Reporting points in halfspaces

JIŘÍ MATOUŠEK
Department of Applied Mathematics
Charles University
Malostranské nám. 25, 118 00 Praha 1, Czechoslovakia

Abstract

We consider the *halfspace range reporting* problem: given a finite set P of points in E^d, preprocess it so that given a query halfspace γ, the points of $P \cap \gamma$ can be reported efficiently. Answering a question posed by M. Sharir, we show that with almost linear storage, this problem can be solved substantially more efficiently that the more general simplex range searching problem. We give a data structure for halfspace range reporting in dimensions $d \geq 4$ using $O(n \log\log n)$ space and $O(n \log n)$ deterministic preprocessing time and with query time $O(n^{1-1/\lfloor d/2 \rfloor}(\log n)^{O(1)} + k)$, where $k = |P \cap \gamma|$ (in dimensions $d = 2,3$, efficient solutions were known). For the *halfspace emptiness problem*, where we only want to know whether $P \cap \gamma = \emptyset$, we can achieve query time $O(n^{1-1/\lfloor d/2 \rfloor} 2^{O(\log^* n)})$ with a linear space and $O(n^{1+\delta})$ preprocessing (where $\delta > 0$ is arbitrary but fixed). Applications of these results will be surveyed elsewhere.

1 Introduction

One of central themes in computational geometry is the development of efficient *range searching algorithms*. We will consider *halfspace range reporting*: Preprocess a set P of n points in E^d so that, given any query halfspace γ, the points in $P \cap \gamma$ can be reported efficiently. (As it is typical in computational geometry, we will consider the space dimension d as a small fixed number, thus something depending on the dimension only will be a constant for us.) This problem is obviously a special case of a more general *simplex range searching problem*, where we want to report all points of P contained in a query simplex σ, or (still more generally) to evaluate $\sum_{p \in P \cap \sigma} w(p)$, where $w : P \to S$ is a function assigning to the points weights belonging to some semigroup. The general simplex range searching problem is now probably close to being solved optimally.

In a *range reporting* problem, we want to list all points contained in a query range. Chazelle [Cha86] proposed a strategy for reporting problems, called "filtering search", whose basic observation is the following: If we report k points of the answer, we must in any case spend $\Omega(k)$ time on this, and thus we can spend $O(k)$ more time on computing the answer without affecting the asymptotic efficiency. Hence for a large k, we may use a less efficient procedure than for a small k. With such an approach, in some situations the halfspace range reporting problem can be solved more efficiently than it follows from general results on the simplex range searching. This only seems to work for halfspaces; no improved reporting results are known e.g. for simplices or slabs.

In this paper we show that in general the halfspace range reporting admits a substantially more efficient solution than guaranteed by the results for the simplex range searching. Before we state our results, let us briefly recall the known results on the above mentioned two problems.

The efficiency of a range searching algorithm depends on the amount of space (and also of preprocessing time) which we are allowed to use for the data structure. For the simplex range searching problem, Chazelle [Cha89] established a lower bound for this tradeoff (which is valid only under some restrictions on the kind of algorithm used, but so far all known algorithms satisfy these restrictions and there does not seem to be much hope for circumventing the lower bound): In dimension d and given m units of storage, the worst-case query time is $\Omega((n/\log n)/m^{1/d})$. In the plane, the bound sharpens to $\Omega(n/\sqrt{m})$; this indicates that for higher dimensions, the logarithmic factor might be only a product of the proof technique.

This tradeoff has nearly been attained by Chazelle et al. [CSW90]. An improved and simpler solution was given in [Mat91], where it is shown that that the simplex range searching problem can be solved with $O(n\log n)$ deterministic preprocessing time, using $O(n)$ space and with $O(n^{1-1/d}(\log n)^{O(1)})$ query time. Also variants of this result are given there, where the polylogarithmic factor in the query time is replaced by a smaller one, paying the price of a slightly longer preprocessing and a more complicated algorithm. We will be able to add one more result of such kind in this paper. The simplex range reporting problem can be solved within the same bounds, with the number of reported points k added to the query time. Of the rich previous literature concerning the simplex range searching problem, let us mention e.g. [Wil82], [YY85], [HW87], [?].

The halfspace range reporting problem was solved it optimally in dimension 2 by Chazelle et al. [CGL85]. With $O(n)$ space and $O(n\log n)$ (deterministic) preprocessing time, one can do the reporting in time $O(\log n + k)$, where k is the number of points reported. For dimension 3, a solution with nearly linear space and $O(\log n + k)$ query time was given by Chazelle and Preparata [CP86], and Aggarwal et at. [AHL90] give an improvement with $O(n)$ space and $O(n\log n)$ preprocessing time.

In higher dimensions, Clarkson [Cla88] showed that with $O(n^{\lfloor d/2\rfloor+\delta})$ space and the same order of magnitude of expected preprocessing time one can answer a query in $O(\log n + k)$ time ($\delta > 0$ is an arbitrarily small but fixed number). A heuristic reason which we will indicate below suggests that this amount of space might be close to optimal. M. Sharir asked whether one can achieve some improvement over the simplex range reporting solution also with linear or almost linear space, in a dimension $d \geq 4$. We show the following:

Theorem 1.1 *Given a set of n points in E^d, $d \geq 4$, one can preprocess it for halfspace range reporting in time $O(n\log n)$, store the results in space $O(n\log\log n)$ and then answer queries in time $O(n^{1-1/\lfloor d/2\rfloor}(\log n)^{O(1)} + k)$, where k is the number of reported points.*

Using a combination of the above result with the Clarkson's data structure for halfspace range reporting with $O(\log n + k)$ query time, we can achieve the following tradeoff:

Corollary 1.2 *Given a set of n points in E^d, $d \geq 4$, and a parameter m, $n \leq m \leq n^{\lfloor d/2\rfloor}$, the halfspace range reporting problem can be solved with space and expected preprocessing time $O(m^{1+\delta})$ (with an arbitrarily small fixed $\delta > 0$) and query time $O(n(\log n)^{O(1)}/m^{1/\lfloor d/2\rfloor} + k)$.* □

We omit the proof of this corollary, since it closely follows the method achieving a similar tradeoff for simplex range searching in [CSW90].

A special case of the reporting problem is an *emptiness* problem; in our case we only want to decide whether a query halfspace contains some point of P. It is well-known that the convex hull of a n-point set in E^d can have $\Omega(n^{\lfloor d/2\rfloor})$ facets. It seems likely that an algorithm deciding the halfspace emptiness problem for such a point set within a polylogarithmic query time should have access to some representation of the convex hull, and thus use a space proportional to its complexity. This is only a heuristic reason, and it would be interesting to prove a lower bound of such kind rigorously, perhaps extending the methods of Chazelle [Cha89]. We even venture to conjecture that with a space m, the query time for the halfspace emptiness problem has to be $\Omega(n/m^{1/\lfloor d/2\rfloor})$, and thus in particular that Theorem 1.1 is close to optimal. For the halfspace emptiness problem, we can improve the result a little:

Theorem 1.3 *Given a set of n points in* E^d, $d \geq 4$, *one can build in time* $O(n^{1+\delta})$ *(where* $\delta > 0$ *is arbitrary but fixed) a linear-sized data structure, which can decide whether a query halfspace contains a point of P in time* $O(n^{1-1/\lfloor d/2 \rfloor} 2^{O(\log^* n)})$.

In [Mat91] some improvements to the simplex range searching with group weights in dimensions 2, 3 are treated, which use efficient halfspace range reporting structures. The proofs given there show that as soon as we have a data structure for halfspace range reporting with $O(n^{1-1/d-\delta} + k)$ query time, where δ is some positive fixed number, we can achieve $O(n^{1-1/d}(\log n)^{O(1)} + k)$ time for reporting segments crossing a query hyperplane, and this result in turn yields the following:

Theorem 1.4 *Given a set of n points in* E^d *equipped by weights belonging to a group (i.e. which can be subtracted), one can solve the simplex range searching problem with* $O(n^{1+\delta})$ *preprocessing, space* $O(n)$ *and query time* $O(n^{1-1/d} 2^{O(\log^* n)})$. □

Applications of our results on other computational geometry problems will be surveyed elsewhere ([AM91]). One obtains improvements of ray shooting algorithms in various situations and a possibility of preprocessing time/query time tradeoffs in such algorithms. For example, Chazelle et al. [CEG89] develop an algorithm for ray shooting in polyhedral terrains with a nearly quadratic space and a polylogarithmic query time. With our results, one can get an algorithm using $O(n^{1+\delta})$ space and $O(n^{1/2+\delta})$ query time.

The methods used to obtain Theorem 1.1 are similar to the methods of [Mat91], only a different version of a "cutting lemma" and of a "partition theorem" are needed; these will be developed in sections 2 and 3, while section 4 describes the range reporting algorithm and discusses the preprocessing issues. In section 5 we bring one application of our auxiliary results (concerning ε-nets with respect to halfspaces in E^3), which turns out to be very similar to an unpublished result of Clarkson [Cla89].

Throughout the paper we will assume that the point sets and collections of hyperplanes considered are in general position where convenient.

2 Shallow cutting lemma

Let us begin by some definitions. We will assume that H is a collection of n hyperplanes in E^d. For a simplex s, let H_s denote the collection of hyperplanes of H intersecting its interior. An *ε-cutting* for H is a collection Ξ of (possibly unbounded) closed d-dimensional simplices with disjoint interiors, which cover E^d and such that $|H_s| \leq \varepsilon n$ for every $s \in \Xi$. The *size* of an ε cutting is the number of its simplices.

The concept of $(1/r)$-cutting originated in applications of random sampling to computational geometry problems (see [HW87], [Cla88]), and explicitly it was introduced in [Mat90]. It turns out to be an useful tool in many computational geometry problems.

Chazelle and Friedman [CF90] proved that for every H and $r \leq n$ there exists a $(1/r)$-cutting of size $O(r^d)$ for H (which is asymptotically the best possible size). We will call this result Cutting lemma.

For the halfspace range reporting results in this paper, we will need a modification of Cutting lemma, where the simplices should cover not all E^d, but only a single cell in the arrangement of H. We will prove a slightly more general result. Let o be a fixed point in E^d; we define the *level* of a point $x \in \mathrm{E}^d$ (relative to H) as the number of hyperplanes of H crossed by the open segment ox. In the literature, the reference point o is usually chosen to be $(0, \ldots, 0, -\infty)$; then the level of a point is just the number of hyperplanes of H lying below it. The general case can be reduced to this special one by a suitable projective transformation.

Clarkson [Cla88] proved that the number of vertices of the arrangement of H of level at most k is $O(n^{\lfloor d/2 \rfloor} k^{\lceil d/2 \rceil})$; there exist arrangements for which this bound is tight.

Let $k \leq n$. We say that a collection Ξ of simplices is an *ε-cutting for the $(\leq k)$-level of H*, provided that the simplices of Ξ cover all points of level at most k and $|H_s| \leq \varepsilon n$ for every $s \in \Xi$.

Theorem 2.1 (Shallow cutting lemma) *Let H be a collection of n hyperplanes in* E^d *and* $k, r \leq n$ *parameters, and let* $q = k(r/n) + 1$. *There exists a* $(1/r)$*-cutting* Ξ *for the* $(\leq k)$*-level of H, consisting of* $O(r^{\lfloor d/2 \rfloor} q^{\lceil d/2 \rceil})$ *simplices. Moreover, we may* *assume that the complement of the union of Ξ is contained in a union of $O(r)$ halfspaces defined by hyperplanes of H.*

In particular, a single cell of an arrangement can be covered by $O(r^{\lfloor d/2 \rfloor})$ simplices, such that the interior of each is intersected by at most n/r hyperplanes.

Let us also remark that although we will assume a general position for the hyperplanes in the proof, this assumption can be removed by a perturbation argument (see e.g. [Ede87]), and the lemma is applicable also for multisets of hyperplanes, or for weighted collections (where a nonnegative weight function w is defined on the hyperplanes and we require that $w(H_s) \leq w(H)/r$ for every simplex s of a $(1/r)$-cutting).

The proof is similar to the proof of Cutting lemma due to Chazelle and Friedman [CF90]. Alternatively, we could also use a slightly different approach of [Cla88].

In the proof one uses a special kind of triangulation of cells of an arrangement of hyperplanes, called *canonical triangulation* (some authors use *bottom-vertex triangulation*). The definition and some properties of the canonical triangulation can be found in [CF90]; here we will recall only the properties directly needed for the proof. For a subcollection $R \subseteq H$ of hyperplanes, let $CT(R)$ denote the set of simplices in the canonical triangulation of the arrangement of R and let $CT_{\leq k}(R)$ denote the set of those simplices of $CT(R)$ whose all points have level at most k (relative to H!).

Let H and the parameter r be chosen. For a simplex s, let the *excess* of s be the number $|H_s|(r/n)$.

The proof is based on the following lemma:

Lemma 2.2 *Let R be a random sample of hyperplanes of H, where each hyperplane of H is drawn independently with probability $p = r/n$. Then the expected number $n(p,t,k)$ of simplices with excess at least t in $CT_{\leq k}(R)$ satisfies $n(p,t,k) = O(2^{-t}n(p/t,0,k))$.*

This lemma essentially follows from general results of [CF90] (with minor modifications, since we use a different probability space).

We show how this lemma implies Shallow cutting lemma. We observe that the number of simplices of $CT_{\leq k}(R)$ is bounded by the number of vertices of the arrangement of R which have level $\leq k$ relative to H. For each vertex of the arrangement of H of level $\leq k$, the probability that it becomes a vertex of the arrangement of R is p^d, and applying the Clarkson's result on the number of such vertices, we get that the expected number $n(p,0,k)$ of simplices in $CT_{\leq k}(R)$ is bounded as follows:

$$n(p,0,k) = O(p^d n^{\lfloor d/2 \rfloor} k^{\lceil d/2 \rceil}). \qquad (1)$$

The desired $(1/r)$-cutting for $(\leq k)$-level of H is constructed by a refinement of an appropriate portion of $CT(R)$. We proceed as follows: Let Ξ_0 be the set of all simplices of $CT(R)$ which contain some point of level $\leq k$ (relative to H). For every such simplex s, let us consider the collection H_s, and let Ξ_s be a $(1/t)$-cutting for H_s of size $O(t^d)$, where $t = t(s)$ is the excess of s (such a cutting exists by Cutting lemma, and actually we would suffice with size $O(t^c)$ for any constant c, which is much easier to obtain than the tight bound). We take the intersection of every simplex $s' \in \Xi_s$ with s and we triangulate it. The collection of simplices appearing in these triangulations for all s will form our cutting Ξ. It is easy to see that is is a $(1/r)$-cutting for $(\leq k)$-level of H, and it remains to bound its size, which amounts to bounding the expected value S of the sum $\sum_{s \in \Xi_0} t(s)^d$. Since a simplex $s \in \Xi_0$ with excess t has all points of level at most $k + tn/r$, we can estimate S by

$$\sum_{t=1}^{\infty} n(p,t,k+tn/r),$$

and using Lemma 2.2 and the bound (1) for $n(p/t,0,k)$, we get the estimates

$$S \leq \sum_{t=1}^{\infty} 2^{-t} O(n(p/t,0,k+\frac{tn}{r})) \leq$$

$$\sum_{t=1}^{\infty} 2^{-t} O((p/t)^d n^{\lfloor d/2 \rfloor}(k+\frac{tn}{r})^{\lceil d/2 \rceil}) \leq$$

$$\leq \sum_{t=1}^{\infty} 2^{-t} O(r^{\lfloor d/2 \rfloor} t^{-\lfloor d/2 \rfloor}(\frac{kr}{nt}+1)^{\lceil d/2 \rceil}) =$$

$$O(r^{\lfloor d/2 \rfloor}(\frac{kr}{n}+1)^{\lceil d/2 \rceil}),$$

which is just the claimed bound.

It remains to show the additional claim about covering of the complement of Ξ in Theorem 2.1.

Obviously the complement is covered by halfspaces defined by the sample R in our construction. We know that the expected number of such halfspaces is r, and it is easy to infer that there exists a sample for which both the number of hyperplanes and the size of the cutting do not exceed the expectations more than (say) four times. □

3 Partition theorem for shallow hyperplanes

Let us repeat a definition from [Mat91]. Let P be a n-point set in $\mathbb{E}^d$. A *simplicial partition* for P is a collection $\Pi = \{(P_1, s_1), \ldots, (P_m, s_m)\}$, where the P_i's are nonempty sets (called the *classes* of Π) forming a partition of P and each s_i is a relatively open simplex (not necessarily full-dimensional) containing P_i.

The number m is the *size* of the partition. We say that a simplicial partition Π is *fine* if $|P_i| \leq 2n/m$ for each i, where m is the size of Π and n is the cardinality of the underlying point set.

If h is a hyperplane and s a simplex, we say that h *crosses* s if $h \cap s \neq \emptyset$ and $s \not\subseteq h$. For a hyperplane h, we define the *crossing number* of h (relative to Π) as the number of simplices among the s_i's crossed by h.

The basis of the efficient algorithms in [Mat91] was so-called Partition theorem, which stated that for every a n-point set $P \subseteq \mathbb{E}^d$ and a parameter $r \leq n$ there exists a fine simplicial partition Π of size $O(r)$ for P, such that the crossing number of every hyperplane relative to Π is $O(r^{1-1/d})$.

In this section we modify the proof of this theorem to obtain a result which allows us to treat halfspace range reporting. We say that a hyperplane h is *k-shallow* (relative to P), if one of the open halfspaces determined by h contains no more than k points of P.

Theorem 3.1 (Partition theorem for shallow hyperplanes) *Let P be a set of n points in $\mathbb{E}^d$, $d \geq 4$, and k, r parameters satisfying $k \leq n/r$. Then there exists a fine simplicial partition Π of P, of size $O(r)$ and such that the crossing number of any k-shallow hyperplane relative to Π is $O(r^{1-1/\lfloor d/2 \rfloor})$.*

The whole proof is quite similar to the one in [Mat91], which in turn uses ideas of Welzl (see [?]), and some its parts will be only sketched here. The heart of the proof is the following lemma:

Lemma 3.2 *Let P, n be as above, $k \leq 2n/r$ and suppose that r exceeds some prescribed constant. Let Q be a set of k-shallow hyperplanes. Then there exists a subset $P' \subseteq P$ of at least $n/2$ points and a simplicial partition $\Pi = \{(P_1, s_1), \ldots, (P_m, s_m)\}$ for P' with $|P_i| = \lfloor n/r \rfloor$ for all i and such that the crossing number of every hyperplane of Q relative to Π is $O(r^{1-1/\lfloor d/2 \rfloor} + \log |Q|)$.*

Proof: We will inductively construct disjoint sets $P_1, P_2, \ldots \subset P$ of size $\lfloor n/r \rfloor$ and simplices $s_1, s_2, \ldots, P_i \subseteq s_i$. The construction finishes when $|P_1 \cup \ldots \cup P_m| \geq n/2$. When $P_1, \ldots, P_{i-1}$ have been constructed, we construct P_i as follows: For a hyperplane $h \in Q$, let $\kappa_i(h)$ denote the number of simplices among $s_1, \ldots, s_{i-1}$ crossed by h. We define a weighted collection Q, w_i by setting $w(h) = 2^{\kappa_i(h)}$ for every $h \in Q$. According to Theorem 2.1 (and the remark following it), we let Ξ_i be a $(1/t)$-cutting for the (≤ 0)-level of Q, w_i such that the total number of all faces of its simplices it at most $r/3$, with $t = \Omega(r^{1/\lfloor d/2 \rfloor})$. Let P_i^* be the set of points of P covered by the simplices of Ξ_i; since the complement of the union of simplices of Ξ_i is contained in a union of at most t halfspaces defined by k-shallow hyperplanes, we get that $|P_i^*| \geq n - tk$. Since we assumed that r was large enough, we may suppose that $tk < n/6$ and thus $|P_i^*| \geq 5n/6$. Let us put $P_i' = P_i^* \setminus (P_1 \cup \ldots \cup P_{i-1})$; this set has more than $n/3$ points and so there is a relatively open face of a simplex of Ξ_i containing at least n/r points of P_i'. We shrink it so that it contains exactly $\lfloor n/r \rfloor$ points and call it s_i, and we put $P_i = P_i' \cap s_i$.

We establish the bound on the crossing numbers of the hyperplanes of Q relative to the simplicial partition $\Pi = \{(P_1, s_1), \ldots, (P_m, s_m)\}$. The weight $w_m(h)$ of a hyperplane $h \in Q$ with crossing number κ is equal to 2^κ. On the other hand, it is easy to see that each newly added simplex s_i is crossed by hyperplanes of Q of total weight $O(w_i(Q)/r^{1/\lfloor d/2 \rfloor})$. For these hyperplanes, the weight is doubled in the i-th step and for the others it remains unchanged, thus $w_{i+1}(Q) \leq w_i(Q)(1+O(1/r^{1/\lfloor d/2 \rfloor}))$. From this one computes that $\kappa \leq \log w_m(h) = O(r^{1-1/\lfloor d/2 \rfloor} + \log |Q|)$ as claimed; see e.g. [Mat91] for more details. □

The next step is to choose a "test set" of hyperplanes Q, guaranteeing that the crossing number of *no* k-shallow hyperplane will exceed the required bound.

Lemma 3.3 (Shallow test set lemma) *Let P, n be as above, let k, r be parameters. Then there exists a set Q of $O(r^d)$ $(k + n/r)$-shallow hyperplanes, such that for any simplicial partition $\Pi = \{(P_1, s_1), \ldots, (P_m, s_m)\}$ satisfying $|P_i| \geq \lfloor n/r \rfloor$ for every i the following holds: If κ_0 is the maximum of crossing numbers of hyperplanes of Q relative to Π, then the crossing number of any k-shallow hyperplane relative to Π is bounded by $(d+1)\kappa_0 + 1$.*

Proof: Let $H = D(P)$ be the collection of hyperplanes dual to the points of P. Let us choose a $(1/r)$-cutting Ξ of size $O(r^d)$ for H, and let V be the set of all vertices of simplices of Ξ, which have of level at most $k + n/r$ or at least $n - k - n/r$. Let us put $Q = D(V)$.

Clearly $|Q| = O(r^d)$, and also all hyperplanes of Q are $(k + n/r)$-shallow. As for the desired test set property, we essentially repeat an argument of Chazelle et al. [CSW90] or [Mat91]. Let h be any k-shallow hyperplane and let G be the set of vertices of a simplex s of Ξ containing the dual point $D(h)$. All the vertices of G are included in V (because h was k-shallow and $D(H)$ is separated from any vertex of G by at most n/r hyperplanes). Now each of the $d+1$ hyperplanes dual to the points of G crosses at most κ_0 simplices of the simplicial partition Π, and it remains to bound the number of simplices of Π which are crossed by h but by no hyperplane of $D(G)$. Such simplices s_i must be completely contained in the zone of h in the arrangement of $D(G)$, and hence this zone must contain also their corresponding point sets P_i. Now any point of P lying in the zone of h in the arrangement of $D(G)$ dualizes to a hyperplane of H crossing the simplex s, and there are at most n/r of such hyperplanes in H. Hence the zone of h may contain at most this many points of P and thus at most one simplex of Π. □

Let us remark that we could get $O(r^{\lfloor d/2 \rfloor})$ test hyperplanes if we used the Shallow cutting lemma, but the above approach will be more convenient for us for algorithmic purposes.

The **proof of Theorem 3.1** is now finished by a repeated application of Lemma 3.2. We begin with the set $P_0 = P$. In i-th step, we have a set $P_i \subseteq P$ of cardinality $n_i \leq n/2^i$ and we set $r_i = r/2^i$. By Shallow test set lemma 3.3, we can find a set Q_i of $O(r_i^d)$ k_i-shallow hyperplanes, $k_i = k + n_i/r_i$, satisfying the conclusion of that lemma with n_i standing for n and r_i standing for r (but k is the same k as in the statement of Theorem 3.1). We then use Lemma 3.2 and find a simplicial partition Π_i of at least half of the set P_i, such that the crossing number of any hyperplane of Q_i (and thus of any k-shallow hyperplane) is $O(r_i^{1-1/\lfloor d/2 \rfloor})$, and with classes of size $\lfloor n_i/r_i \rfloor \leq \lfloor n/r \rfloor$. We let P_{i+1} be the set of points of P_i not contained in Π_i and repeat the construction, until r_{i+1} becomes too small to apply Lemma 3.2. Then the number of points in P_{i+1} is only $O(n/r)$, and we can cover these remaining points by $O(1)$ additional large simplices, assigning n/r points to each of these simplices. We combine this simplicial partition with the ones built in the previous steps, and we obtain a simplicial partition Π of P. The crossing number of any k-shallow hyperplane relative to Π will be $O(1) + \sum_i O(r_i^{1-1/\lfloor d/2 \rfloor}) = O(r^{1-1/\lfloor d/2 \rfloor})$. The size of Π is at most $O(1) + r_0/2 + r_1/2 + \ldots = O(1) + r/2 + r/4 + \ldots \leq 2r$ and we also get that Π is fine. □

4 Reporting points in halfspaces

We are now ready to describe a data structure for the algorithm from Theorem 1.1. It will be a kind of partition tree. Each node v of the tree will correspond to some subset of $P_v \subseteq P$, the root will correspond to the whole P. The sets corresponding to children of a node will form a partition of the set of their father, and the sets for leaves will have size not exceeding a suitably chosen constant.

Let $\alpha > 0$ be a suitable constant smaller than $1/d$ (we will put more requirements on it when discussing the preprocessing algorithm). We begin the construction in the root of the partition tree and we proceed recursively down the tree. For a node v, we set $m = |P_v|$, $r = m^\alpha$ and we choose (according to Theorem 3.1) a fine simplicial partition Π_v of size $O(r)$ for P_v, such that the crossing number of any (m/r)-shallow hyperplane will be at most $\kappa_v = O(r^{1-1/\lfloor d/2 \rfloor})$. The sets in Π_v will correspond to $O(r)$ children of v. We moreover associate with the node v a linear-sized data structure which allows to report points of P_v in a query halfspace in time $O(m^{1-1/d}(\log m)^{O(1)} + k)$, where k is the number of points reported; we use the simplex range searching result of [Mat91], mentioned in the introduction.

Since each level of the tree uses $O(n)$ space and the depth of the tree is $O(\log\log n)$, the total space is $O(n \log\log n)$. The query answering algorithm is very simple. We start in the root, and being in a node v, we perform the following recursive procedure: We detect all simplices of Π_v crossed by the boundary hyperplane of the query halfspace γ. If the number of such simplices is greater than κ_v (in that case we know that the hyperplane is not (m/r)-shallow relative to P_v), we use the auxiliary data structure of that node to report the points of $P_v \cap \gamma$. Otherwise we report all points in simplices completely contained in γ, and we proceed recursively down the tree for the simplices crossing the boundary hyperplane, eventually stopping at the leaves, where we answer the query trivially by inspecting all points.

First we estimate the work done in reporting using the auxiliary data structures. Should such a data structure be used in a node v with $|P_v| = m$, it must be $k = |P_v \cap \gamma| \geq m/r = m^{1-\alpha}$ with $\alpha < 1/d$, and the auxiliary data structure uses time $O(m^{1-1/d}(\log m)^{O(1)} + k) = O(k)$, thus the total work is proportional to k. The time spent on the other parts of the algorithm satisfies the recurrence

$$T(m) = O(1) \text{ for } m = O(1),$$

$$T(m) \leq O(r) + O(r^{1-1/\lfloor d/2 \rfloor})T(2m/r)$$

with $r = m^\alpha$, whose solution is $T(n) = O(n^{1-1/\lfloor d/2 \rfloor}(\log n)^{O(1)})$.

Let us now turn to the issue of preprocessing. It is easy to check that it is enough to show the following: For a n-point set P, $r = n^\alpha$ (with a sufficiently small fixed $\alpha > 0$) and $k = n/r$, a simplicial partition as in Shallow partition theorem 3.1 can be found in time $O(n \log n)$. With a suitable choice of α, we are free to use steps of complexity $O(r^{O(1)})$ and also $O(n^{1-\delta} r^{O(1)})$ ($\delta > 0$ fixed) in our algorithm.

Let us look at the components of the proof of Partition theorem for shallow hyperplanes. In Shallow test set lemma, we should find a $(1/r)$-cutting for a collection of n hyperplanes, which can be done in time $O(n \log r)$ for the values of r considered according to [Mat91]. Then we should discard the vertices whose corresponding dual hyperplanes are not $(2n/r)$-shallow; since we have polynomially many vertices in r, we may use any halfspace range counting algorithm with $O(n \log n)$ time preprocessing and a sublinear query time ($O(n^{1-\delta})$).

Then in the construction of the simplicial partition itself, all steps can be performed in time polynomial in r (for the construction of a shallow cutting, one may apply the method of conditional probabilities), the only exception being the counting of the number of points in the simplices of a cutting. Again it is enough to have a simplex range counting structure with suitable parameters – see [Mat91], where the same problem is addressed. This finishes the proof of Theorem 1.1. □

The key to the improvement for the emptiness problem (Theorem 1.3) is making the partition tree much more shallow. Namely, in a node with $|P_v| = m$, we choose $r = m/(\log m)^C$, where C is a suitable constant. We also will not need the auxiliary data structures for simplex range reporting in P_v, but we will use another auxiliary data structure instead. We consider the set V_v of all vertices of simplices of the simplicial partition associated with the node v, and we build the halfspace range reporting structure described in Theorem 1.1 for it. This auxiliary structure has size $O(r \log\log r) = O(m/\log m)$ in a m-point node, and this implies that the whole data structure uses only $O(n)$ space.

The query answering for the emptiness problem for a query halfspace γ again starts in the root, and in a node v the algorithm is as follows: We set $t_v = f(|V_v|, (d+1)\kappa_v)$, where $f(n,k) = O(n^{1-1/\lfloor d/2 \rfloor}(\log n)^{O(1)} + k)$ is an upper bound on the complexity of the query answering algorithm from Theorem 1.1. We run the query answering algorithm for reporting points of V_v in γ, terminating after t_v time units if it has not finished yet. Now if it has run at least t_v time units, we know that $|V_v \cap \gamma| > (d+1)\kappa_v$, and thus γ cannot be empty (either too many simplices cross the bounding hyperplane, or some simplex is completely contained in γ). On the other hand, if all points of V_v in the query halfspace were reported, we can list all simplices crossing the query halfspace γ. We thus check if some of them is completely contained in γ (then we may stop), and if not we perform a recursion for the simplices crossing the boundary of γ.

The value of r is chosen in such a way that the work spent in a m-point node is $o(m^{1-1/\lfloor d/2 \rfloor})$, and so the running time will satisfy

$$T(m) \leq o(m^{1-1/\lfloor d/2 \rfloor}) + O(r^{1-1/\lfloor d/2 \rfloor})T(2m/r),$$

with $r = m/(\log m)^c$, which gives $T(n) = O(n^{1-1/\lfloor d/2 \rfloor} 2^{O(\log^* n)})$.

The above described data structure needs simplicial partitions with r quite close to the number of points ($r = m/(\log m)^c$), which we are no longer able to compute in $O(n \log n)$ time. We can obviously compute them in a polynomial time. To reduce this to $O(n^{1+\delta})$ time, we may refine the simplicial partition a constant number of times; this is discussed in [Mat91] in more detail. □

5 Linear-sized ε-nets in dimension 3

One of interesting open problems in combinatorial geometry concerns ε-nets with respect to halfspaces. Let $X \subseteq \mathrm{E}^d$ be a finite point set. Let ε be a real number, $0 \leq \varepsilon < 1$. A subset $N \subseteq X$ is called an *ε-net for X with respect to halfspaces,* if every halfspace containing more than $\varepsilon|X|$ points of X also contains a point of N. The notion of ε-net can actually be defined in a more general setting, and from general results of Haussler and Welzl [HW87] one gets that for every fixed d and for every ε and X, there exists an ε-net for X with respect to halfspaces, of size $O((1/\varepsilon)\log(1/\varepsilon))$. The problem is whether this bound can be sharpened, e.g. if ε-nets of size $O(1/\varepsilon)$ with respect to halfspaces always exist.

A partial result was given by Seidel et al. [SWM90], who proved (by a quite complicated geometric argument) that for every finite point set $X \subseteq \mathrm{E}^3$ and every $\varepsilon \in (0,1)$, there exists an ε-net for X with respect to halfspaces, of size $O(1/\varepsilon)$.

Later K. Clarkson gave an alternative probabilistic proof of this result in an unpublished note [Cla89]. By an easy application of the Shallow cutting lemma, we get a proof quite similar to Clarkson's.

It will be more convenient for us to consider a dual version of this problem, where we have a collection H of hyperplanes and we want to select a subset $N \subseteq H$ such that any point of level $> n/r$ is separated from the reference point o by some hyperplane of N.

Let us apply Shallow cutting lemma with $d = 3$, $k = 2n/r$ and $2r$ instead of r, choosing a $(1/2r)$-cutting Ξ of size $O(r)$ for the ($\leq 2n/r$)-level of H. Now for every simplex $s \in \Xi$ which is separated from the reference point o by at least one hyperplane $h \in H$ (i.e. s and o lie in opposite halfspaces determined by h), let us choose one such separating hyperplane h_s. Let all such hyperplanes h_s form a set N with $O(r)$ elements. We claim that N has the desired property. Indeed, it suffices to show that for every point of level greater than n/r but smaller than $2n/r$ there is a hyperplane of N separating it from the reference point o. Let us consider one such point x and let s be the simplex of Ξ containing it. There are at least n/r hyperplanes separating s from o, and at most $n/2r$ of them may cross the simplex s; thus the remaining ones of them separate s from o. Therefore there is a hyperplane separating s (and also x) from o in N. □

References

[AM91] P. K. Agarwal and J. Matoušek. Ray shooting and parametric search. Manuscript, 1991.

[AHL90] A. Aggarwal, M. Hansen, and T.Leighton. Solving query-retrieval problems by compacting Voronoi diagrams. In *Proc. 21. ACM Symposium on Theory of Computing*, pages 331–340, 1990.

[Cha86] B. Chazelle. Filtering search: A new approach to query-answering. *SIAM Journal on Computing*, 15(3):703–724, 1986.

[Cha89] B. Chazelle. Lower bounds on the complexity of polytope range searching. *J. Amer. Math. Soc*, 2(4):637–666, 1989.

[CGL85] B. Chazelle, L. Guibas, and D. T. Lee. The power of geometric duality. *BIT*, 25(1), 1985.

[CEG89] B. Chazelle, H. Edelsbrunner, and L. Guibas. Lines in space: Combinatorics, algorithms and applications. In *Proc. 21. ACM Symposium on Theory of Computing*, pages 389–392, 1989.

[CF90] B. Chazelle and J. Friedman. A deterministic view of random sampling and its use in geometry. *Combinatorica*, 10(3):229–249, 1990.

[CP86] B. Chazelle and F. P. Preparata. Halfspace range searching: An algorithmic application of k-sets. *Discrete & Computational Geometry*, 1:3–93, 1986.

[CSW90] B. Chazelle, M. Sharir, and E. Welzl. Quasi-optimal upper bounds for simplex range searching and new zone theorems. In *Proc. 6. ACM Symposium on Computational Geometry*, pages 23–33, 1990.

[CW89] B. Chazelle and E. Welzl. Quasi-optimal range searching in spaces of finite VC-dimension. *Discrete & Computational Geometry*, 4:467–490, 1989.

[Cla88] K. L. Clarkson. Applications of random sampling in computational geometry II. In *Proc. 4. ACM Symposium on Computational Geometry*, pages 1–11, 1988.

[Cla89] K. Clarkson, unpublished proof outline, December 1989.

[Ede87] H. Edelsbrunner. *Algorithms in combinatorial geometry.* Springer, 1987.

[HW87] D. Haussler and E. Welzl. ε-nets and simplex range queries. *Discrete & Computational Geometry*, 2:127–151, 1987.

[Mat90] J. Matoušek. Cutting hyperplane arrangements. In *Proc. 6. ACM Symposium on Computational Geometry*, pages 1–9, 1990. Full version: Discr. & Comput. Geometry 1991, to appear.

[Mat91] J. Matoušek. Efficient partition trees. In *Proc. 7. ACM Symposium on Computational Geometry*, 1991.

[SWM90] R. Seidel, E. Welzl, and J. Matoušek. Netting a lot with a little: Small ε-nets for disks and halfspaces. In *Proc. 6. ACM Symposium on Computational Geometry*, pages 16–22, 1990.

[Wil82] D. E. Willard. Polygon retrieval. *SIAM Journal on Computing*, 11:149–165, 1982.

[YY85] F. F. Yao and A. C. Yao. A general approach to geometric queries. In *Proc. 17. ACM Symposium on Theory of Computing*, pages 163–168, 1985.