SIAM J. COMPUT.
Vol. 38, No. 6, pp. 2411–2425

RANDOM HYPERPLANE SEARCH TREES*

LUC DEVROYE†, JAMES KING†, AND COLIN MCDIARMID‡

Abstract. A hyperplane search tree is a binary tree used to store a set S of n d-dimensional data points. In a random hyperplane search tree for S, the root represents a hyperplane defined by d data points drawn uniformly at random from S. The remaining data points are split by the hyperplane, and the definition is used recursively on each subset. We assume that the data are points in general position in $\mathbb{R}^d$. We show that, uniformly over all such data sets S, the expected height of the hyperplane tree is not worse than that of the k-d tree or the ordinary one-dimensional random binary search tree, and that, for any fixed $d \geq 3$, the expected height improves over that of the standard random binary search tree by an asymptotic factor strictly greater than one.

Key words. binary search tree, data structures, expected time analysis, height of a tree, hyperplane tree, random tree, large deviation theory, random sampling

AMS subject classifications. 68P05, 68W20, 68W40, 60E15

DOI. 10.1137/060678609

1. Introduction and results.

Hyperplane search trees. A hyperplane search tree is defined as follows. Given is a set $S = \{x_1, \ldots, x_n\}$ of points in general position[1] in $\mathbb{R}^d$. The root node is formed by $X_1, \ldots, X_d$, obtained by uniform random sampling without replacement from $x_1, \ldots, x_n$. The hyperplane through these points is denoted by $H = H(X_1, \ldots, X_d)$. It partitions $\mathbb{R}^d \backslash H$ into two sets H^+ and H^-, with some rule to choose which is which. The $n - d$ remaining data points are split according to membership in H^+ and H^-. The subtrees are defined recursively from there and are randomly labeled as the left and right subtrees of the root. A set of cardinality less than d is not split: it occupies a leaf in the tree. Leaves correspond, thus, to collections of cardinality between 0 and $d - 1$. For $d = 1$, therefore, all points in S lie in internal nodes and all leaves are empty. Figure 1 shows a hyperplane tree in $\mathbb{R}^2$ and the partition of the plane into disjoint polygons defined by it.

For a given set S of points in $\mathbb{R}^d$, $d \geq 2$, let $T(S)$ denote the random hyperplane search tree based on S. For $d = 1$, the hyperplane tree depends only on $|S|$ and not on the elements of S. Thus, it makes sense to drop the set and simply write $T_{|S|}$ or T_n. With this definition, the structure[2] of the usual random binary search tree on n distinct random keys is the same as that of T_n.

Results. For random variables X and Y, we say that Y stochastically dominates X, or $X \leq^s Y$, if $\mathbf{P}\{X \geq t\} \leq \mathbf{P}\{Y \geq t\}$ for any value of t. If this is the case, there

*Received by the editors December 27, 2006; accepted for publication (in revised form) November 25, 2008; published electronically March 27, 2009. This research was supported by NSERC grant A3456.
http://www.siam.org/journals/sicomp/38-6/67860.html
†School of Computer Science, McGill University, Montreal, H3A 2A7, Canada (luc@cs.mgcill.ca, jking@cs.mcgill.ca).
‡Department of Statistics, University of Oxford, 1 South Parks Road, Oxford, OX1 3TG, United Kingdom (cmcd@stats.ox.ac.uk).
[1] We say a set of points is in general position in a d-dimensional vector space W if and only if no $k + 2$ points are contained in a k-dimensional affine subspace (or flat) of W for any $k < d$. The vector space will typically be $\mathbb{R}^d$.
[2] In the random binary search tree, unlike the hyperplane search tree, children of a node are deterministically and meaningfully labeled as left or right children. However, this distinction is irrelevant to our analysis of height and depth.

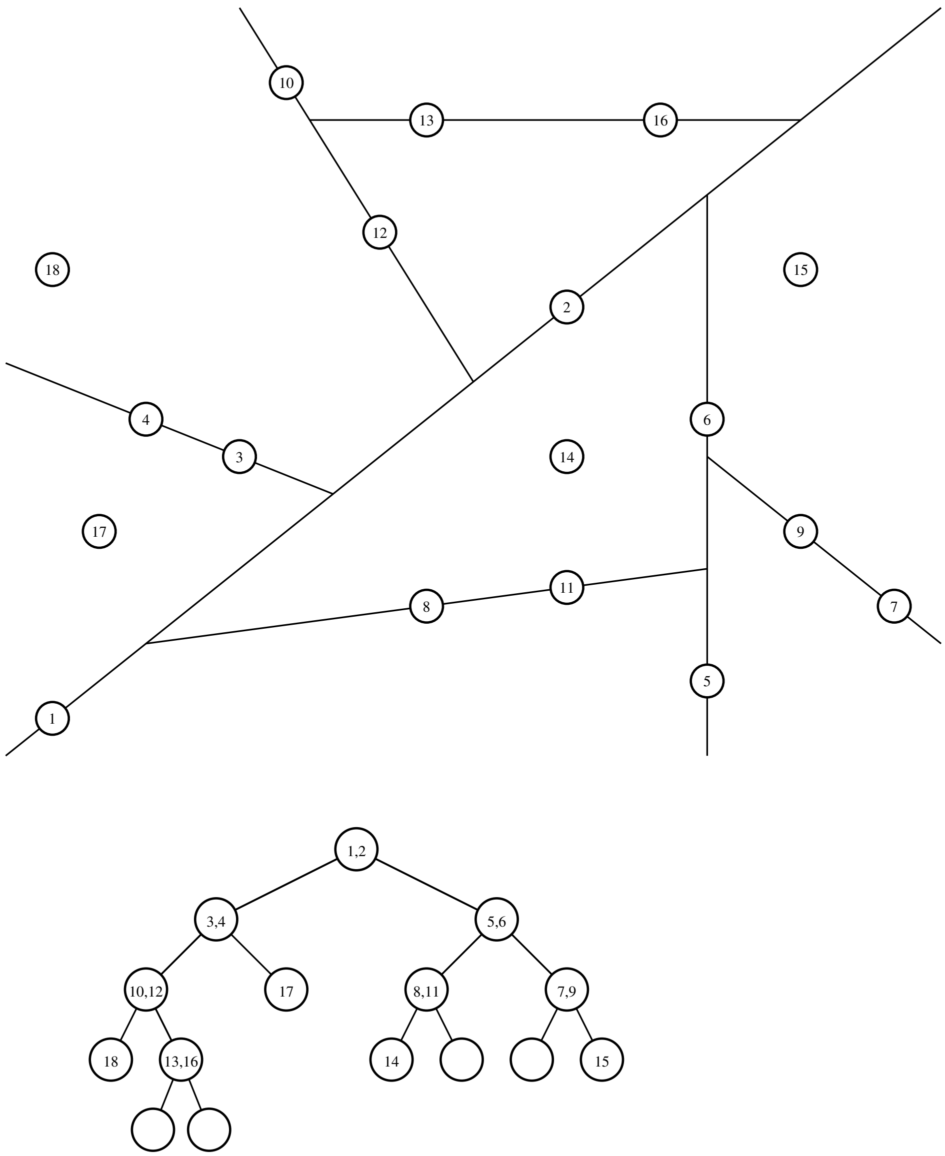


FIG. 1. *The figure shows the partition of the plane induced by the hyperplane tree. Internal nodes of the tree house two data points each.*

exists a coupling of X and Y such that X is never greater than Y.

Our first result compares the depths of points in $T(S)$ and T_n, recalling that S is a set of points in general position in $\mathbb{R}^d$, $d \geq 2$. Let $D(S)$ be the depth of a point (not a tree node) in $T(S)$ chosen uniformly at random, and let D_n be the depth of a point in T_n chosen uniformly at random. Define $\mathcal{S}_{n,d} = \{S : S \subseteq \mathbb{R}^d, |S| = n, S \text{ is in general position}\}$.

THEOREM 1. *For each set $S \in \mathcal{S}_{n,d}$, we have $D(S) \leq^s D_n$.*

We prove this theorem in section 4. There are similar results for other measures to compare the trees $T(S)$ and T_n that follow trivially from this result. For one example, we consider the internal path length (IPL) of a tree, defined as the sum of the depths of all points (not nodes) stored in the tree.

COROLLARY 1. $\sup_{S\in\mathcal{S}_{n,d}} \mathbf{E}\{\mathrm{IPL}(T(S))\} \le \mathbf{E}\{\mathrm{IPL}(T_n)\}$.

We also have an ordering on the moments.

COROLLARY 2. *For all* $r > 0$,

$$\sup_{S\in\mathcal{S}_{n,d}} \mathbf{E}\left\{(D(S))^r\right\} \le \mathbf{E}\left\{(D_n)^r\right\}$$

and

$$\sup_{S\in\mathcal{S}_{n,d}} \mathbf{E}\left\{e^{rD(S)}\right\} \le \mathbf{E}\left\{e^{rD_n}\right\}.$$

The rest of our results concern the height of hyperplane search trees. Let $H(S)$ and H_n be the respective heights of $T(S)$ and T_n. For $c \ge 2$, let $\eta(c) = 1 - c\ln\frac{2e}{c}$. Let $\alpha = 4.31107\ldots$ be the unique solution at least 2 to $\eta(c) = 0$. It is known (Robson [31], Devroye [6]) that $H_n/\ln n \to \alpha$ in probability as $n \to \infty$.

COROLLARY 3. *Let* $c > \alpha$. *Then* $\eta(c) > 0$ *and for each* $d \ge 2$ *and* $n \ge d$

$$\sup_{S\in\mathcal{S}_{n,d}} \mathbf{P}\{H(S) \ge c\ln n\} \le n^{-\eta(c)}.$$

Proof. By a union bound, for any $S \in \mathcal{S}_{n,d}$, $\mathbf{P}\{H(S) \ge t\} \le n\mathbf{P}\{D(S) \ge t\} \le n\mathbf{P}\{D_n^* \ge t\}$, where D_n^* is the depth of the last node inserted in a random binary search tree on n nodes. By Devroye [8], D_n^* is distributed as $\sum_{i=2}^n B_i$, where B_i is Bernoulli$(2/i)$ and the B_i are independent. Thus, for $s > 0$, by the Chernoff bound,

$$\begin{aligned}\mathbf{P}\{D_n^* \ge t\} &\le e^{-st}\prod_{i=2}^n \mathbf{E}\left\{e^{sB_i}\right\}\\ &= e^{-st}\prod_{i=2}^n (1 + 2(e^s - 1)/i)\\ &\le e^{-st}\prod_{i=2}^n \exp(2(e^s-1)/i)\\ &\le \exp(2(e^s-1)\ln n - st),\end{aligned}$$

and, upon taking s such that $2e^s \ln n = t$, we get the bound

$$\mathbf{P}(H(S) \ge t) \le n \cdot \exp\left(t - 2\ln n - t\ln\left(\frac{t}{2\ln n}\right)\right).$$

Put $t = c\ln n$ for constant c. The upper bound then becomes

$$\exp\left(\ln n\left(c - c\ln\left(\frac{c}{2}\right) - 1\right)\right) = n^{c - c\ln\left(\frac{c}{2}\right) - 1} = n^{-\eta(c)}.$$

Finally, we note that $\eta(\alpha) = 0$ and $\eta'(x) = \ln\frac{x}{2} > 0$ for $x > 2$, so $\eta'(c) > 0$. ◻

With the goal of proving tighter bounds for hyperplane search trees, let us consider first the plane $\mathbb{R}^2$. Let us say that a set of points in $\mathbb{R}^2$ is in *exposed position* if each point in the set is an extreme point of the convex hull. Let S_n be a set of n points in exposed position in $\mathbb{R}^2$ for $n = 2, 3, \ldots$. Then it is not hard to show that, as $n \to \infty$, if our hyperplane tree is based on S_n,

$$H(S_n)/\ln n \to \alpha \text{ in probability.}$$

Thus, in the plane we see that hyperplane search trees need not be better than random binary search trees. In many cases, however, hyperplane search trees are strictly better than random binary search trees. The following theorem deals with point sets in general position in $\mathbb{R}^d$, $d \geq 3$.

THEOREM 2. *There exist constants $c < \alpha$ and $\epsilon > 0$ such that, for each $d \geq 3$ and each $n \geq d$,*

$$\sup_{S \in \mathcal{S}_{n,d}} \mathbf{P}\{H(S) \geq c \ln n\} \leq n^{-\epsilon}.$$

As mentioned above, hyperplane search trees built on point sets in $\mathbb{R}^2$ need not be better than random binary search trees. However, perhaps the most interesting case in the plane is when the points in the set S are sampled uniformly from some convex body K. In this case, hyperplane search trees are strictly better than random binary search trees.

THEOREM 3. *There exist constants $c < \alpha$ and $\epsilon > 0$ such that the following holds. Let $n \geq 3$, and let S be a set of n points sampled uniformly at random from a convex body K in $\mathbb{R}^2$. Then, with $H(S)$ being the height of a hyperplane search tree built on S,*

$$\mathbf{P}\{H(S) \geq c \ln n\} \leq n^{-\epsilon}.$$

The next section contains a brief list of relevant references on multidimensional search trees. In section 3 we analyze the distribution that generates the sizes of the root's subtrees in $T(S)$. In section 4 we prove Theorem 1, and in section 5 we prove Theorems 2 and 3. In section 6 we prove Lemmas 3 and 4, which are required for Theorems 2 and 3, respectively.

2. Related work. Application areas of multidimensional search trees include graphics, computational geometry, pattern recognition, and tree classification. The *k-d tree* (Bentley [3]), obtained by letting data points define splits that are perpendicular to one of the axes, creates a structure that is exactly distributed like the ordinary one-dimensional binary search tree if split points are picked randomly from the data. The properties are independent of the underlying distribution.

Quadtrees (Samet [32]) are also based on the premise that one data point defines a split. However, the tree is 2^d-ary, as each quadrant defined by the data point corresponds to a subset of the tree. The properties of these trees depend heavily on the distribution. For the uniform density on the unit hypercube, it is known that the height H_n satisfies

$$\frac{H_n}{\ln n} \to \frac{\alpha}{d}$$

almost surely, where $\alpha = 4.311\ldots$ is as for the one-dimensional binary search tree (Devroye [7]). For additional analysis, see Flajolet et al. [12] or Devroye and Laforest ([10]). Hyperplane trees have a formidable property: their shapes are invariant under rotations and, indeed, under linear transformations in general. Rotations do alter the form of k-d trees or quadtrees, for example. This may be important in statistical applications where often one applies an appropriate linear transformation to the data to make them more manageable—this is not needed here. Furthermore, queries such as point location (see Mehlhorn [25] for definitions) can be performed in $O(\log n)$ time on the average. Unfortunately, while insertion is rather simple and deletion in $O(\log n)$ expected amortized time is achievable via lazy delete (see Cormen, Leiserson, and Rivest [5] for definitions), ordinary deletion in $O(\log n)$

expected time may take some extra care. Nevertheless, this too can be handled in logarithmic expected time per operation. If a constraint check (to see on which side of a hyperplane a point falls) is performed in one unit of time, then Theorem 1 shows that hyperplane trees are more interesting than ordinary k-d trees. This would no longer be true if constraint checks would cost d time units. Under a suitable vector calculus model, hyperplane trees may, thus, lead to improved search times, as both the quadtree and the k-d tree are based on coordinatewise comparisons.

The earliest instance of a hyperplane tree we could find is Mizoguchi, Kizawa, and Shimura [27], where its use in pattern recognition is highlighted. *Tree classifiers* based on hyperplane splitting (sometimes with the hyperplanes restricted to be parallel with one of the axes) are used and analyzed by Anderson and Fu [1], Bartolucci, Swain, and Wu [2], Breiman et al. [4], Friedman [13], Gelfand, Ravishankar, and Delp [16], Gordon and Olshen [17], Gustafson, Gelfand, and Mitter [18], Kurzynski [22], Lin and Fu [23], Meisel and Michalopoulos [26], Qing-Yun and Fu [30], Swain and Hauska [36], and You and Fu [38].

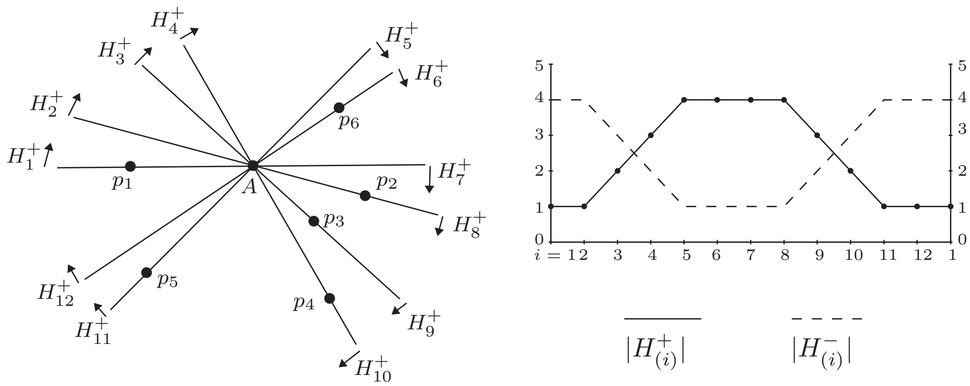
In *computational geometry*, trees based upon partitions of space by means of hyperplanes are ubiquitous. See, for example, the survey of Edelsbrunner and Van Leeuwen [11] or the work of Haussler and Welzl [19] on ϵ-nets. One may also consult Overmars and Van Leeuwen [29], Willard [37], or Mulmuley [28]. The trees obtained in ϵ-nets generalize hyperplane trees very nicely. Instead of taking d points at random to partition a convex set into two parts, one selects $s > d$ points at random and considers the partition defined by all $\binom{s}{d}$ hyperplanes defined by subsets of size d from the s points. The sets in the partition are further partitioned in the same manner. The expected height of such trees appears not to have been studied to date.

Fuchs, Kedem, and Naylor [15] introduce the "binary space partition trees" (BSP trees) for use in graphics applications. The space is split in two linear half-spaces; each half-space may in turn be split by a linear hyperplane, and so forth. If a viewer sits in a given polyhedral set in this partition and wants to project the world onto his/her view plane, the BSP tree aids in establishing the order in which the polyhedral cells must be drawn so as not to cause visibility problems. Basically, one should consider polyhedra in depth-first search order, where the depth-first search first visits half-spaces that would not contain the viewer so that polyhedra are visited from "far" to "near" (this is called the painter's algorithm). For more on the hidden surface elimination with the aid of BSP trees, see Samet [33, 34] or Fuchs, Abram, and Grant [14]. While BSP trees are not hyperplane trees (because we do not take data points to generate the partition), they are intimately related and indicate interesting applications of hyperplane trees in hidden surface elimination and beam tracing. See also Sung and Shirley [35] and Kaplan [21].

3. Safety and subtree sizes.

Definition of safety. Assume we are given a set S of $n \geq d+2$ points in $\mathbb{R}^d$ in general position, along with a constant $0 \leq \sigma \leq 1/2$. Let $A \subset S$ be the first $d-1$ points chosen for the first hyperplane when building a hyperplane search tree. We say that S is σ-safe if, for any $b \in S \setminus A$, the hyperplane defined by $A \cup \{b\}$ is such that both the corresponding open half-spaces contain at least $\sigma(n-d)$ points of S.

In section 6 we prove that if S is a set of $n \geq 4$ points randomly sampled from a convex body in $\mathbb{R}^2$ then S is σ-safe with probability at least δ for strictly positive constants σ and δ. We also prove that, for $d \geq 3$, the same is true for any deterministic set S of $n \geq d+2$ points in general position in $\mathbb{R}^d$. Observe that if the set S is in exposed position in $\mathbb{R}^2$, then S cannot be σ-safe for any $\sigma > 0$.



(a) *A line indicates each $H_{(i)}$. Each $H_{(i)}^+$ is indicated by an arrow and a label. "A" contains only the central labeled point.*

(b) *The walk done by $|H_{(i)}^+|$ is indicated with a solid line. The corresponding walk done by $|H_{(i)}^-|$ is shown with a dashed line.*

FIG. 2. *An example of a point set in $\mathbb{R}^2$ with $n = 7$.*

Intuitively, if S is σ-safe with probability δ for larger values of σ and δ, a hyperplane search tree on S will be more evenly balanced. A formalization of this idea follows.

Distribution of subtree sizes. We consider a hyperplane search tree storing a set S of n points. Let N_1 and N_2 be the number of points (not tree nodes) stored in the left and right subtrees of the root, respectively.

Let A be the set containing the first $d-1$ points chosen for the hyperplane at the root. Imagine a hyperplane H rotating around the axis defined by A. Let H^+ and H^- be the open half-spaces defined by H, where H^+ and H^- rotate along with H. As H rotates in a fixed direction it intersects the points in $S \setminus A$ one after another in a rotational order. In this order, label these points $p_1, \dots, p_{n-d+1}$. Use $H_{(i)}$ to denote the hyperplane passing through $A \cup \{p_i\}$, and use $H_{(i)}^+$ (resp., $H_{(i)}^-$) to denote the intersection of $S \setminus A$ and H^+ (resp., H^-) when $H = H_{(i)}$. Our definitions are rotational and can, therefore, be cyclic, so if we extend our definitions of $H_{(i)}^+$ and $H_{(i)}^-$ to allow indices greater than $n-d+1$, we have $H_{(i)}^+ = H_{(n-d+1+i)}^-$.

Consider the difference between $H_{(i)}^+$ and $H_{(i+1)}^+$. When H rotates from $H_{(i)}$ to $H_{(i+1)}$, p_i is added to one side of H (because H no longer passes through it) while p_{i+1} is removed from one side of H (because H now passes through it). We can now see that, over the values of i from 1 to $2(n-d+1)$, $|H_{(i)}^+|$ does a walk taking only steps of $+1$, 0, or -1. It starts, by convention, at its minimum value with $i = 1$, goes up to its maximum value at $i = n-d+1$, and returns to at most 1 plus its minimum value at $i = 2(n-d+1)$. Note that this means $|H_{(i)}^+|$ hits each value strictly between its minimum and its maximum at least twice. See Figure 2 for an example of a point set and the corresponding walk.

Given A, N_1 is distributed as $|H_{(I)}^+|$, where I is uniform on $1, \dots, 2(n-d+1)$. Therefore, for integer k we have

$$\mathbf{P}\{\max(N_1, N_2) = k\} \geq \begin{cases} \frac{1}{n-d+1}, & k = \frac{n-d}{2}, \\ \frac{2}{n-d+1}, & \frac{n-d}{2} < k < \max_i |H_{(i)}^+|. \end{cases} \tag{1}$$

For a random binary search tree on n nodes, let N_1^* and N_2^* be the number of nodes in the left and right subtrees of the root, respectively. N_1^* and N_2^* are both uniform on the integers $0, \ldots, n-1$, with the additional constraint that $N_1^* + N_2^* = n-1$. Therefore, again for integer k, we have

$$\mathbf{P}\{\max(N_1^*, N_2^*) = k\} = \begin{cases} \frac{1}{n}, & k = \frac{n-1}{2}, \\ \frac{2}{n}, & \frac{n-1}{2} < k \leq n-1. \end{cases} \tag{2}$$

Note that (1) and (2) are valid even if $n-d$ and $n-1$, respectively, are odd. The following lemma, which is used in the proof of Theorem 1, is now trivial.

LEMMA 1. $\max(N_1, N_2) \leq^s \max(N_1^*, N_2^*)$.

We now introduce a random variable X, whose purpose is to provide a sufficiently tight bound on N_1/n when our hyperplane search tree is built on a (possibly random) set that is σ-safe with probability δ.

DEFINITION 1. *For given constants $0 \leq \sigma \leq \frac{1}{2}$ and $0 \leq \delta \leq 1$, X is a continuous random variable on the range $[0,1]$ defined as follows. X is symmetric about $\frac{1}{2}$, and, for $t \leq \frac{1}{2}$,*

$$\mathbf{P}\{X \leq t\} = \begin{cases} t(1-\delta), & 0 \leq t < \sigma, \\ t, & \sigma \leq t \leq \frac{1}{2}. \end{cases}$$

From the definition of X we obtain the distribution function

$$\mathbf{P}\{\max(X, (1-X)) \leq t\} = \begin{cases} 0, & t < \frac{1}{2}, \\ 2t-1, & \frac{1}{2} \leq t < 1-\sigma, \\ \delta + (1-\delta)(2t-1), & 1-\sigma \leq t \leq 1. \end{cases} \tag{3}$$

Bounding the distribution of $\max(N_1, N_2)$ can also be done in a way that is more suitable in the context of safety. Assume our hyperplane search tree is built on a set that is σ-safe with probability at least δ. Now, for real valued t we obtain from (1),

$$\mathbf{P}\left\{\max\left(\frac{N_1}{n}, \frac{N_2}{n}\right) \leq t\right\} \geq \begin{cases} 2t-1, & \frac{n-d+1}{2n} \leq t < \frac{\lfloor(1-\sigma)(n-d)\rfloor}{n}, \\ \delta + (1-\delta)(2t-1), & \frac{\lfloor(1-\sigma)(n-d)\rfloor}{n} \leq t \leq 1. \end{cases} \tag{4}$$

To see this, note that, for $\frac{n-d+1}{2n} \leq t < \max_i |H_{(i)}^+|$, the number of integers k in the interval $[\frac{n-d+1}{2}, nt]$ times $\frac{2}{n-d+1}$ is at least $(nt - \frac{n-d+1}{2})\frac{2}{n-d+1} \geq 2t-1$.

LEMMA 2. *Let X be as defined in Definition 1 with σ and δ such that S is σ-safe with probability at least δ. Then $\max\left(\frac{N_1}{n}, \frac{N_2}{n}\right) \leq^s \max(X, 1-X)$.*

Proof. We need to prove that, for any t,

$$\mathbf{P}\left\{\max\left(\frac{N_1}{n}, \frac{N_2}{n}\right) \leq t\right\} \geq \mathbf{P}\{\max(X, 1-X) \leq t\}.$$

This is true for $t < \frac{1}{2}$ since $\max(X, 1-X) \geq \frac{1}{2}$. Now, assuming $t \geq \frac{1}{2}$ we need to take note of a few more trivial bounds, namely, $0 \leq 2t-1 \leq \delta + (1-\delta)(2t-1)$, $\frac{n-d+1}{2n} \leq \frac{1}{2}$, and $\frac{\lfloor(1-\sigma)(n-d)\rfloor}{n} \leq 1-\sigma$. Verifying the lemma case by case from (3) and (4) is now trivial. ∎

4. Bounding $D(S)$.

Proof of Theorem 1. For any $S \in \mathcal{S}_{n,d}$ and $t \leq 0$, we have $\mathbf{P}\{D(S) \geq t\} = \mathbf{P}\{D_n \geq t\} = 1$. We now need to prove that, for all $t \geq 0$,

$$\sup_{S \in \mathcal{S}_{n,d}} \mathbf{P}\{D(S) \geq t+1\} \leq \mathbf{P}\{D_n \geq t+1\}.$$

We do this by induction on n. Consider the root of the hyperplane search tree. For $t \geq 0$, $n \geq d$,

$$\mathbf{P}\{D(S) \geq t+1\} = \mathbf{E}\left\{\mathbf{P}\{D(S_1) \geq t | S_1\} \frac{|S_1|}{n} + \mathbf{P}\{D(S_2) \geq t | S_2\} \frac{|S_2|}{n}\right\}$$

where the expectation is over S_1 and S_2, the two subsets of S sent to the subtrees of the root. By the above,

$$\sup_{S \in \mathcal{S}_{n,d}} \mathbf{P}\{D(S) \geq t+1\}$$
$$\leq \mathbf{E}\left\{\sup_{S_1 \in \mathcal{S}_{N_1,d}} \mathbf{P}\{D(S_1) \geq t\} \frac{N_1}{n} + \sup_{S_2 \in \mathcal{S}_{N_2,d}} \mathbf{P}\{D(S_2) \geq t\} \frac{N_2}{n}\right\}$$

where the expectation is over the random sizes N_1 and N_2 of the subtrees of the root (with respect to *points*, not tree nodes).

We argue by induction on n that for all $t \geq 0$,

$$\sup_{S \in \mathcal{S}_{n,d}} \mathbf{P}\{D(S) \geq t+1\} \leq \mathbf{P}\{D_n \geq t+1\}.$$

For $n \leq d-1$, this is obvious, as $\mathbf{P}\{D(S) \geq t+1\} = 0$. For $n \geq d$, we have by the induction hypothesis

$$\sup_{S \in \mathcal{S}_{n,d}} \mathbf{P}\{D(S) \geq t+1\} \leq \mathbf{E}\left\{\mathbf{P}\{D_{N_1} \geq t\} \frac{N_1}{n} + \mathbf{P}\{D_{N_2} \geq t\} \frac{N_2}{n}\right\}.$$

We argue that this is bounded from above by

$$\mathbf{E}\left\{\mathbf{P}\left\{D_{N_1^*} \geq t\right\} \frac{N_1^*}{n} + \mathbf{P}\left\{D_{N_2^*} \geq t\right\} \frac{N_2^*}{n}\right\}, \tag{5}$$

where now, N_1^* and N_2^* are the sizes of the subtrees of the root of a random binary search tree (so that $N_1^* + N_2^* = n-1$; compare with $N_1 + N_2 = n-d$). If this bound is true, then we are done, because (5) is $\mathbf{P}\{D_n \geq t\}$.

Consider first the function

$$\psi(n) = n\mathbf{P}\{D_n \geq t\} = n\frac{1}{n}\sum_{i=1}^{n} \mathbf{P}\{D_i^* \geq t\} = \sum_{i=1}^{n} \mathbf{P}\{D_i^* \geq t\},$$

where D_i^* is the depth of the last node inserted in a random binary search tree on i nodes. Clearly ψ is increasing. Further, $\psi(n+1) - \psi(n) = \mathbf{P}\{D_{n+1}^* \geq t\}$, which is trivially increasing in n for fixed t, implying that ψ is also convex.

We know from Lemma 1 that $\max(N_1, N_2) \leq^s \max(N_1^*, N_2^*)$. We also know that $N_1 + N_2 \leq N_1^* + N_2^*$. Since ψ is convex, this suffices to conclude (see, e.g., Marshall and Olkin [24, pp. 3–7]) that

$$\mathbf{E}\{\psi(N_1) + \psi(N_2)\} \leq \mathbf{E}\{\psi(N_1^*) + \psi(N_2^*)\}.$$

This proves the bound in (5), completing the proof. $\square$

5. Proof of Theorems 2 and 3. Let $\lambda > 1$ be a parameter (which we later set as $\alpha - 1 \approx 3.311$). Theorems 2 and 3 hold trivially for $n \le d+2$. Now, either let $d \ge 3$, let $n \ge d+2$, and let $S \in \mathcal{S}_{n,d}$ (for the proof of Theorem 2) or let $d = 2$, let $n \ge 4$, and let S be a set of n points sampled uniformly from a convex body K in $\mathbb{R}^2$ (for the proof of Theorem 3). Let $H(S)$ be the height of the hyperplane search tree in dimension d. We index the levels of an infinite tree by setting the root level to zero. The tree is infinite and complete, with the understanding that internal nodes hold d data points and leaf nodes hold between 1 and $d-1$ data points. All other nodes correspond to zero data points and are just added to make the tree formally complete and infinite. For a node u, we call N_u the number of data points (not nodes!) in the subtree rooted at u, and let $|u|$ denote its level number.

Observe that if $H(S)$ denotes the height, then by the union bound and Markov's inequality

$$\begin{aligned}\mathbf{P}\{H(S) \ge k\} &\le \mathbf{P}\{\max_{u:|u|=k} N_u \ge 1\} \\ &\le \sum_{u:|u|=k} \mathbf{P}\{N_u \ge 1\} \\ &\le \sum_{u:|u|=k} \mathbf{E}\left\{N_u^\lambda\right\}.\end{aligned}$$

By Lemmas 3 and 4 (see section 6) there exist $\sigma > 0$ and $\delta > 0$ (not depending on d) such that S is σ-safe with probability at least δ. We show that the corresponding $[0,1]$-valued random variable X (see Definition 1) acts as what is commonly called a split vector in the random search tree literature (see, e.g., Devroye [9]) and show by induction that

$$\mathbf{E}\left\{\sum_{u:|u|=k} N_u^\lambda\right\} \le n^\lambda \left(2\mathbf{E}\left\{X^\lambda\right\}\right)^k. \tag{6}$$

Furthermore, we have that for each $\lambda > 1$,

$$\mathbf{E}\left\{X^\lambda\right\} < \frac{1}{\lambda+1}. \tag{7}$$

We prove (6) and (7) below, but for now we assume they are true to complete the proofs. For $x > 0$, let $f(x) = \rho^x \left(\frac{2}{\alpha}\right)^{x-\alpha}$. Then f is continuous and $f(\alpha) = \rho^\alpha < 1$. Hence, there exist $c < \alpha$ (with $c > 0$) and $\epsilon > 0$ such that $f(c) = e^{-\epsilon} < 1$.

Consider $k = \lceil c \ln n \rceil$ and $\lambda = \alpha - 1$. Define the ratio ρ of $\mathbf{E}\{X^{\alpha-1}\}$ and $1/(\lambda+1)$:

$$\rho = \alpha \mathbf{E}\left\{X^{\alpha-1}\right\} < 1.$$

Then the right-hand side (RHS) of (6) equals

$$\begin{aligned}n^{-1} e^{\alpha \ln n} \left(\frac{2\rho}{\alpha}\right)^k &= n^{-1} \left(\frac{2e}{\alpha}\right)^{\alpha \ln n} \left(\frac{2}{\alpha}\right)^{k - \alpha \ln n} \rho^k \\ &= \left(\frac{2}{\alpha}\right)^{k-\alpha \ln n} \rho^k\end{aligned}$$

since $(\frac{2e}{\alpha})^{\alpha \ln n} = e^{\ln n} = n$. Thus, using also $\frac{2\rho}{\alpha} \leq 1$, the RHS of (6) equals

$$\left(\frac{2\rho}{\alpha}\right)^k \left(\frac{2}{\alpha}\right)^{-\alpha \ln n} \leq \left(\frac{2\rho}{\alpha}\right)^{c \ln n} \left(\frac{2}{\alpha}\right)^{-\alpha \ln n} = \left(\rho^c \left(\frac{2}{\alpha}\right)^{c-\alpha}\right)^{\ln n} = e^{-\epsilon \ln n} = n^{-\epsilon}.$$

To complete the proof of Theorems 2 and 3 we must now prove (6) and (7).

Proof of (6) *and* (7). We recall from Lemma 2 that

$$\max\left(\frac{N_1}{n}, \frac{N_2}{n}\right) \leq^s \max(X, 1-X).$$

We also know that

$$\frac{N_1}{n} + \frac{N_2}{n} = \frac{n-d}{n} < X + (1-X).$$

Therefore, in the sense of Marshall and Olkin [24], the vector $\left(\frac{N_1}{n}, \frac{N_2}{n}\right)$ is dominated by the vector $(X, 1-X)$. Note, in particular, that there are couplings such that this domination is on the same probability space, in other words, such that $\max\left(\frac{N_1}{n}, \frac{N_2}{n}\right) \leq \max(X, 1-X)$ with probability one. Because of this domination and because the function u^λ is convex in u on the positive half-line, we have

$$\begin{aligned}\mathbf{E}\left\{N_1^\lambda + N_2^\lambda\right\} &= n^\lambda \mathbf{E}\left\{\left(\frac{N_1}{n}\right)^\lambda + \left(\frac{N_2}{n}\right)^\lambda\right\} \\ &\leq n^\lambda \mathbf{E}\left\{X^\lambda + (1-X)^\lambda\right\} \\ &= 2n^\lambda \mathbf{E}\left\{X^\lambda\right\}.\end{aligned}$$

Applying this at level k and conditioning on subtree sizes (note that conditional on subtree sizes, all subtrees rooted at one level are independent), we have

$$\mathbf{E}\left\{\sum_{u:|u|=k} N_u^\lambda\right\} \leq \mathbf{E}\left\{\sum_{u:|u|=k-1} N_u^\lambda\right\} \times \left(2\mathbf{E}\{X^\lambda\}\right).$$

By induction on k, we, thus, obtain

$$\mathbf{E}\left\{\sum_{u:|u|=k} N_u^\lambda\right\} \leq n^\lambda \times \left(2\mathbf{E}\{X^\lambda\}\right)^k.$$

This proves (6).

To show (7), note that X has a distribution best described as follows: it has density 1 on $[\sigma, 1-\sigma]$, density $1-\delta$ on $[0, \sigma)$ and $(1-\sigma, 1]$, and point masses equal to $\sigma\delta$ at σ and $1-\sigma$. Then

$$\begin{aligned}\mathbf{E}\{X^\lambda\} &= \sigma\delta\left(\sigma^\lambda + (1-\sigma)^\lambda\right) + \frac{1}{\lambda+1} - \delta\frac{\sigma^{\lambda+1}}{\lambda+1} - \delta\left(\frac{1-(1-\sigma)^{\lambda+1}}{\lambda+1}\right) \\ &= \frac{\delta}{\lambda+1}\left(\sigma^{\lambda+1}\lambda + (1-\sigma)^{\lambda+1} + \sigma(1-\sigma)^\lambda(\lambda+1) - 1\right) + \frac{1}{\lambda+1}.\end{aligned}$$

Consider the first term as a function of σ. At $\sigma = 0$, its value is zero. Its derivative is easily seen to be nonpositive for $0 \leq \sigma \leq 1/2$, and since σ is clearly in this range, the first term is nonpositive. It is, in fact, strictly negative for $\sigma \in (0, 1/2]$ and $\delta > 0$ when $\lambda > 1$. This proves (7). ∎

6. Proofs of safety lemmas.

LEMMA 3. *There exist $\sigma > 0$ and $\delta > 0$ such that, for any $d \geq 3$ and any $S \subseteq \mathbb{R}^d$ with $|S| \geq d+2$ in general position, S is σ-safe with probability at least δ.*

LEMMA 4. *There exist $\sigma > 0$ and $\delta > 0$ such that the following holds. For a convex body K in $\mathbb{R}^2$, let S be a set of $n \geq 4$ random points drawn independently and uniformly from K. Then S is σ-safe with probability at least δ.*

We shall first prove Lemma 3, using three further lemmas. Let W be a d-dimensional real vector space (that is, a copy of $\mathbb{R}^d$) where $d \geq 2$. We say a hyperplane $H \subset W$ *separates* two points $a, b \in W$ if H does not contain either point but H intersects the line segment connecting them, i.e., if a and b are strictly on opposite sides of H. A *cross* in W is a set S of $d-1$ points and a disjoint set T of 3 points (with $S \cup T$ in general position) such that every hyperplane H containing S separates two points of T. Thus, in $\mathbb{R}^2$ a cross is a pair $\{a\}, \{b, c, d\}$ such that a is inside triangle bcd, and in $\mathbb{R}^3$ a cross is a pair $\{a, b\}, \{c, d, e\}$ such that the line passing through a and b intersects triangle cde.

LEMMA 5. *Let W be a three-dimensional real vector space (that is, a copy of $\mathbb{R}^3$). From any set Q of 5 points in general position in W we may form at least one cross.*

Proof. Let the polytope P be the convex hull of Q. The graph with vertex set Q and edge set $E(P)$ is planar, and so it cannot be the complete graph K_5. So some pair $\{a, b\} \subset Q$ does not form an edge of G, and then $\{a, b\}, Q \setminus \{a, b\}$ is a cross. ∎

LEMMA 6. *For any $d \geq 3$ let W be a d-dimensional real vector space (that is, a copy of $\mathbb{R}^d$). Let Q be a set of $d+2$ points in general position in W. Then for any set $A \subset Q$ with $|A| = d-3$ there is a cross $S, Q \setminus S$ with $Q \supset S \supset A$.*

Proof. We prove this by induction on d, where the last lemma handles the base case with $d = 3$. We now assume $d > 3$ and the lemma holds for all dimensions at least 3 and strictly smaller than d.

Consider the polytope P defined as the convex hull of $Q \setminus A$, recalling that $Q \setminus A$ is a set of 5 points in general position. For $d > 3$, P is a nondegenerate 4-simplex. We consider two cases based on whether or not $A \subset P$.

If $A \subset P$, then P must be full-dimensional; thus, $d = 4$ and $|A| = 1$. Also A must be in the interior of P (since Q is in general position). Any hyperplane H containing A must cut P into two full-dimensional parts, each of which must contain at least one vertex of P that is not on H. This implies that $A \cup \{p_1, p_2\}$, $Q \setminus (A \cup \{p_1, p_2\})$ is a cross for any pair of points $\{p_1, p_2\} \subset Q \setminus A$, proving the lemma in this case.

Otherwise some point $p \in A$ must be strictly outside P. Let H be a hyperplane that separates p from every point in $Q \setminus A$. Suppose that we were to translate all points by a fixed vector in H. Then crosses stay crosses, and sets of points in general position stay in general position. Thus, we may assume without loss of generality that H contains the origin. Now, H is a $(d-1)$-dimensional subspace of W (a copy of $\mathbb{R}^{d-1}$). We label $Q \setminus \{p\} = \{z_1, \ldots, z_{d+1}\}$ and define $Q_H = \{y_1, \ldots, y_{d+1}\}$, where y_i is the point where the line passing through p and z_i intersects H for $1 \leq i \leq d+1$. Note that a flat contains p and y_i if and only if it contains p and z_i. Define $A_H \subset Q_H$ as $\{y_i : z_i \in A \setminus \{p\}\}$. We note that Q_H is a set of $d+1$ points in general position in H for the following reason. For any $k < d-1$, if $k+2$ points $y_{s(1)}, \ldots, y_{s(k+2)}$ in Q_H lie on a k-dimensional flat for some permutation s of $1, \ldots, d+1$, then $p, y_{s(1)}, \ldots, y_{s(k+2)}$ lie on a $(k+1)$-dimensional flat, and this flat then contains $k+3$ points from Q, namely, $p, z_{s(1)}, \ldots, z_{s(k+2)}$. This cannot happen since Q is in general position, so Q_H must, indeed, be in general position in H.

Now, by our induction hypothesis, Q_H contains a cross $S_H, Q_H \setminus S_H$ in the $(d-1)$-dimensional space H, where $Q_H \supset S_H \supset A_H$. Define $S \subset Q$ as $\{z_i : y_i \in S_H\}$, and note that H separates p from S. We argue that $S \cup \{p\}, Q \setminus (S \cup \{p\})$ is a cross in our original d-dimensional space W.

Any hyperplane H' containing $S \cup \{p\}$ intersects H at a corresponding $(d-2)$-dimensional flat H'_H containing S_H. Since $S_H, Q_H \setminus S_H$ is a cross, H'_H separates $y_i, y_j \in Q_H \setminus S_H$ for some indices i and j. But y_i and y_j are not in H' (since they are in H and not in H'_H), and, thus, H' separates them. The two open half-lines starting at p and, respectively, passing through y_i and y_j must, respectively, include z_i and z_j since H separates p from z_i and z_j. Since these half-lines do not intersect H', H' must separate z_i from z_j just as it separates y_i from y_j. Therefore, $S \cup \{p\}, Q \setminus (S \cup \{p\})$ is a cross in our original d-dimensional space W. ∎

Given a set S of at least d points in $\mathbb{R}^d$, for each $A \subseteq S$ with $|A| = d-1$, let $f(A)$ be the minimum over all $b \in S \setminus A$ of the smaller number of points of S in one of the open half-spaces determined by the hyperplane spanned by $A \cup \{b\}$.

LEMMA 7. *Let $d \geq 3$, let $n \geq d+2$, and let S be a set of n points in $\mathbb{R}^d$ in general position. Then if X is chosen uniformly at random from the $(d-1)$-subsets of S,*

$$\mathbf{E}\{f(X)\} \geq \frac{n-d+1}{30}.$$

Proof. By the last lemma, the total number of crosses we may form from a set Q of $d+2$ points in $\mathbb{R}^d$ is at least

$$\frac{\binom{d+2}{d-3}}{\binom{d-1}{d-3}} = \frac{\binom{d+2}{5}}{\binom{d-1}{2}} = \frac{d(d+1)(d+2)}{60}.$$

So the total number of crosses we may form from S is at least

$$\frac{d(d+1)(d+2)}{60}\binom{n}{d+2}.$$

Let B be a $(d-1)$-subset of S. Let $g(B)$ denote the number of crosses with first set B. Let H be any hyperplane containing B and some other point of S. Suppose the numbers of points of S in the two corresponding half-spaces are h_1 and h_2, respectively. Then $h_1 + h_2 = n - d$, and

$$g(B) = h_1\binom{h_2}{2} + h_2\binom{h_1}{2} \leq \frac{h_1 h_2 (n-d)}{2}.$$

So $h_1 h_2 \geq \frac{2g(B)}{n-d}$. But $\max(h_1, h_2) \leq n - d - 1$ if $g(B) > 0$, so

$$\min(h_1, h_2) \geq \frac{2g(B)}{(n-d)(n-d-1)}.$$

Thus, $f(B)$ is at least the RHS above. Hence, by averaging over sets B,

$$\begin{aligned}\mathbf{E}\{f(X)\} &\geq \frac{2\mathbf{E}\{g(X)\}}{(n-d)(n-d-1)} \\ &\geq \frac{2d(d+1)(d+2)}{60(n-d)(n-d-1)}\binom{n}{d+2}\frac{1}{\binom{n}{d-1}} \\ &= \frac{n-d+1}{30}\end{aligned}$$

as required. ∎

LEMMA 8. *Lemma* 3 *holds with* $\sigma = \frac{1}{60}$ *and* $\delta = \frac{1}{30}$.

Proof. Let X denote a random $(d-1)$-subset of S. It must be the case that $\mathbf{P}\{f(X) \geq (n-d+1)/60\} \geq 1/30$; for if not, then we would have

$$\mathbf{E}\{f(X)\} < \left(\frac{29}{30}\right)\frac{n-d+1}{60} + \left(\frac{1}{30}\right)\frac{n-d}{2} < \frac{n-d+1}{30},$$

which would contradict Lemma 7. ∎

We now prove Lemma 4, for which we need one preliminary lemma.

LEMMA 9. *Let* $X_1, X_2, X_3, \ldots$ *be independent identically distributed random variables with a common density function* f *in* $\mathbb{R}^2$. *Let* β *be the probability that* X_1 *is in the convex hull of* X_2, X_3, X_4. *Then for all* $n \geq 4$,

$$\mathbf{P}\left\{X_1 \text{ is } \frac{\beta}{6}\text{-safe in } \{X_1, \ldots, X_n\}\right\} \geq \frac{\beta}{3}.$$

Proof. From $X_2, \ldots, X_n$, form $\left\lfloor \frac{n-1}{3} \right\rfloor$ disjoint sets A of size 3. Let Z be the number of these sets A such that X_1 is in the convex hull of A. Thus, $\mathbf{E}\{Z\} = \beta \left\lfloor \frac{n-1}{3} \right\rfloor$. It suffices to show that $\mathbf{P}\{Z \geq \frac{\beta}{6}(n-2)\} \geq \frac{\beta}{3}$. But if this were false, we would have

$$\beta \left\lfloor \frac{n-1}{3} \right\rfloor = \mathbf{E}\{Z\} < \left(1 - \frac{\beta}{3}\right)\frac{\beta}{6}(n-2) + \frac{\beta}{3}\left\lfloor \frac{n-1}{3} \right\rfloor,$$

and then

$$4\left\lfloor \frac{n-1}{3} \right\rfloor < \left(1 - \frac{\beta}{3}\right)(n-2) < n-2,$$

which is false for $n \geq 4$. ∎

Proof of Lemma 4. Let β_0 be the value of β for the uniform distribution over the unit ball in $\mathbb{R}^2$. Clearly $\beta_0 > 0$ (it can actually be shown that $\beta_0 \geq 1/(9\pi)$). John's theorem [20] tells us that, for any convex body K in the plane, there is a nonsingular linear transformation τ such that τK contains a ball of area $\frac{1}{4}$ that of τK (note that crosses remain crosses under nonsingular linear transformations of a point set). Hence, the value β_K for K satisfies $\beta_K \geq (\frac{1}{4})^4 \beta_0$. Lemma 4 now follows from Lemma 9. ∎

Acknowledgments. The main ideas in this paper were discussed at Universidad Simon Bolivar during a workshop held in January 1993. Discussions with Bruce Reed are gratefully acknowledged. We are grateful to the anonymous referees whose comments led us to improve and clarify the paper.

REFERENCES

[1] A. C. ANDERSON AND K. S. FU, *Design and Development of a Linear Binary Tree Classifier for Leukocytes*, Technical report, Purdue University, West Lafayette, IN, 1979.

[2] L. A. BARTOLUCCI, P. H. SWAIN, AND C. WU, *Selective radiant temperature mapping using a layered classifier*, IEEE Trans. Geosci. Electron., GE-14 (1976), pp. 101–106.

[3] J. L. BENTLEY, *Multidimensional binary search trees used for associative searching*, Commun. ACM, 18 (1975), pp. 509–517.

[4] L. BREIMAN, J. H. FRIEDMAN, R. A. OLSHEN, AND C. J. STONE, *Classification and Regression Trees*, Wadsworth International, Belmont, CA, 1984.

[5] T. H. CORMEN, C. E. LEISERSON, AND R. L. RIVEST, *Introduction to Algorithms*, MIT Press, Boston, MA, 1990.

[6] L. Devroye, *A note on the height of binary search trees*, J. ACM, 33 (1986), pp. 489–498.
[7] L. Devroye, *Branching processes in the analysis of the heights of trees*, Acta Inform., 24 (1987), pp. 277–298.
[8] L. Devroye, *Applications of the theory of records in the study of random trees*, Acta Inform., 26 (1988), pp. 123–130.
[9] L. Devroye, *Universal limit laws for depths in random trees*, SIAM J. Comput., 28 (1998), pp. 409–432.
[10] L. Devroye and L. Laforest, *An analysis of random d-dimensional quad trees*, SIAM J. Comput., 19 (1990), pp. 821–832.
[11] H. Edelsbrunner and J. van Leeuwen, *Multidimensional Data Structures and Algorithms: A Bibliography*, Technical report, Technische Universität Graz, Graz, Austria, 1983.
[12] P. Flajolet, G. Gonnet, C. Puech, and J. M. Robson, *The analysis of multidimensional searching in quad-trees*, in Proceedings of the Second Annual ACM-SIAM Symposium on Discrete Algorithms, Philadelphia, 1991, pp. 100–109.
[13] J. H. Friedman, *A tree-structured approach to nonparametric multiple regression*, in Smoothing Techniques for Curve Estimation, Lecture Notes in Math. 757, T. Gasser and M. Rosenblatt, eds., Springer-Verlag, Heidelberg, 1979, pp. 5–22.
[14] H. Fuchs, G. D. Abram, and E. D. Grant, *Near real-time shaded display of rigid objects*, Computer Graphics, 17 (1983), pp. 65–72.
[15] H. Fuchs, Z. M. Kedem, and B. Naylor, *On visible surface generation by a priori tree structures*, Comput. Graph., 14 (1980), pp. 124–133.
[16] S. B. Gelfand, C. S. Ravishankar, and E. J. Delp, *An iterative growing and pruning algorithm for classification tree design*, IEEE Trans. Pattern Anal. Mach. Intelligence, 13 (1991), pp. 163–174.
[17] L. Gordon and R. A. Olshen, *Asymptotically efficient solutions to the classification problem*, Ann. Statist., 6 (1978), pp. 515–533.
[18] D. E. Gustafson, S. Gelfand, and S. K. Mitter, *A nonparametric multiclass partitioning method for classification*, in Proceedings of the Fifth International Conference on Pattern Recognition, Miami, FL, International Association for Pattern Recognition, 1980, pp. 654–659.
[19] D. Haussler and E. Welzl, *Epsilon-nets and simplex range queries*, Discrete Comput. Geom., 2 (1987), pp. 127–151.
[20] F. John, *Extremum problems with inequalities as subsidiary conditions*, in Studies and Essays, Courant Anniversary Volume, Interscience, New York, 1948, pp. 187–204.
[21] M. R. Kaplan, *The uses of spatial coherence in ray tracing*, in Proceedings of SIGGRAPH '85 Course Notes, San Francisco, CA, ACM, 1985, pp. 22–26.
[22] M. W. Kurzynski, *The optimal strategy of a tree classifier*, Pattern Recognition, 16 (1983), pp. 81–87.
[23] Y. K. Lin and K. S. Fu, *Automatic classification of cervical cells using a binary tree classifier*, Pattern Recognition, 16 (1983), pp. 69–80.
[24] A. W. Marshall and I. Olkin, *Inequalities: Theory of Majorization and its Applications*, Academic Press, New York, 1979.
[25] K. Mehlhorn, *Data Structures and Algorithms* 3: *Multi-Dimensional Searching and Computational Geometry*, Springer-Verlag, Berlin, 1984.
[26] W. S. Meisel and D. A. Michalopoulos, *A partitioning algorithm with application in pattern classification and the optimization of decision trees*, IEEE Trans. Comput., C-22 (1973), pp. 93–103.
[27] R. Mizoguchi, M. Kizawa, and M. Shimura, *Piecewise linear discriminant functions in pattern recognition*, Syst. Comput. Control, 8 (1977), pp. 114–121.
[28] K. Mulmuley, *Randomized multidimensional search trees: Dynamic sampling*, in Proceedings of the 7th Annual Symposium on Computational Geometry, North Conway, NH, ACM, 1991, pp. 121–131.
[29] M. H. Overmars and J. van Leeuwen, *Dynamic multidimensional data structures based on quad- and k-d trees*, Acta Inform., 17 (1982), pp. 265–287.
[30] S. Qing-Yun and K. S. Fu, *A method for the design of binary tree classifiers*, Pattern Recognition, 16 (1983), pp. 593–603.
[31] J. M. Robson, *The height of binary search trees*, Australian Comput. J., 11 (1979), pp. 151–153.
[32] H. Samet, *The quadtree and related hierarchical data structures*, Comput. Surv., 16 (1984), pp. 187–260.
[33] H. Samet, *Applications of Spatial Data Structures*, Addison–Wesley, Reading, MA, 1990.

[34] H. Samet, *The Design and Analysis of Spatial Data Structures*, Addison–Wesley, Reading, MA, 1990.
[35] K. Sung and P. Shirley, *Ray tracing with the bsp tree*, in Graphics Gems III, D. Kirk, ed., Academic Press, Boston, 1992, pp. 271–274.
[36] P. H. Swain and H. Hauska, *The decision tree classifier: Design and potential*, IEEE Trans. Geosci. Electron., GE-15 (1977), pp. 142–147.
[37] D. E. Willard, *Polygon retrieval*, SIAM J. Comput., 11 (1982), pp. 149–165.
[38] K. C. You and K. S. Fu, *An approach to the design of a linear binary tree classifier*, in Proceedings of the Third Symposium of Machine Processing of Remotely Sensed Data, West Lafayette, IN, IEEE, 1976, pp. 3A-10.