)$ -time algorithm, which is $O(nD^2)$ in this bivariate case, it is still a far cry from what we are aiming at. Let us remark that in a different setting, for polynomials over finite fields (and over certain more general finite rings), there is a remarkable method for multievaluation by Kedlaya and Umans [15] achieving $O((n + D^d) \log q)^{1+\varepsilon}$ running time, where q is the cardinality of the field.

4. Crossing a polynomial partition with a range. In this section we define the crossing number of a polynomial partition and describe an algorithm for computing the cells of a polynomial partition that are crossed by a semialgebraic range, both of which will be crucial for our range-searching data structures. We begin by recalling a few results on arrangements of algebraic surfaces. We refer the reader to [26] for a comprehensive review of such arrangements.

Let Σ be a finite set of algebraic surfaces in \mathbb{R}^d . The *arrangement* of Σ , denoted by $\mathcal{A}(\Sigma)$, is the partition of \mathbb{R}^d into maximal relatively open connected subsets, called

cells, such that all points within each cell lie in the same subset of surfaces of Σ (and in no other surface). If \mathcal{F} is a set of d -variate polynomials, then with a slight abuse of notation, we use $\mathcal{A}(\mathcal{F})$ to denote the arrangement $\mathcal{A}(\{Z(f) \mid f \in \mathcal{F}\})$ of their zero sets. We need the following result on arrangements, which follows from Proposition 7.33 and Theorem 16.18 in [5].

THEOREM 4.1 (Basu, Pollack, and Roy [5]). *Let $\mathcal{F} = \{f_1, \dots, f_s\}$ be a set of s real d -variate polynomials, each of degree at most Δ . Then the arrangement $\mathcal{A}(\mathcal{F})$ in \mathbb{R}^d has at most $O(1)^d (s\Delta)^d$ cells, and it can be computed in time at most $T = s^{d+1} \Delta^{O(d^4)}$. Each cell is described as a semialgebraic set using at most T polynomials of degree bounded by $\Delta^{O(d^3)}$. Moreover, the algorithm supplies an explicitly computed point in each cell.*

A key ingredient for the analysis of our range-searching data structure is the following recent result of Barone and Basu [3], which is a refinement of a series of previous studies; see, e.g., [4, 5]:

THEOREM 4.2 (Barone and Basu [3]). *Let V be a k -dimensional algebraic variety in \mathbb{R}^d defined by a finite set \mathcal{G} of d -variate polynomials, each of degree at most Δ , and let \mathcal{F} be a set of s polynomials of degree at most $D \geq \Delta$. Then the number of cells of $\mathcal{A}(\mathcal{F} \cup \mathcal{G})$ (of all dimensions) that are contained in V is bounded by $O(1)^d \Delta^{d-k} (sD)^k$.*

The crossing number of polynomial partitions. Let P be a set of n points in \mathbb{R}^d , and let f be an r -partitioning polynomial for P . Recall that the *polynomial partition* $\Omega = \Omega(f)$ induced by f is the partition of \mathbb{R}^d into the zero set $Z(f)$ and the connected components $\omega_1, \omega_2, \dots, \omega_t$ of $\mathbb{R}^d \setminus Z(f)$. As already noted, Warren’s theorem [30] implies that $t = O(r)$. We call $\omega_1, \dots, \omega_t$ the *cells* of Ω (although they need not be cells in the sense typical, e.g., in topology; they need not even be simply connected). Ω also induces a partition P^*, P_1, \dots, P_t of P , where $P^* = P \cap Z(f)$ is the *exceptional part*, and $P_i = P \cap \omega_i$, for $i = 1, \dots, t$, are the *regular parts*. By construction, $|P_i| \leq n/r$ for every $i = 1, 2, \dots, t$, but we have no control over the size of P^* —this will be the source of most of our technical difficulties.

Next, let γ be a range in $\Gamma_{d,\Delta,s}$. We say that γ *crosses* a cell ω_i if neither $\omega_i \subseteq \gamma$ nor $\omega_i \cap \gamma = \emptyset$. The *crossing number* of γ is the number of cells of Ω crossed by γ , and the *crossing number* of Ω (with respect to $\Gamma_{d,\Delta,s}$) is the maximum of the crossing numbers of all $\gamma \in \Gamma_{d,\Delta,s}$. Similar to many previous range-searching algorithms [6, 20, 21], the crossing number of Ω will determine the query time of our range-searching algorithms described in sections 5 and 7.

LEMMA 4.3. *If Ω is a polynomial partition induced by an r -partitioning polynomial of degree at most D , then the crossing number of Ω with respect to $\Gamma_{d,\Delta,s}$, with $\Delta \leq D$, is at most $Cs\Delta D^{d-1}$, where C is a suitable constant depending only on d .*

Proof. Let $\gamma \in \Gamma_{d,\Delta,s}$; then γ is a Boolean combination of up to s sets of the form $\gamma_j := \{x \in \mathbb{R}^d \mid g_j(x) \geq 0\}$, where g_1, \dots, g_s are polynomials of degree at most Δ . If γ crosses a cell ω_i , then at least one of the ranges γ_j also crosses ω_i , and thus it suffices to establish that the crossing number of any range γ , defined by a single d -variate polynomial inequality $g(x) \geq 0$ of degree at most Δ , is at most $C\Delta D^{d-1}$. \square

We apply Theorem 4.2 with $V := Z(g)$, which is an algebraic variety of dimension $k \leq d - 1$, and with $s = 1$ and $\mathcal{F} = \{f\}$, where f is the r -partitioning polynomial. Then, for each cell ω_i crossed by γ , $\omega_i \cap Z(g)$ is a nonempty union of some of the cells in $\mathcal{A}(\mathcal{F} \cup \{g\}) = \mathcal{A}(\{f, g\})$ that lie in V . Thus, the crossing number of γ is at most $O(1)^d \Delta D^{d-1}$. \square

Algorithmic issues. We need to perform the following algorithmic primitives (for d fixed as usual) for the range-searching algorithms that we will later present:

- (A1) Given an r -partitioning polynomial f of degree $D = O(r^{1/d})$, compute (a suitable representation of) the partition Ω and the induced partition of P into P^*, P_1, \dots, P_t .

By computing $\mathcal{A}(\{f\})$, using Theorem 4.1, and then testing the membership of each point $p \in P$ in each cell ω_i in time polynomial in r , the above operation can be performed in $O(nr^c)$ time,⁶ where $c = d^{O(1)}$.

- (A2) Given (a suitable representation of) Ω as in (A1) and a query range $\gamma \in \Gamma_{d,\Delta,1}$, i.e., a range defined by a single d -variate polynomial g of degree $\Delta \leq D$, compute which of the cells of Ω are crossed by γ and which are completely contained in γ .

We already have the arrangement $\mathcal{A}(\{f\})$, and we compute $\mathcal{A}(\{f, g\})$. For each cell of $\mathcal{A}(\{f, g\})$ contained in $Z(g)$, we locate its representative point in $\mathcal{A}(\{f\})$, and this gives us the cells crossed by γ . For the remaining cells, we want to know whether they are inside γ or outside, and for that, it suffices to determine the sign of g at the representative points. Using Theorem 4.1, the above task can thus be accomplished in time $O(r^c)$, with $c = d^{O(1)}$.

5. Constant fan-out partition tree. We are now ready to describe our first data structure for $\Gamma_{d,\Delta,s}$ -range searching, which is a constant fan-out (branching degree) partition tree, and which works for points in general position.

Proof of Theorem 1.2. Let P be a set of n points in \mathbb{R}^d , and let Δ, s be constants. We choose r as a (large) constant depending on d, Δ, s , and the prespecified parameter ε . We assume P to be in D_0 -general position for some sufficiently large constant $D_0 \gg r^{1/d}$. We construct a partition tree \mathcal{T} of fan-out $O(r)$ as follows. We first construct an r -partitioning polynomial f for P using Theorem 1.1, and compute the partition Ω of \mathbb{R}^d induced by f , as well as the corresponding partition $P = P^* \cup P_1 \cup \dots \cup P_t$ of P , where $t = O(r)$. Since r is a constant, the (A1) operation, discussed in section 4, performs this computation in $O(n)$ time. We choose D_0 so as to ensure that it is at least $\deg f$, and then our assumption that P is in D_0 -general position implies that the size of $P^* = P \cap Z(f)$ is bounded by D_0 .

We set up the root of \mathcal{T} , where we store

- (i) the partitioning polynomial f , and a suitable representation of the partition Ω ;
- (ii) a list of the points of the exceptional part P^* ; and
- (iii) $w(P_i)$, the sum of the weights of the points of the regular part P_i , for each $i = 1, 2, \dots, t$.

The regular parts P_i are not stored explicitly at the root. Instead, for each P_i we recursively build a subtree representing it. The recursion terminates, at leaves of \mathcal{T} , as soon as we reach point sets of size smaller than a suitable constant n_0 . The points of each such set are stored explicitly at the corresponding leaf of \mathcal{T} .

Since each node of \mathcal{T} requires only a constant amount of storage and each point of P is stored at only one node of \mathcal{T} , the total size of \mathcal{T} is $O(n)$. The preprocessing time is $O(n \log n)$ since \mathcal{T} has depth $O(\log_r n)$ and each level is processed in $O(n)$ time.

⁶Of course, this is somewhat inefficient, and it would be nice to have a fast point-location algorithm for the partition Ω —this would be the second step, together with an improved construction of an r -partitioning polynomial f (concretely, an improved multipoint evaluation procedure for f) as discussed at the end of section 3, needed to improve the preprocessing time in Theorem 1.4.

To process a query range $\gamma \in \Gamma_{d,\Delta,s}$, we start at the root of \mathcal{T} and maintain a global counter which is initially set to 0. Among the cells $\omega_1, \dots, \omega_t$ of the partition Ω stored at the root, we find, using the (A2) operation, those completely contained in γ , and those crossed by γ . Actually, we compute a superset of the cells that γ crosses, namely, the cells crossed by the zero set of at least one of the (at most s) polynomials defining γ . For each cell $\omega_i \subseteq \gamma$, we add the weight $w(P_i)$ to the global counter. We also add to the global counter the weights of the points in $P^* \cap \gamma$, which we find by testing each point of P^* separately. Then we recurse in each subtree corresponding to a cell ω_i crossed by γ (in the above weaker sense). The leaves, with point sets of size $O(1)$, are processed by inspecting their points individually. By Lemma 4.3, the number of cells crossed by any of the polynomials defining γ at any interior node of \mathcal{T} is at most $Cs\Delta D^{d-1} \leq C'r^{1-1/d}$, where $C' = C'(d, s, \Delta)$ is a constant independent of r .

The query time $Q(n)$ obeys the following recurrence:

$$Q(n) \leq \begin{cases} C'r^{1-1/d}Q(n/r) + O(1) & \text{for } n > n_0, \\ O(n) & \text{for } n \leq n_0. \end{cases}$$

It is well known (see., e.g., [20]), and easy to check, that the recurrence solves to $Q(n) = O(n^{1-1/d+\varepsilon})$, for every fixed $\varepsilon > 0$, with an appropriate sufficiently large choice of r as a function of C' and ε , and with an appropriate choice of n_0 . This concludes the proof of Theorem 1.2. \square

Boundary-fuzzy range searching: Proof of Corollary 1.3. Now we consider the case where the points of P are not necessarily in D_0 -general position. As was mentioned in the introduction, we apply a general perturbation scheme of Yap [32] to the previous range-searching algorithm.

Yap's scheme is applicable to an algorithm whose input is a sequence of real numbers (in our case, the dn point coordinates plus the coefficients in the polynomials specifying the query range). It is assumed that the algorithm makes decision steps by way of evaluating polynomials with rational coefficients taken from a finite set \mathcal{P} , where the input parameters are substituted for the variables. The algorithm makes a 3-way branching depending on the sign of the evaluation. The set \mathcal{P} does not depend on the input. The input is considered degenerate if one of the signs in the tests is 0.

Yap's scheme provides a black box for evaluating the polynomials from \mathcal{P} that, whenever the actual value is 0, also supplies a nonzero sign, +1 or -1, which the algorithm may use for the branching, instead of the zero sign. Thus, the algorithm never "sees" any degeneracy. Yap's method guarantees that these signs are consistent, i.e., for every input, the branching done in this way corresponds to some infinitesimal perturbation of the input sequence, and so does the output of the algorithm (in our case, the answer to a range-searching query).

For us, it is important that if the degrees of the polynomials in \mathcal{P} are bounded by a constant, the black box also operates in time bounded by a constant (which is apparent from the explicit specification in [32]). Thus, applying the perturbation scheme influences the running time only by a multiplicative constant.

It can be checked that the range-searching algorithm presented above is of the required kind, with all branching steps based on the sign of suitable polynomials in the coordinates of the input points and in the coefficients of the polynomials in the query range, and the degrees of these polynomials are bounded by a constant. For producing the partitioning polynomial f , we solve systems of linear equations, and

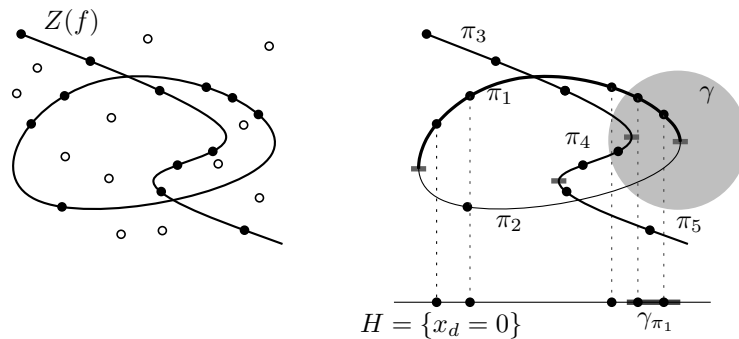


FIG. 2. The zero set of the partitioning polynomial (left), and its decomposition into monotone patches that project to the hyperplane H bijectively. Only the 1-dimensional patches are labeled.

thus the coefficients of f are given by certain determinants obtained from Cramer's rule. The computation of the polynomial partition and locating points in it is also based on the signs of suitable bounded-degree polynomials, as can be checked by inspecting the relevant algorithms, and similarly for intersecting a polynomial partition with the query range. The key fact is that all computations in the algorithm are of constant-bounded depth—each of the values ever computed is obtained from the input parameters by a constant number of arithmetic operations.

We also observe that when Yap's scheme is applied, the algorithm never finds more than D_0 points in the exceptional set P^* (in any of the nodes of the partition tree). Indeed, if $D_0 + 1$ input points lie in the zero set of a polynomial f as in the algorithm, then a certain polynomial in the coordinates of these $D_0 + 1$ points vanishes (see, e.g., [13, Lemma 6.3]). Thus, assuming that the algorithm found $D_0 + 1$ points on $Z(f)$, it could test the sign of this polynomial at such points, and the black box would return a nonzero sign, which would contradict the consistency of Yap's scheme.

After applying Yap's scheme, the preprocessing cost, storage, and query time remain asymptotically the same as in Theorem 1.2 (but with larger constants of proportionality). Since the output of the algorithm corresponds to some infinitesimally perturbed version of the input (point set and query range), we obtain a boundary-fuzzy answer for the original point set. \square

6. Decomposing a surface into monotone patches. As mentioned in the introduction, if we construct an r -partitioning polynomial f for an arbitrary point set P , the exceptional set $P^* = P \cap Z(f)$ may be large, as is schematically indicated in Figure 2 (left). Since P^* is not partitioned by f in any reasonable sense, it must be handled differently, as described below.

Following the terminology in [12, 25], we call a direction $v \in \mathbb{S}^{d-1}$ *good* for f if, for every $a \in \mathbb{R}^d$, the polynomial $p(t) = f(a + vt)$ does not vanish identically; that is, any line in direction v intersects $Z(f)$ at finitely many points. As argued in [25, pp. 304–305 and pp. 314–315], a random direction is good for f with probability 1. By choosing a good direction and rotating the coordinate system, we assume that the x_d -direction, referred to as the *vertical* direction, is good for f .

In order to deal with P^* , we partition $Z(f)$ into finitely many pieces, called *patches*, in such a way that each of the patches is *monotone* in the vertical direction, meaning that every line parallel to the x_d -axis intersects it at most once. This is illustrated, in the somewhat trivial 2-dimensional setting, in Figure 2 (right): there

are five 1-dimensional patches π_1, \dots, π_5 , plus four 0-dimensional patches. Then we treat each patch π separately: We project the point set $P^* \cap \pi$ orthogonally to the coordinate hyperplane $H := \{x_d = 0\}$, and we preprocess the projected set, denoted P_π^* , for range searching with suitable ranges. These ranges are projections of ranges of the form $\gamma \cap \pi$, where $\gamma \in \Gamma_{d,\Delta,s}$ is one of the original ranges. In Figure 2 (right), the patch π_1 is drawn thick, a range γ is depicted as a gray disk, and the projection γ_{π_1} of $\gamma \cap \pi_1$ is shown as a thick segment in H .

The projected range γ_π is typically more complicated than the original range γ (it involves more polynomials of larger degrees), but, crucially, it is only $(d-1)$ -dimensional, and $(d-1)$ -dimensional queries can be processed somewhat more efficiently than d -dimensional ones, which makes the whole scheme work. We will discuss this in more detail in section 7, but first we recall the notion of *cylindrical algebraic decomposition* (CAD, or also *Collins decomposition*), which is a tool that allows us to decompose $Z(f)$ into monotone patches, and also to compute the projected ranges γ_π .

Given a finite set $\mathcal{F} = \{f_1, \dots, f_s\}$ of d -variate polynomials, a *cylindrical algebraic decomposition adapted to \mathcal{F}* is a way of decomposing \mathbb{R}^d into a finite collection of relatively open *cells*, which have a simple shape (in a suitable sense), and which refine the arrangement $\mathcal{A}(\mathcal{F})$. We refer, e.g., to [5, Chap. 5.12] for the definition and construction of the “standard” CAD. Here we will use a simplified variant, which can be regarded as the “first stage” of the standard CAD, and which is captured by [5, Theorem 5.14, Algorithm 12.1]. We also refer to [25, Appendix A] for a concise treatment, which is perhaps more accessible at first encounter.

Let \mathcal{F} be as above. To obtain the first-stage CAD for f , one constructs a suitable collection $\mathcal{E} = \mathcal{E}(\mathcal{F})$ of polynomials in the variables x_1, \dots, x_{d-1} (denoted by $\text{Elim}_{x_d}(\mathcal{F})$ in [5]). Roughly speaking, the zero sets of the polynomials in \mathcal{E} , viewed as subsets of the coordinate hyperplane H (which is identified with \mathbb{R}^{d-1}), contain the projection onto H of all intersections $Z(f_i) \cap Z(f_j)$, $1 \leq i < j \leq s$, as well as the projection of the loci in $Z(f_i)$ where $Z(f_i)$ has a vertical tangent hyperplane, or a singularity of some kind. The actual construction of \mathcal{E} is somewhat more complicated, and we refer to the aforementioned references for more details.

Having constructed \mathcal{E} , the first-stage CAD is obtained as the arrangement $\mathcal{A}(\mathcal{F} \cup \mathcal{E})$ in \mathbb{R}^d , where the polynomials in \mathcal{E} are now considered as d -variate polynomials (in which the variable x_d is not present). In geometric terms, we erect a “vertical wall” in \mathbb{R}^d over each zero set within H of a $(d-1)$ -variate polynomial from \mathcal{E} , and the CAD is the arrangement of these vertical walls plus the zero sets of f_1, \dots, f_s . The first-stage CAD is illustrated in Figure 3, for the same (single) polynomial as in Figure 2 (left).

In our algorithm, we are interested in the cells of the CAD that are contained in some of the zero sets $Z(f_i)$; these are going to be the monotone patches alluded to above. We note that using the first-stage CAD for the purpose of decomposing $Z(f)$ into monotone patches seems somewhat wasteful. For example, the number of patches in Figure 2 is considerably smaller than the number of patches in the CAD in Figure 3. But the CAD is simple and well known, and (as will follow from the analysis in section 7) possible improvements in the number of patches (e.g., using the vertical-decomposition technique [26]) do not seem to influence our asymptotic bounds on the performance of the resulting range-searching data structure. The following lemma summarizes the properties of the first-stage CAD that we will need; we refer to [5, Theorem 5.14, Algorithm 12.1] for a proof.

LEMMA 6.1 (single-stage CAD). *Given a set $\mathcal{F} = \{f_1, \dots, f_s\} \subset \mathbb{R}[x_1, \dots, x_d]$ of polynomials, each of degree at most D , there is a set $\mathcal{E} = \mathcal{E}(\mathcal{F})$ of $O(s^2 D^3)$ polynomials in $\mathbb{R}[x_1, \dots, x_{d-1}]$, each of degree $O(D^2)$, which can be computed in time $s^2 D^{O(d)}$, such*

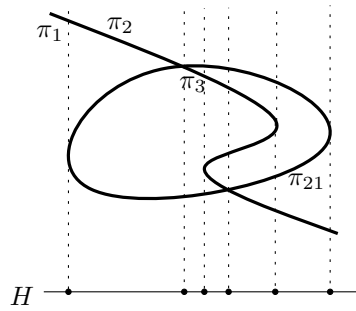


FIG. 3. A schematic illustration of the first-stage cylindrical algebraic decomposition.

that the first-stage CAD defined by these polynomials, i.e., the arrangement $\mathcal{A}(\mathcal{F} \cup \mathcal{E})$ in \mathbb{R}^d , has the following properties:

- (i) (“Cylindrical” cells) For each cell σ of $\mathcal{A}(\mathcal{F} \cup \mathcal{E})$, there exists a unique cell τ of the $(d - 1)$ -dimensional arrangement $\mathcal{A}(\mathcal{E})$ in H , such that one of the following possibilities occur:
 - (a) $\sigma = \{(x, \xi(x)) \mid x \in \tau\}$, where $\xi: \tau \rightarrow \mathbb{R}$ is a continuous semialgebraic function (that is, σ is the graph of ξ over τ).
 - (b) $\sigma = \{(x, t) \mid x \in \tau, t \in (\xi_1(x), \xi_2(x))\}$, where each ξ_i , $i = 1, 2$, is either a continuous semialgebraic real-valued function on τ , or the constant function $\tau \rightarrow \{\infty\}$, or the constant function $\tau \rightarrow \{-\infty\}$, and $\xi_1(x) < \xi_2(x)$ for all $x \in \tau$ (that is, σ is a portion of the “cylinder” $\tau \times \mathbb{R}$ between two consecutive graphs).
- (ii) (Refinement property) If $\mathcal{F}' \subseteq \mathcal{F}$, then $\mathcal{E}' = \mathcal{E}(\mathcal{F}') \subseteq \mathcal{E}$, and thus each cell of $\mathcal{A}(\mathcal{F}' \cup \mathcal{E}')$ is fully contained in some cell of $\mathcal{A}(\mathcal{F} \cup \mathcal{E})$.

Returning to the problem of decomposing the zero set of the partitioning polynomial f into monotone patches, we construct the first-stage CAD for $\mathcal{F} = \{f\}$, and the patches are the cells of $\mathcal{A}(\mathcal{F} \cup \mathcal{E})$ contained in $Z(f)$. If the x_d -direction is good for f , then every cell of $\mathcal{A}(\mathcal{F} \cup \mathcal{E})$ lying in $Z(f)$ is of type (a), and so if any cell of type (b) lies in $Z(f)$, we choose another random direction and construct the first-stage CAD in that direction. Putting everything together and using Theorem 4.1 to bound the complexity of $\mathcal{A}(\mathcal{F} \cup \mathcal{E})$, we obtain the following lemma.

LEMMA 6.2. Let f be a d -variate polynomial of degree D , and let us assume that the x_d -direction is good for f . Then $Z(f)$ can be decomposed, in $D^{O(d^4)}$ time, into $D^{O(d)}$ monotone patches, and each patch can be represented semialgebraically by $D^{O(d^4)}$ polynomials of degree $D^{O(d^3)}$.

The first-stage CAD can also be used to compute the projection of the intersection of a range in $\Gamma_{d,\Delta,s}$ with a monotone patch of f .

LEMMA 6.3. Let Π be the decomposition of the zero set of a d -variate polynomial f of degree D into monotone patches, as described in Lemma 6.2, and let γ be a semialgebraic set in $\Gamma_{d,\Delta,s}$, with $\Delta \leq D$. For every patch $\pi \in \Pi$, the projection of $\gamma \cap \pi$ in the x_d -direction can be represented as a member of $\Gamma_{d-1,\Delta_1,s_1}$, i.e., by a Boolean combination of at most s_1 polynomial inequalities in $(d - 1)$ variables, each of degree at most Δ_1 , where $\Delta_1 = D^{O(d^3)}$ and $s_1 = (Ds)^{O(d^4)}$. The representation can be computed in $(Ds)^{O(d^4)}$ time.

Proof. The task of computing γ_π , the projection of $\gamma \cap \pi$, is similar to the operation (A2) discussed in section 4. In more abstract terms, it can also be viewed

as a quantifier elimination task: we can represent $\gamma \cap \pi$ by a quantifier-free formula $\Phi(x_1, \dots, x_d)$ (a Boolean combination of polynomial inequalities); then γ_π is represented by $\exists x_d \Phi(x_1, \dots, x_d)$, and by eliminating $\exists x_d$ we obtain a quantifier-free formula describing γ_π . More concretely, we use a procedure based on the first-stage CAD (Lemma 6.1) and the arrangement construction (Theorem 4.1).

By definition, γ is a Boolean combination of inequalities of the form $g_1 \geq 0, \dots, g_s \geq 0$, where g_1, \dots, g_s are d -variate polynomials, each of degree at most $\Delta \leq D$. We set $\tilde{\mathcal{F}} := \{f, g_1, \dots, g_s\}$, we compute the set $\tilde{\mathcal{E}} = \mathcal{E}(\tilde{\mathcal{F}})$ of $(d - 1)$ -variate polynomials as in Lemma 6.1, and the first-stage CAD is then computed as the d -dimensional arrangement $\mathcal{A}(\tilde{\mathcal{F}} \cup \tilde{\mathcal{E}})$ according to Theorem 4.1. Since by Lemma 6.1(ii), $\mathcal{A}(\tilde{\mathcal{F}} \cup \tilde{\mathcal{E}})$ refines $\mathcal{A}(\{f\} \cup \mathcal{E}(\{f\}))$ (the first-stage CAD from the preprocessing phase), each patch $\pi \in \Pi$ is decomposed into subpatches. Since the sign of each g_i is constant on each cell of $\mathcal{A}(\tilde{\mathcal{F}})$, and thus on each cell of $\mathcal{A}(\tilde{\mathcal{F}} \cup \tilde{\mathcal{E}})$, $\gamma \cap \pi$ is a disjoint union of subpatches. The projections of these subpatches into H are cells of $\mathcal{A}(\tilde{\mathcal{E}})$, and thus we obtain, in time $(Ds)^{O(d^4)}$, a representation of γ_π as a member of $\Gamma_{d-1, \Delta_1, s_1}$ by Theorem 4.1, where $\Delta_1 = D^{O(d^3)}$ and $s_1 = (Ds)^{O(d^4)}$. \square

7. Large fan-out partition tree: Proof of Theorem 1.4. We now describe our second data structure for $\Gamma_{d, \Delta, s}$ -range searching. Compared to the first data structure from section 5, this one works on arbitrary point sets, without the D_0 -general position assumption, or, alternatively, without the fuzzy boundary constraint on the output, and has slightly better performance bounds. The data structure is built recursively, and this time the recursion involves both n and d .

7.1. The data structure. Let P be a set of n points in \mathbb{R}^d , and let Δ and s be parameters (not assumed to be constant). The data structure for $\Gamma_{d, \Delta, s}$ -range searching on P is obtained by constructing a partition tree \mathcal{T} on P recursively, as above, except that now the fan-out of each node is larger (and nonconstant), and each node also stores an auxiliary data structure for handling the respective exceptional part. We need to set two parameters: $n_0 = n_0(d, \Delta, s)$ and $r = r(d, \Delta, s, n)$. Neither of them is a constant in general; in particular, r is typically going to be a tiny power of n . The specific values of these parameters will be specified later, when we analyze the query time.

We also note that there is yet another parameter in Theorem 1.4, namely, the arbitrarily small constant $\varepsilon > 0$ entering the preprocessing time bound. However, ε enters the construction solely by the requirement that r should be chosen smaller than $n^{\varepsilon/c}$, for a sufficiently large constant c . It will become apparent later in the analysis that $r \leq n^{\varepsilon/c}$ can be assumed, provided that some other parameters are taken sufficiently large; we will point this out at suitable moments.

When constructing the partition tree \mathcal{T} on an n -point set P , we distinguish two cases. For $n \leq n_0$, \mathcal{T} consists of a single leaf storing all points of P . For $n > n_0$, we construct an r -partitioning polynomial f of degree $D = O(r^{1/d})$, the partition Ω of \mathbb{R}^d induced by f , and the partition of P into the exceptional part P^* and regular parts P_1, \dots, P_t , where $t = O(r)$. Set $n^* = |P^*|$ and $n_i = |P_i|$, for $i = 1, \dots, t$. The root of \mathcal{T} stores f , Ω , and the total weight $w(P_i)$ of each regular part P_i of P , as before. Still in the same way as before, we recursively preprocess each regular part P_i for $\Gamma_{d, \Delta, s}$ -range searching (or stop if $|P_i| \leq n_0$), and attach the resulting data structure to the root as a respective subtree.

Handling the exceptional part. A new feature of the second data structure is that we also preprocess the exceptional set P^* into an auxiliary data structure, which

is stored at the root. Here we recurse on the dimension, exploiting the fact that P^* lies on the algebraic variety $Z(f)$ of dimension at most $d - 1$.

We choose a random direction v and rotate the coordinate system so that v becomes the direction of the x_d -axis. We construct the first-stage CAD adapted to $\{f\}$, according to Lemma 6.1 and Theorem 4.1. We check whether all the patches are x_d -monotone, i.e., of type (a) in Lemma 6.1(i); if it is not the case, we discard the CAD and repeat the construction, with a different random direction. This yields a decomposition of $Z(f)$ into a set Π of $D^{O(d)}$ monotone patches, and the running time is $D^{O(d^4)}$ with high probability.

Next, we distribute the points of P^* among the patches: for each patch $\pi \in \Pi$, let P_π^* denote the projection of $P^* \cap \pi$ onto the coordinate hyperplane $H = \{x \in \mathbb{R}^d \mid x_d = 0\}$. We preprocess each set P_π^* for $\Gamma_{d-1, \Delta_1, s_1}$ -range searching. Here $s_1 = (Ds)^{O(d^4)}$ is the number of polynomials defining a range and $\Delta_1 = D^{O(d^3)}$ is their maximum degree; the constants hidden in the $O(\cdot)$ notation are the same as in Lemma 6.3. For simplicity, we treat all patches as being $(d-1)$ -dimensional (although some may be of lower dimension); this does not influence the worst-case performance analysis.

The preprocessing of the sets P_π^* is done recursively, using an r_1 -partitioning polynomial in \mathbb{R}^{d-1} , for a suitable value of r_1 . The exceptional set at each node of the resulting “ $(d-1)$ -dimensional” tree is handled in a similar manner, constructing an auxiliary data structure in $d-2$ dimensions, based on a first-stage CAD, and storing it at the corresponding node. The recursion on d bottoms out at dimension 1, where the structure is simply a standard binary search tree over the resulting set of points on the x_1 -axis. We remark that the treatment of the top level of recursion on the dimension will be somewhat different from that of deeper levels, in terms of both the choice of parameters and the analysis; see below for details.

This completes the description of the data structure, except for the choice of r and n_0 , which will be provided later as we analyze the performance of the algorithm.

Answering a query. Let us assume that, for a given P , the data structure for $\Gamma_{d, \Delta, s}$ -range searching, as described above, has been constructed, and consider a query range $\gamma \in \Gamma_{d, \Delta, s}$. The query is answered in the same way as before, by visiting the nodes of the partition tree \mathcal{T} in a top-down manner, except that, at each node that we visit, we also query with γ the auxiliary data structure constructed on the exceptional set P^* for that node.

Specifically, for each patch π of the corresponding collection Π , we compute w_π , the weight of $P^* \cap (\gamma \cap \pi)$. If $\gamma \cap \pi = \emptyset$, then $w_\pi = 0$, and if $\gamma \cap \pi = \pi$, then w_π is the total weight of $P^* \cap \pi$. Otherwise, i.e., if γ crosses π , then w_π is the same as the weight of $P_\pi^* \cap \gamma_\pi$, where γ_π is the x_d -projection of $\gamma \cap \pi$, because π is x_d -monotone. By Lemma 6.3, $\gamma_\pi \in \Gamma_{d-1, \Delta_1, s_1}$ and can be constructed in $(Ds)^{O(d^4)}$ time. We can find the weight of $\gamma_\pi \cap P_\pi^*$ by querying the auxiliary data structure for P_π^* with γ_π . We then add w_π to the global count maintained by the query procedure. This completes the description of the query procedure.

7.2. Performance analysis. The analysis of the storage requirement and preprocessing time is straightforward, and will be provided later. We begin with the more intricate analysis of the query time. For now we assume that n_0 and r have been fixed; the analysis will later specify their values.

Let $Q_d(n, \Delta, s)$ denote the maximum overall query time for $\Gamma_{d, \Delta, s}$ -range searching on a set of n points in \mathbb{R}^d . For $n \leq n_0$ and $d \geq 1$, $Q_d(n, \Delta, s) = O(n)$. For $d = 1$

and $n > n_0$, $Q_1(n, \Delta, s) = O(\Delta s \log n)$ because any range in $\Gamma_{1,\Delta,s}$ is the union of at most Δs intervals. Finally, for $d > 1$ and $n > n_0$, an analysis similar to the one in section 5 gives the following recurrence for $Q_d(n, \Delta, s)$:

$$(7.1) \quad Q_d(n, \Delta, s) \leq C \Delta s r^{1-1/d} Q_d(n/r, \Delta, s) + \sum_{\pi \in \Pi} Q_{d-1}(n_\pi, \Delta_1, s_1) + r^c,$$

where $c = d^{O(1)}$, C is a constant depending on d , $\sum_\pi n_\pi \leq n$, and both $|\Pi|$ and $\Delta_1 s_1$ are bounded by $(Ds)^{a_d}$ with $D = O(r^{1/d})$ and $a_d = O(d^4)$. (These are rather crude estimates, but we prefer simplicity.) The leading term of the recurrence relies on the crossing-number bound given in Lemma 4.3. In order to apply that lemma, we need that $r \geq \Delta^d$, which will be ensured by the choice of r given below. The second term corresponds to querying the auxiliary data structures for the exceptional set P^* . The last term covers the time spent in computing the cells of the polynomial partition crossed by the query range γ and for computing the projections γ_π for every $\pi \in \Pi$; here we assume that the choice of r will be such that $r \geq Ds$.

Ultimately, we want to derive that if Δ, s are constants, the recurrence (7.1) implies, with a suitable choice of r and n_0 at each stage,

$$(7.2) \quad Q_d(n, \Delta, s) \leq n^{1-1/d} \log^{B(d,\Delta,s)} n,$$

where $B(d, \Delta, s)$ is a constant depending on d, Δ , and s .

However, as was already mentioned, even if Δ, s are constants initially, later in the recursion they are chosen as tiny powers of n , and this makes it hard to obtain a direct inductive proof of (7.2). Instead, we proceed in two stages. First, in Lemma 7.1 below we derive, without assuming Δ, s to be constants, a weaker bound for $Q_d(n, \Delta, s)$, for which the induction is easier. Then we obtain the stronger bound (7.2) for constant values of Δ, s by using the weaker bound for the $(d - 1)$ -dimensional queries on the exceptional parts, i.e., for the second term in the recurrence (7.1).

A weaker bound for lower-dimensional queries.

LEMMA 7.1. *For every $\nu > 0$ there exists $A_{d,\nu}$ such that, with a suitable choice of r and n_0 ,*

$$(7.3) \quad Q_d(n, \Delta, s) \leq (\Delta s)^{A_{d,\nu}} n^{1-1/d+\nu}$$

for all d, n, Δ, s (with $\Delta s \geq 2$, say).

Remarks. (i) This lemma may look similar to our first result on $\Gamma_{d,\Delta,s}$ -range searching, Theorem 1.2, but there are two key differences—the lemma works for arbitrary point sets, with no general position assumption, and Δ and s are not assumed to be constants.

(ii) Since query time $O(n)$ is trivial to achieve, we may assume $\nu < 1/d$, for otherwise, the bound (7.3) in the lemma exceeds n .

Proof. The case $d = 1$ is trivial because $Q_1(n, \Delta, s) \leq C \Delta s \log_2 n$ clearly implies (7.3), assuming that $A_{d,\nu} \geq 1 + \log_2 C$ and that n is sufficiently large so that $\log_2 n \leq n^\nu$. We assume that (7.3) holds up to dimension $d - 1$ (for all $\nu > 0$, Δ, s , and n), and we establish it for dimension d by induction on n . We consider $A_{d,\nu}$ yet unspecified but sufficiently large; from the proof below one can obtain an explicit lower bound that $A_{d,\nu}$ should satisfy. We set

$$n_0 = n_0(d, \Delta, s, \nu) := (\Delta s)^{dA_{d,\nu}} \quad \text{and} \quad r = (2C \Delta s)^{1/\nu}.$$

This value of n_0 is roughly the threshold where the bound (7.3) becomes smaller than n . Since we assume $\nu < 1/d$, our choice of r satisfies the assumptions $r \geq \Delta^d$ and $r \geq Ds$, as needed in (7.1).

In the inductive step, for $n \leq n_0$,

$$Q_d(n, \Delta, s) \leq n \leq n_0^{1/d} n^{1-1/d} = (\Delta s)^{A_{d,\nu}} n^{1-1/d} \leq (\Delta s)^{A_{d,\nu}} n^{1-1/d+\nu}.$$

So we assume that $n > n_0$ and that the bound (7.3) holds for all $n' < n$. Using the induction hypothesis, i.e., plugging (7.3) into the recurrence (7.1), we obtain

$$(7.4) \quad Q_d(n, \Delta, s) \leq C \Delta s r^{1-1/d} (\Delta s)^{A_{d,\nu}} (n/r)^{1-1/d+\nu} + |\Pi| (\Delta_1 s_1)^{A_{d-1,\nu}} n^{1-1/(d-1)+\nu} + r^c.$$

By the choice of r , the first term of the right-hand side of (7.4) can be bounded by

$$C \Delta s r^{-\nu} (\Delta s)^{A_{d,\nu}} n^{1-1/d+\nu} = \frac{1}{2} (\Delta s)^{A_{d,\nu}} n^{1-1/d+\nu},$$

which is half of the bound we are aiming for.

Next, we bound the second term. We use the estimates $\Delta_1 s_1 \leq (Ds)^{a_d}$, $|\Pi| \leq (Ds)^{a_d}$, and $Ds \leq r$. Then

$$(7.5) \quad \begin{aligned} |\Pi| (\Delta_1 s_1)^{A_{d-1,\nu}} n^{1-1/(d-1)+\nu} &\leq r^{a_d(A_{d-1,\nu}+1)} n^{1-1/(d-1)+\nu} \\ &\leq \frac{r^{a_d(A_{d-1,\nu}+1)}}{n^{1/d(d-1)}} \cdot n^{1-1/d+\nu}. \end{aligned}$$

We choose

$$(7.6) \quad A_{d,\nu} = \frac{d-1}{\nu} a' a_d (A_{d-1,\nu} + 1),$$

where $a' = \log_2(4C)$; i.e., we choose $A_{d,\nu} = d^{\Theta(d)}/\nu^d$. Since $n \geq n_0 = (\Delta s)^{dA_{d,\nu}}$ and $r = (2C\Delta s)^{1/\nu}$, the fraction in (7.5) can be bounded by

$$\frac{r^{a_d(A_{d-1,\nu}+1)}}{n^{1/d(d-1)}} \leq \frac{(2C\Delta s)^{A_{d,\nu}/a'(d-1)}}{(\Delta s)^{A_{d,\nu}/(d-1)}} \leq \left(\frac{2C}{(\Delta s)^{a'-1}} \right)^{A_{d,\nu}/a'(d-1)} \leq 1$$

because $\Delta s \geq 2$.

Finally, recall that $c = d^{O(1)}$, so our choice of $A_{d,\nu}$ (again, choosing a' sufficiently large) ensures that $r^c < n^{1-1/d}$. Hence, the right-hand side in (7.4) is bounded by

$$\frac{1}{2} (\Delta s)^{A_{d,\nu}} n^{1-1/d+\nu} + 2n^{1-1/d+\nu} \leq (\Delta s)^{A_{d,\nu}} n^{1-1/d+\nu},$$

as desired. This establishes the induction step and thereby completes the proof of the lemma. \square

The improved bound for the query time. Now we want to obtain the improved bound (7.2), i.e., $Q_d(n, \Delta, s) \leq n^{1-1/d} \log^B n$, with $B = B(d, \Delta, s)$, assuming that Δ, s are constants and $n > 2$. To this end, in the top-level (d -dimensional) partition tree, we set $r := n^\delta$, where $\delta > 0$ is a suitable small constant to be specified later. Then we use the result of Lemma 7.1 with $\nu := \frac{1}{2d(d-1)}$ for processing the $(d-1)$ -dimensional queries on the sets P_π^* . Thus, in the forthcoming proof, we do induction only on n , while d is fixed throughout.

We choose $n_0 = n_0(d, \Delta, s)$ sufficiently large (we will specify this more precisely later on), and we assume that $n > n_0$ and that the desired bound (7.2) holds for all $n' < n$. In the inductive step we estimate, using the recurrence (7.1), the induction hypothesis, and the bound in (7.3),

$$Q_d(n, \Delta, s) \leq C\Delta sr^{1-1/d}(n/r)^{1-1/d} \log^B(n/r) + |\Pi|(\Delta_1 s_1)^{A_{d-1,\nu}} n^{1-1/(d-1)+\nu} + r^c.$$

The first term simplifies to $(1 - \delta)^B C\Delta s n^{1-1/d} \log^B n$. Thus, if we choose B depending on δ (which is a small positive constant still to be determined) so that $(1 - \delta)^B C\Delta s \leq \frac{1}{2}$, then the first term will be at most half of the target value $n^{1-1/d} \log^B n$. Thus, it suffices to set δ so that the remaining two terms are negligible compared to this value.

For the r^c term, any $\delta \leq 1/2c$ will do. The second term can be bounded, as in the proof of Lemma 7.1, by

$$r^{a_d(A_{d-1,\nu}+1)} n^{1-1/(d-1)+\nu} = \frac{r^{\nu A_{d,\nu}/(a'(d-1))}}{n^\nu} \cdot n^{1-1/d}.$$

Thus, with $\delta \leq a'(d - 1)/A_{d,\nu}$, the term is at most $n^{1-1/d}$. Again, this establishes the induction step and concludes the proof of the final bound for the query time. We remark that our choice of δ requires us to choose

$$B \approx \frac{1}{\delta} \ln(2C\Delta s) \approx \ln(2C\Delta s) \cdot d^{\Theta(d)},$$

making its dependence on d superexponential.

Analysis of storage and preprocessing. Let $S_d(n, \Delta, s)$ denote the size of the data structure on n points in \mathbb{R}^d for $\Gamma_{d,\Delta,s}$ -range searching, with the settings of r and n_0 as described above. For $n \leq n_0 = n_0(d, \Delta, s)$ we have $S_d(n, \Delta, s) = O(n)$. For larger values of n , the space occupied by the root of the partition tree, not counting the auxiliary data structure for the exceptional part P^* , is bounded by r^c , where $c = d^{O(1)}$. Furthermore, since $S_d(n, \Delta, s)$ is at least linear in n , the total size of the auxiliary data structure constructed on P^* is $\sum_{\pi \in \Pi} S_{d-1}(n_\pi, \Delta_1, s_1) \leq S_{d-1}(n^*, \Delta_1, s_1)$, where $n^* = |P^*|$. We thus obtain the following recurrence for $S_d(n, \Delta, s)$:

$$S_d(n, \Delta, s) \leq \sum_{i=1}^t S_d(n_i, \Delta, s) + S_{d-1}(n^*, \Delta_1, s_1) + O(r^c)$$

for $n > n_0 = n_0(d, \Delta, s)$, and $S_d(n, \Delta, s) = O(n)$ for $n \leq n_0$. Using $n_i \leq n/r$, $n^* + \sum_i n_i \leq n$, and $r^c = o(n)$, for both types of choices of r , the recurrence easily leads to

$$S_d(n, \Delta, s) = O(n),$$

where the constant of proportionality depends on d .

It remains to estimate the preprocessing time; here, finally, the parameter $\varepsilon > 0$ in Theorem 1.4 comes into play. Let δ^* be a constant such that $r \leq n^{\delta^*}$ (at all stages of the algorithm). As was remarked in the preceding analysis of the query time, we can make δ^* arbitrarily small, by adjusting various constants (and, generally speaking and as already remarked above, the smaller δ^* , the worse constant $B(d, \Delta, s)$ we obtain in the query time bound).

Let $T_d(n, \Delta, s, \delta^*)$ denote the maximum preprocessing time of our data structure for $\Gamma_{d,\Delta,s}$ -range searching on n points, with $\delta^* > 0$ a constant as above. Using the operation (A1) of section 4, we spend $O(nr^c)$ time to compute $\Omega(f)$ and the partition of P into the exceptional part and the regular parts, and we spend additional $O(nr^c)$ time to compute Π and P_π^* for every $\pi \in \Pi$, where $c = d^{O(1)}$. The total time spent in constructing the secondary data structures for all patches of Π is bounded by $T_{d-1}(n^*, \Delta_1, s_1, \delta^*)$. Hence, we obtain the recurrence

$$T_d(n, \Delta, s, \delta^*) \leq \sum_{i=1}^t T_d(n_i, \Delta, s, \delta^*) + T_{d-1}(n^*, \Delta_1, s_1, \delta^*) + O(nr^c)$$

for $n > n_0$, and $T_d(n, \Delta, s, \delta^*) = O(n)$ for $n \leq n_0$. Using the properties $n_i \leq n/r$ and $n^* + \sum_i n_i \leq n$, a straightforward calculation shows that

$$T_d(n, \Delta, s, \delta^*) = O(n^{1+c\delta^*}),$$

where the constant of proportionality depends on d . Hence, by choosing $\delta^* = \varepsilon/c$, the preprocessing time is $O(n^{1+\varepsilon})$. This concludes the proof of Theorem 1.4. \square

8. Open problems. We conclude this paper by mentioning a few open problems.

(i) A very interesting and challenging problem is, in our opinion, the fast-query case of range searching with constant-complexity semialgebraic sets, where the goal is to answer a query in $O(\log n)$ time using roughly n^d space. There are actually two, apparently distinct, issues. The standard approach to fast-query searching is to parameterize the ranges in Γ by points in a space of a suitable dimension, say t ; then the n points of P correspond to n algebraic surfaces in this t -dimensional “parameter space,” and a query is answered by locating the point corresponding to the query range in the arrangement of these surfaces.

First, the arrangement has $O(n^t)$ combinatorial complexity, and one would expect to be able to locate points in it in polylogarithmic time with storage about n^t . However, such a method is known only up to dimension $t = 4$, and in higher dimension, one again gets stuck at the arrangement decomposition problem, which was the bottleneck in the previously known solution of [2] for the low-storage variant, as was mentioned in the introduction. It would be nice to use polynomial partitions to obtain a better point location data structure for such arrangements, but unfortunately, so far all of our attempts in this direction have failed.

The second issue is, whether the point location approach just sketched is actually optimal. This question is exhibited nicely already in the simple instance of range searching with disks in the plane. The best known solution that guarantees logarithmic query time uses point location in \mathbb{R}^3 and requires storage roughly n^3 , but it is conceivable that roughly quadratic storage might suffice.

(ii) Our range-searching data structure for arbitrary point sets—the one with large fan-out—is so complex and has a rather high exponent in the polylogarithmic factor, because we have difficulty with handling highly degenerate point sets, where many points lie on low-degree algebraic surfaces. This issue appears even more strongly in combinatorial applications, and in that setting it has been dealt with only in rather specific cases (e.g., in dimension 3); see [14, 28, 33] for initial studies. It would be nice to find a construction of suitable “multilevel polynomial partitions” that would cater to such highly degenerate input sets, as touched upon in [14, 33].

(iii) Another open problem, related to the construction of polynomial partitions, is the fast evaluation of a multivariate polynomial at many points, as briefly discussed at the end of section 3.

Acknowledgment. We thank the anonymous referees for their useful comments on this paper.

REFERENCES

- [1] P. K. AGARWAL AND J. ERICKSON, *Geometric range searching and its relatives*, in Advances in Discrete and Computational Geometry, B. Chazelle, J. E. Goodman, and R. Pollack, eds., AMS, Providence, RI, 1998, pp. 1–56.
- [2] P. K. AGARWAL AND J. MATOUŠEK, *On range searching with semialgebraic sets*, Discrete Comput. Geom., 11 (1994), pp. 393–418.
- [3] S. BARONE AND S. BASU, *Refined bounds on the number of connected components of sign conditions on a variety*, Discrete Comput. Geom., 47 (2012), pp. 577–597.
- [4] S. BASU, R. POLLACK, AND M.-F. ROY, *On the number of cells defined by a family of polynomials on a variety*, Mathematika, 43 (1996), pp. 120–126.
- [5] S. BASU, R. POLLACK, AND M.-F. ROY, *Algorithms in Real Algebraic Geometry*, Algorithms Comput. Math. 10, Springer-Verlag, Berlin, 2003.
- [6] T. M. CHAN, *Optimal partition trees*, Discrete Comput. Geom., 47 (2012), pp. 661–690.
- [7] B. CHAZELLE, *Lower bounds on the complexity of polytope range searching*, J. Amer. Math. Soc., 2 (1989), pp. 637–666.
- [8] B. CHAZELLE, H. EDELSBRUNNER, L. J. GUIBAS, AND M. SHARIR, *A singly exponential stratification scheme for real semi-algebraic varieties and its applications*, Theoret. Comput. Sci., 84 (1991), pp. 77–105. Also in Proceedings of the 16th International Colloquium on Automata, Languages and Programming, Lecture Notes in Comput. Sci. 372, Springer, Berlin, 1989, pp. 179–193.
- [9] GY. ELEKES, H. KAPLAN, AND M. SHARIR, *On lines, joints, and incidences in three dimensions*, J. Combin. Theory Ser. A, 118 (2011), pp. 962–977.
- [10] L. GUTH AND N. H. KATZ, *Algebraic methods in discrete analogs of the Kakeya problem*, Adv. Math., 225 (2010), pp. 2828–2839.
- [11] L. GUTH AND N. H. KATZ, *On the Erdős distinct distances problem in the plane*, arXiv:1011.4105, 2010.
- [12] H. HIRONAKA, *Triangulations of algebraic sets*, in Proceedings of Symposia Pure Math 29, R. Hartshorne, ed., AMS, Providence, RI, 1975, pp. 165–185.
- [13] H. KAPLAN, J. MATOUŠEK, AND M. SHARIR, *Simple proofs of classical theorems in discrete geometry via the Guth–Katz polynomial partitioning technique*, Discrete Comput. Geom., 48 (2012), pp. 499–517.
- [14] H. KAPLAN, J. MATOUŠEK, Z. SAFERNOVÁ, AND M. SHARIR, *Unit distances in three dimensions*, Combin. Probab. Comput., 21 (2012), pp. 597–610.
- [15] K. KEDLAYA AND CH. UMANS, *Fast modular composition in any characteristic*, in Proceedings of the 49th Annual IEEE Symposium in Foundations of Computer Sciences, 2008, pp. 146–155.
- [16] CH. KNAUER, H. R. TIWARI, AND D. WERNER, *On the computational complexity of ham-sandwich cuts, Helly sets, and related problems*, in Proceedings of the 28th Annual Symposium of Theoretical Aspects of Computer Science, 2011, pp. 649–660.
- [17] V. KOLTUN, *Almost tight upper bounds for vertical decompositions in four dimensions*, J. ACM, 51 (2004), pp. 699–730.
- [18] K. G. LARSEN, *On range searching in the group model and combinatorial discrepancy*, in Proceedings of the 52nd Annual IEEE Symposium Foundation of Computer Sciences, 2011, pp. 542–549.
- [19] C.-Y. LO, J. MATOUŠEK, AND W. STEIGER, *Ham-sandwich cuts in \mathbb{R}^d* , in Proceedings of the 24th Annual ACM Symposium on the Theory of Computer Science, 1992, pp. 539–545.
- [20] J. MATOUŠEK, *Efficient partition trees*, Discrete Comput. Geom., 8 (1992), pp. 315–334.
- [21] J. MATOUŠEK, *Range searching with efficient hierarchical cuttings*, Discrete Comput. Geom., 10 (1993), pp. 157–182.
- [22] J. MATOUŠEK, *Geometric range searching*, ACM Comput. Surveys, 26 (1994), pp. 421–461.
- [23] J. MATOUŠEK, *Using the Borsuk–Ulam Theorem*, in Lectures on Topological Methods in Combinatorics and Geometry Series, Springer-Verlag, Berlin, 2003.

- [24] M. NÜSKEN AND M. ZIEGLER, *Fast multipoint evaluation of bivariate polynomials*, in Proceedings of the 12th Annual European Symposiums on Algorithms, 2004, Lecture Notes in Comput. Sci. 3221, Springer, Berlin, pp. 544–555.
- [25] J. T. SCHWARTZ AND M. SHARIR, *On the Piano Movers' problem: II. General techniques for computing topological properties of real algebraic manifolds*, Adv. in Appl. Math., 4 (1983), pp. 298–351.
- [26] M. SHARIR AND P. K. AGARWAL, *Davenport Schinzel Sequences and Their Geometric Applications*, Cambridge University Press, Cambridge, UK, 1995.
- [27] M. SHARIR AND H. SHAUL, *Semi-algebraic range reporting and emptiness searching with applications*, SIAM J. Comput., 40 (2011), pp. 1045–1074.
- [28] J. SOLYMOSI AND T. TAO, *An incidence theorem in higher dimensions*, Discrete Comput. Geom., 48 (2012), pp. 255–280.
- [29] A. H. STONE AND J. W. TUKEY, *Generalized "sandwich" theorems*, Duke Math. J., 9 (1942), pp. 356–359.
- [30] H. E. WARREN, *Lower bound for approximation by nonlinear manifolds*, Trans. Amer. Math. Soc., 133 (1968), pp. 167–178.
- [31] A. C. YAO AND F. F. YAO, *A general approach to d -dimensional geometric queries*, in Proceedings of the 17th Annual ACM Symposium on the Theory of Computing, pp. 163–168.
- [32] C. K. YAP, *A geometric consistency theorem for a symbolic perturbation scheme*, J. Comput. System Sci., 40 (1990), pp. 2–18.
- [33] J. ZAHL, *An improved bound on the number of point-surface incidences in three dimensions*, arXiv:1104.4987, 2011.