Original Lecture #13: Tuesday, May 24, 1994
Topics: Range Searching

# Efficient Partition Trees for Range Searching

We have seen the use of partition trees for range searching earlier. In this lecture an efficient partition scheme which achieves the optimal query time within polylog factors is described. See [4] and [5] for more details. There are several new ideas on which this scheme relies. The existence of "good" *cuttings* are central to the algorithm. In the next section some of the definitions and results on cuttings are reviewed.

## Cuttings

**Definition 1.** *A* cutting *is a collection of $d$-dimensional closed simplices with disjoint interiors whose union is the entire space $E^d$.*

Let $H$ be a collection of $n$ hyperplanes in $E^d$. Let $\Xi$ be a cutting. For a simplex $\Delta \in \Xi$ denote by $H_\Delta$ the subset of $H$ which cut the interior of $\Delta$. See figure 1.

**Definition 2.** *A cutting $\Xi$ is an $\varepsilon$-cutting for $H$ if $|H_\Delta| \leqslant \varepsilon n$ for every $\Delta \in \Xi$.*

The notion of a cutting can be extended to *weighted* collection of hyperplanes. A weighted collection of hyperplanes is a pair $(H, w)$, where $H$ is a collection of hyperplanes and $w : H \to R^+$ is a wight function on $H$. The notation $w(X)$ stands for $\sum_{h \in X} w(h)$.

**Definition 3.** *A cutting $\Xi$ is an $\varepsilon$-cutting for $(H, w)$ if $w(H_\Delta) \leqslant \varepsilon w(H)$ for every $\Delta \in \Xi$.*

The main results on cuttings are summarized in the following theorem.

**Theorem 4.**

1. *For any collection of $n$ hyperplanes and a parameter $r \leqslant n$, there exists a $(1/r)$-cutting of size $O(r^d)$ (which is asymptotically the best possible size) [2].*

2. *A cutting as in 1 can be computed in $O(nr^{d-1})$ time by a randomized algorithm [2] or by a deterministic algorithm [1].*

3. *Any algorithm computing $(1/r)$-cutting for unweighted collecttions of hyperplanes can be converted into one computing $(1/r)$-cutting for weighted collections, with the same asymptotic bounds on the running time and on the size of the resulting cutting [3].*

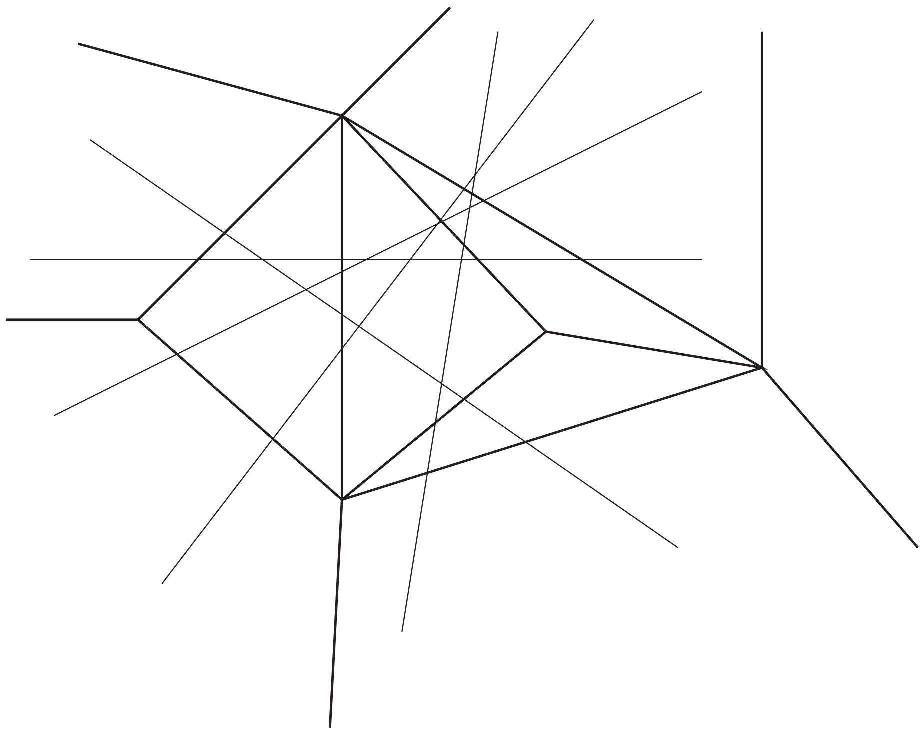

Figure 1: Set of lines (light lines) and a cutting (thick lines)

We saw earlier a randomized algorithm for cuttings in the plane. Pick a random sample $r$ out of the $n$ lines and do a vertical decomposition of their arrangement. We proved that the expected number of lines crossing any trapezoid is less than $4n/r$. The number of trapezoids is $O(r^2)$. There are techniques to make the result worst case for every trapezoid both in a randomized and deterministic sense. Also, we need simplices and not trapezoids. There is a so called *canonical triangulation* that can be applied once we have the vertical decomposition. See [5] for more details. A simple proof of why a $(1/r)$-cutting needs $\Omega(r^d)$ simplices in the worst case, is given below. Let $\mathcal{A}(H)$ denote the arrangement of the $n$ hyperplanes. There are $O(n^d)$ vertices in this arrangement. It is clear that, if a vertex of an arrangement is in the interior of a simplex $\Delta$ all $d$ hyperplanes incident on that vertex cut $\Delta$. Since no simplex is cut my more than $O(n/r)$ hyperplanes, the number of vertices of the arrangement in the interior of any simplex $\Delta$ is upper bounded by $O(n^d/r^d)$. Since the simplices subdivide the plane there are at least $\Omega(n^d/(n^d/r^d)) = \Omega(r^d)$ simplices in the cutting.

## Simplicial Partitions

The *ham-sandwich* cut theorem is a way to partition point sets into equal sized parts. The partition schemes which depend on *subdividing* the space are too restrictive. Do we really need to subdivide the space itself to partition the point set? We need two properties for a partition, to have good query time. The first property is that the points should be evenly divided among

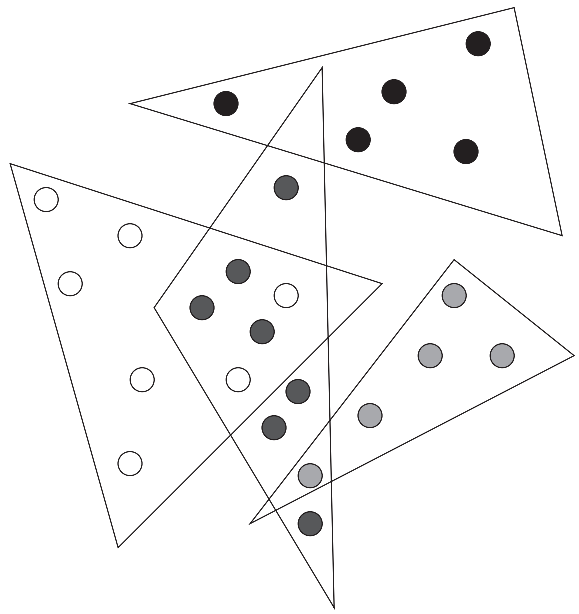

Figure 2: A Simplicial Partition

the parts for efficient divide and conquer. The second is that no hyperplane should intersect more than a fraction of the total number of parts. We define a partition which does not have the restriction that its parts should divide space.

**Definition 5.** *A* simplicial partition *for a point set $P$ is a collection of pairs*

$$\Pi = \{(P_1, \Delta_1), (P_2, \Delta_2), \ldots, (P_m, \Delta_m)\} \tag{1}$$

*where $P_i$'s form a partition of $P$ and $\Delta_i$ is a simplex containing the set $P_i$.*

Note that the simplices can *overlap*, and that a point can be in the interior of more than one simplex, but each point *belongs* to exactly one simplex. See figure 2. As remarked earlier tha simplices need not subdivide the space. We say that the simplices *cover* the point set. In addition we would like our partition to satisfy the two properties mentioned above. The first property is enforced by requiring our partition to satisfy

$$\max\{|P_i|\} < 2\min\{|P_i|\}. \tag{2}$$

We would like to make the notion of a hyperplane not cutting too many simplices more precise.

**Definition 6.** *The* crossing number *of a hyperplane $h$ (relative to $\Pi$), denoted by $\kappa_\pi(h)$, is defined as the number of simplices among the $\Delta_i$'s cut by $h$. The crossing number of the partition $\Pi$ itself, is defined to be the maximum of crossing numbers over all hyperplanes $h$, and is denoted by $\kappa_\pi$.*

For efficient partition trees we would like our simplicial partition to have as small a crossing number as possible. The main result which leads to efficient partition trees is the following.

**Theorem 7. Partition Theorem :** *Let $P$ be a point set in $E^d (d \geqslant 2)$. Let $r$ be an integer parameter such that $2 \leqslant r < n$. There exists a simplicial partition $\Pi$ for $P$ such that $n/r \leqslant |P_i| < 2n/r$ for $1 \leqslant i \leqslant m$ and whose crossing number is $O(r^{1-1/d})$. This bound is asymptotically tight in the worst case.*

Before we prove the above theorem, and show how to construct the partition, let's see what kind of bounds we can get for query and space using partition trees based on such partitions. Denote by $Q(n)$ and $S(n)$ the query and space bounds for the *counting* version. For a query *half-space* defined by a hyperplane $h$, we find all the parts of $\Pi$ cut by $h$. Since there are $O(r)$ simplices, checking for intersection with the query hyperplane $h$ takes $O(r)$ time (since $d$ is fixed, intersection of a hyperplane with a simplex is assumed to be O(1) ). All the parts which are not intersected lie entirely within one of the two half-space's defined by $h$, handling them is trivial. From the bound on the crossing number, $h$ intersects only $O(r^{1-1/d})$ of the simplices. We have to recurse on each of them. Since the number of points in each simplex is bounded by $2n/r$, the following recurrence holds for query time.

$$Q(n) = O(r) + O(r^{1-1/d})Q(\frac{2n}{r}) \tag{3}$$

If $r$ is chosen to be equal to $n^{1-1/d}$ the depth of the recursion tree will be $O(\log\log n)$ and the solution to the above recurrence is

$$Q(n) = O(n^{1-1/d} 2^{O(\log\log n)}) = O(n^{1-1/d} \log n) \tag{4}$$

The space requirements satisfy the following recurrence (this is true only for the *counting* version of the problem). We need $O(r)$ space to store the number of points in each part, and to store the partition itself.

$$S(n) = O(r) + O(r)S(\frac{2n}{r}) \tag{5}$$

It can be verified that for $r = O(n^{1-1/d})$ that $S(n) = O(n)$.

The *reporting* case can be handled with the above data structure as follows. With each *leaf* of the data structure, the set of points associated with that leaf are stored. This does not increase the space complexity. To answer queries, when a class is to be reported in its entirety, we just go down the sub-tree rooted at that class and report all the points stored in the leaves. As we saw before, the size of the tree is linear in the number of points (from the space recurrence). Therefore reporting takes time proportional to the number of points reported, in addition to the traversal cost. Thus the reporting version can be solved with the same data structure, in time $O(n^{1-1/d} \log(n) + k)$, where $k$ is the number of points reported.

Though the above description used *half-spaces*, the same arguments can be applied to *simplices*. Since a simplex has complexity depending only on $d$ (the dimension), it is bound by $O(1)$ halfspaces each of which intersects atmost $O(r^{1-1/d})$ simplices of $\Pi$. Also intersection between two simplices takes $O(1)$ time. We conclude that same time and space bounds hold for simplex range searching. These results can be summarized in the following theorem.

**Theorem 8.** *Let $P$ be a set on $n$ points in $E^d$. There exists a partition tree of size $O(n)$ that can be used for answering any simplex range query (half-space is a special case) in $O(n^{1-1/d}\log n)$ time. In report mode , the query time is $O(n^{1-1/d}\log n + k)$, where $k$ is the size of the answer.*

It is to be noted that the above results are not the best known for *half- space* range *reporting*, though they are within polylog factors of the lower bounds due to Chazelle, for the *counting* case. With slightly more than linear space, one can achieve better report time for half-spaces. An ingeneous technique called *filtering search* due to Chazelle is used to speed up the report time. The following theorem is stated without proof.

**Theorem 9.** *Let $P$ be a set on $n$ points in $E^d$. There exists a partition tree of size $O(n\log\log n)$ that can be used for answering any half-space reporting queries in $O(n^{1-1/\lfloor d/2\rfloor}\log n + k)$ time.*

It remains to prove the Partition theorem. The proof is constructive, and the algorithm to construct the partition falls out of it. The proof relies on two ideas. The first idea shows that, in order for a partition Π, to have a low crossing number, it is enough to show that it has a low crossing number with respect to a small set of *test* hyperplanes. The proof of this also gives us a way to find such a set. The second idea is to use weighted cuttings to find a partition, which has a low crossing number with respect to a *given* set of hyperplanes. Combining these two ideas, we can construct a partition which satisfies the conditions of the Partition theorem.

## Test Set of hyperplanes

Let $P$ be a point set in $E^d$. Suppose we are given a simplicial partition Π for $P$, and an integer parameter $r$. How fast can we *verify* that it has a small crossing number *i. e.* the crossing number is $O(r^{1-1/d})$? It seems impossible at first thought, since the number of different hyperplanes is infinite. Can we find a small set of *test* hyperplanes such that, it is enough to verify Π against those, instead of every hyperplane? The answer turns out to be positive. The real need of the test hyperplanes is not to *verify* but to *construct* good partitions. If we have a small set, we can efficiently build a partition which has a small crossing number with respect to that set, and this will ensure that the partition is good for all hyperplanes. The existence of a small *test set* of hyperplanes is shown by the following lemma. Define for a set of hyperplanes $Q$, $\kappa_\pi(Q)$ as $\max_{h\in Q}\{\kappa_\pi(h)\}$.

**Lemma 10. Test Set lemma :** *For an $n$-point set $P$ in $E^d$ and a parameter $r$ there exists a set $T$ of at most $r$ hyperplanes, such that, for any simplicial partition Π for $P$, satisfying $|P_i| \geqslant n/r$ for every $i$, the following holds*

$$\kappa_\pi \leqslant (d+1)\kappa_\pi(T) + O(r^{1-1/d}) \tag{6}$$

*Proof.* Let $\mathcal{D}(X)$ denote the *dual* objects of a collection of objects $X$ in the primal space. Let $H = \mathcal{D}(P)$ be the set of hyperplanes dual to the point set $P$. From the results on cuttings, we can choose a $(1/t)$-cutting Ξ of $H$ where $t = r^{1/d}$. This will ensure that Ξ will have $t^d = O(r)$ vertices. Let $V$ be the set of all vertices of the simplices of Ξ. Set $T = \mathcal{D}(V)$, *i. e.* $T$ is the set

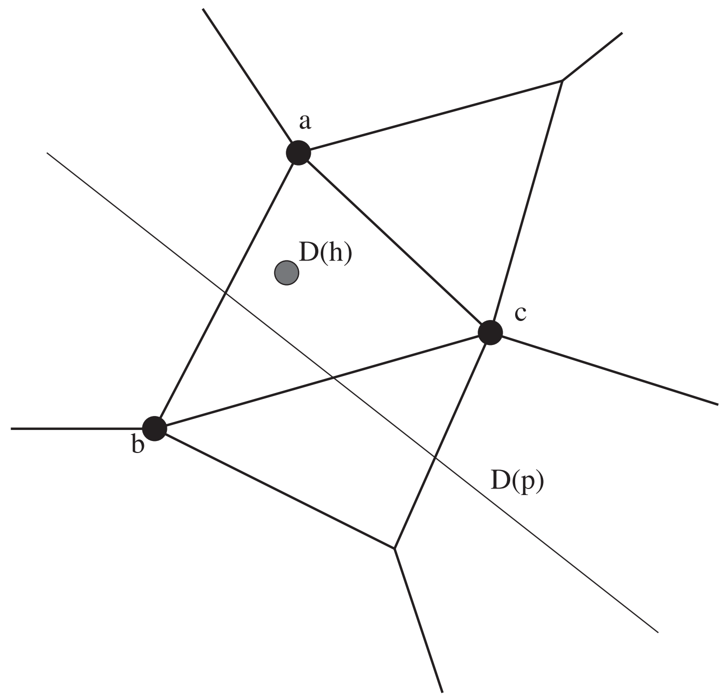


Figure 3: The Dual. $a, b, c$ are the vertices of $\Delta_h$

of hyperplanes dual to the vertices $V$. Since $V$ is in the dual, $T$ is in the primal and has $O(r)$ hyperplanes. The claim is that $T$ as constructed, has the desired property. See figures 3 and 4.

Let $h$ be any hyperplane, and $\Delta_h$ be the simplex of $\Xi$ (in the dual) containing the point $\mathcal{D}(h)$. Let $G$ be the vertices of $\Delta_h$, and $\mathcal{D}(G)$ be the set of hyperplanes in the primal, dual to vertices in $G$. Note that $\mathcal{D}(G)$ is a subset of $T$, and since $G$ is a simplex, it's cardinality is bounded by $d+1$. Consider the simplices of $\Pi$ cut by $h$. We classify them into two types, and count the number of simplices of each type separately. The first class contains simplices which are cut by $h$, and *some* hyperplane in $\mathcal{D}(G)$. How many such simplices are there? Since each hyperplane in $\mathcal{D}(G)$ cuts atmost $\kappa_\pi(T)$ simplices of $\Pi$ (by definition of $\kappa_\pi(T)$) and there are atmost $d+1$ of those, their number is bounded by $(d+1)\kappa_\pi(T)$. The second class of simplices are those cut by $h$, but not by any of the hyperplanes in $\mathcal{D}(G)$. Suppose $\Delta_i$ is such a simplex. It should be that, $\Delta_i$ is contained *entirely* within in the *zone* of $h$, in the arrangement of $\mathcal{D}(G)$ (see figure 4), for otherwise it could not have intersected $h$. Consider any point $p \in P_i$, where $(P_i, \Delta_i) \in \Pi$. Since $\Delta_i$ is in the zone of $h$, $p$ is also within the zone of $h$ in the arrangement of $\mathcal{D}(G)$. From duality properties, the hyperplane $\mathcal{D}(p)$, cuts the simplex $\Delta_h$, in the dual. Since $\Delta_h$ is a simplex of a $(1/t)$-cutting $\Xi$, there are atmost $O(n/t)$ hyperplanes cutting it. This implies that number of points $p$ contained in simplices of the second type, is $O(n/t)$. Since each simplex of $\Pi$ has $\Omega(n/r)$ points, the number of simplices of the second type is $O(r/t) = O(r^{1-1/d})$. Since every simplex cut by $h$ has to be one of the two types, adding up the bounds on the number of simplices of each type gives us the desired result.

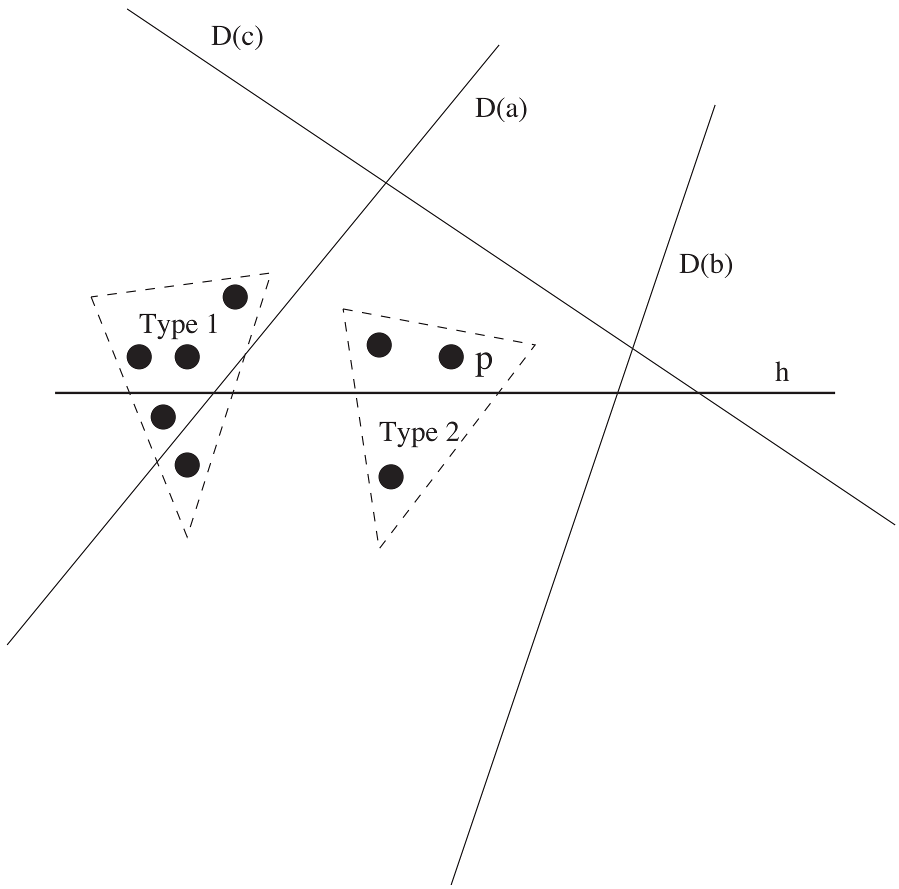


Figure 4: The Primal. Hyperplane $h$ in the zone of $\mathcal{D}(a), \mathcal{D}(b), \mathcal{D}(c)$

## Finding a Good Partition

From the test set lemma, we know that, if $\kappa_\Pi(T)$ is $O(r^{1-1/d})$ for our partition $\Pi$, $\kappa_\pi$ is $O(r^{1-1/d})$. The goal of this section is to find partitions, which have good crossing number with respect to *any given* set of hyperplanes $Q$, and in particular our test set $T$. This will finish the proof of the Partition theorem.

How can we find a partition which has a low crossing number for a given set of hyperplanes? The idea is to build simplices incrementally, each containing roughly $O(n/r)$ points. As we add simplices, we have to ensure that no hyperplane $h \in Q$ cuts too many simplices. Suppose that we have already constructed a few simplices, and that a hyperplane $h$ has cut many of the simplices. When we are adding a new simplex, we have to somehow make the probability, that $h$ cuts our new simplex, small. This will ensure that, at the end of the construction, we will have no hyperplane cutting too many simplices. We have a choice over what sort of function we use, to bound the probabilty. Exponential functions are usually chosen, because they give us good bounds on the final values, and also because they give us a lot of flexibility. The way to do this deterministically, is to use the cutting results for *weighted* hyperplanes.

The following lemma and it's proof give us the algorithm to construct partitions with low crossing number.

**Lemma 11.** *Let $P$ be a $n$-point set and $r$ an integer parameter and let $Q$ be a set of hyperplanes. Then, we can construct a simplicial partition $\Pi$ for $P$, whose classes satisfy $n/r \leqslant |P_i| \leqslant 2n/r$ for every $i$, and such that $\kappa_\pi(Q) = O(r^{1-1/d} + \log|Q|)$.*

*Proof.* The construction is incremental. Suppose we have already constructed the disjoint sets $P_1, P_2, ..., P_i$ and their corresponding simplices $\Delta_1, \Delta_2, ..., \Delta_i$. Let $P_i' = P - \cup_{k=1}^{i} P_k$ be the set of remaining points, and $n_i = |P_i'|$. If $n_i < 2n/r$ we make $P_{i+1} = P_i'$ and $\Delta_{i+1} = E^d$ and we are done. Otherwise, choose $r_i$ such that $n/r \leqslant n_i/r_i < 2n/r$. Let $t_i = 1/r_i^{1/d}$.

Define a weight function $w_i : Q \to R^+$ as follows. For every hyperplane $h \in Q$, let $h(i)$ be the number of simplices in $\Delta_1, \Delta_2, .., \Delta_i$ crossed by $h$. Set $w_i(h) = 2^{h(i)}$. From the results on cuttings mentioned earlier, we can construct a $(1/t_i)$ cutting $\Xi$ of $(Q, w_i)$. This cutting $\Xi$ will have $O(r_i)$ simplices. By the pigeon hole principle, some simplex of the cutting $\Xi$, has $\geqslant n_i/r_i$ points of $P_i'$ (note that a cutting is a subdivision of space). Let $\Delta_{i+1}$ be one such simplex. We choose some $n/r$ points of $P_i'$ contained in $\Delta_{i+1}$, to form $P_{i+1}$. This finishes the description of the construction. The main idea in the construction is *reweighing* the hyperplanes. By making the weight of a hyperplane *exponential* in the number of simplices it has already cut, we ensure that it does not cut too many of them during the incremental construction.

Now we establish bounds on the crossing number $\kappa_\pi(Q)$. The crossing number $\kappa_\pi(h)$ of a hyperplane $h \in Q$, is related to the weight of $h$ at the end of the construction, by the following equation.

$$w_m(h) = 2^{\kappa_\pi(h)} \tag{7}$$

The second observation is concerning the growth of the weight function $w_i$. We note that the $w_{i+1}(h) = 2w_i(h)$ iff $h$ cuts $\Delta_{i+1}$, otherwise $w_{i+1}(h) = w_i(h)$. Denote by $Q_{i+1}$, the set of hyperplanes which intersect $\Delta_{i+1}$. Summing the increase in weight of the hyperplanes in $Q_{i+1}$ gives

$$w_{i+1}(Q) - w_i(Q) = \sum_{h \in Q_{i+1}} w_i(h) \tag{8}$$

We can bound the right hand side of the above equation, by noting that $\Delta_{i+1}$ is a simplex of a $(1/t_i)$-cutting of $Q$. Therefore the sum on the right hand side is bounded by the cutting result by $w_i(Q)/t_i$. Rewriting the above

$$w_x i + 1(Q) \leqslant w_i(Q)\left(1 + \frac{1}{t_i}\right) \tag{9}$$

From above equation we can bound the weight of the set $Q$ at the end of the construction by unwinding the recurrence,

$$w_m(Q) \leqslant w_0(Q) \prod_{i=0}^{m-1} \left(1 + \frac{1}{t_i}\right) \tag{10}$$

We also have $w_0(Q) = |Q|$ and $t_i = 1/r_i^{1/d}$ and $r_i = c(r-i)$ for some constant $c$. Taking logarithms on both sides and using the inequality $\ln(1+x) \leqslant x$ we get

$$\log w_m(Q) \leqslant \log Q + \frac{1}{c}\sum_{i=0}^{m-1} \frac{1}{(r-i)^{1/d}} \tag{11}$$

From construction, $m = \lfloor r \rfloor$.We can bound the harmonic sum on the right hand side by an integral, to get

$$\log w_m(Q) = O(\log Q + r^{1-1/d}) \tag{12}$$

Since $w_m(h) \leqslant w_m(Q)$ and $\kappa_\pi(h) = \log(w_m(h))$ we have the desired result.

## Conclusion

From the previous two sections the proof of the Partition theorem is very simple. Using the test set lemma, for a given parameter $r$, we construct a test set of hyperplanes $T$, of size $r$. Then using the previous lemma, we can construct a partition $\Pi$, whose simplices have $\Theta(n/r)$ points each, and such that no hyperplane $h \in T$, cuts more than $O(r^{1-1/d})$ of it's simplices. From the test set lemma, it follows that no hyperplane cuts more than $(d+1)O(r^{1-1/d}) + O(r^{1-1/d}) = O(r^{1-1/d})$ simplices of $\Pi$.

# References

[1] B. Chazelle. Cutting hyperplanes for divide-and-conquer. *Discrete Comput. Geom.*, 9(2):145–158, 1993.

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

[3] J. Matoušek. Cutting hyperplane arrangements. *Discrete Comput. Geom.*, 6:385–406, 1991.

[4] J. Matoušek. Efficient partition trees. *Discrete Comput. Geom.*, 8:315–334, 1992.

[5] K. Mulmuley. *Computational Geometry: An Introduction Through Randomized Algorithms*. Prentice Hall, New York, 1993.