A New Lower Bound Technique for Decision Trees [1]

by

Rudolf Fleischer [2]

ABSTRACT In this paper, we prove two general lower bounds for algebraic decision trees which test membership in a set $S \subseteq \mathbb{R}^n$ which is defined by linear inequalities. Let $rank(S)$ be the maximal dimension of a linear subspace contained in the closure of S.

First we prove that any decision tree which uses multilinear functions (i.e. arbitrary products of linear functions) must have depth at least $n - rank(S)$. This solves an open question raised by A.C. Yao ([Y89]) and can be used to show that multilinear functions are not really more powerful than simple comparisons between the input variables when computing the largest k elements of n given numbers. Yao could only prove this result in the special case when products of at most two linear functions are used. Our proof is based on a dimension argument. It seems to be the first time that such an approach yields good lower bounds for nonlinear decision trees.

Surprisingly, we can use the same methods to give an alternative proof for Rabin's fundamental Theorem ([Rab]), namely that the depth of any decision tree using arbitrary analytic functions is at least $n - rank(S)$. Since we show that Rabin's original proof is incorrect, our proof of Rabin's Theorem is not only the first correct one but also generalizes the Theorem to a wider class of functions.

1. Introduction

Among other algebraic complexity measures (e.g. [MST], [ST]), the algebraic decision tree and algebraic computation tree models have have turned out to be very useful in proving lower bounds for elementary combinatorial or geometric problems like maximum finding, set equality, set disjointness and sorting (see [Be] for more examples) or even more complicated problems like convex polygon inclusion ([Ram]) and motion planning ([OD]).

The algebraic decision tree model is an abstraction of "real" algorithms where only comparisons between input variables or functions of input variables are counted whereas all other time-consuming operations like data-management, function evaluation or other control structures have zero cost. For complex problems, this simplification can make the problem considerably easier; for example, the knapsack problem which is known to be NP-complete has a polynomial solution in the decision tree model (e.g. [MadH]). Therefore, lower bounds in the decision tree model can only be tight for quite simple problems.

---

[1] This work was (partially) supported by the ESPRIT Basic Research Action No. 7141 (ALCOM II)

[2] Max-Planck-Institut für Informatik, W-6600 Saarbrücken, Germany, e-mail: rudolf@mpi-sb.mpg.de

A *decision problem* is a disjoint partition $\mathbb{R}^n = S_1 \cup \cdots \cup S_q$ of $\mathbb{R}^n$ into arbitrary sets S_i. A *decision tree* T for a decision problem is a binary tree whose internal nodes are labeled by predicates defined on $\mathbb{R}^n$, whose outgoing edges of an internal node are labeled by $true$ or $false$, and whose leaves are labeled by one of the S_i (Fig. 3.1 shows an example). The evaluation of T on input $x \in \mathbb{R}^n$ starts at the root and then proceeds downwards by evaluating the predicate at an internal node and taking the appropriate of the two outgoing edges. Finally, a leaf with label S_x is reached. S_x is the result of the computation, the path x followed is the *computation path* of x, and T is correct if $x \in S_x$ for all x. The worst-case running time of T is the length of the longest computation path in T.

T is called an *algebraic decision tree* if all functions evaluated at internal nodes are defined by polynomials. The most restricted algebraic decision trees are *comparison trees* where only comparisons between two input variables are allowed ([FG], [MT]). *Linear decision trees* ([Sn82]) where linear functions of the input variables can be used ([BLY], [DL], [FG], [MT], [RY]) are more powerful. Products of linear functions were used in [Y89]) and arbitrary polynomials of bounded degree in [Be] and [Y92]. Finally, arbitrary analytic functions were allowed in [Ja] and [Rab]. Of course, this classification of algebraic decision trees is not exhaustive and many less natural restrictions on the functions can be found in the literature (e.g. [DL], [MMS]).

In our paper, we only prove lower bounds for algebraic decision trees, but we also want to mention the *algebraic computation tree* model where additional computing nodes exist which evaluate elementary functions on all input variables and previously computed functions ([Be], [OD]). Of course, this can be simulated by an algebraic decision tree which uses polynomials of increasing degree, but this model is much closer to the model of "real" algorithms (e.g. [PS]). Probabilistic and nondeterministic decision trees have also been studied ([MT], [Sn85]).

From now on, we assume that the decision problem is a membership problem, i.e. we want to decide whether an input $x \in \mathbb{R}^n$ is in a set $S \subseteq \mathbb{R}^n$ (we call S the *target set*), but the lower bounds mentioned below can easily be transformed into similar bounds for arbitrary decision problems. Several lower bound techniques are known for algebraic decision trees. One of the first and most widely used arguments was that the logarithm of the number of connected components of S is a lower bound for the depth of linear decision trees ([DL]. This was generalized in [Be] to bounded degree decision trees. And recently Ramanan has shown how this technique can sometimes give strong lower bounds even for sets S which do not have many connected components by intersecting S with another easily computable set and thus increasing the number of components ([Ram]). Another old approach is to use an adversary argument to show that there must be at least one long path in the tree ([DL]) or proving the existence of many disjoint subtrees ([FG]).

An interesting topological lower bound, at least for linear decision trees, was proved by Rivest and Yao. They showed that the logarithm of the number of s-dimensional faces of S for arbitrary s is a lower bound ([RY]). And very recently, Yao et. al. have found a way to exploit even a much more complex topological property of S, i.e. they showed that the logarithm of the Euler characteristic of S is a lower bound not only for linear decision trees ([BLY]) but also for bounded degree decision trees ([Y92]).

The only known results about analytic decision trees are due to Rabin ([Rab]) and Jaromczyk ([Ja]). Rabin proved the fundamental Theorem that any analytic decision tree for S must have depth $|H|$ if S is defined by a set H of independent linear inequalities. Jaromczyk tried to generalize Rabin's Theorem to sets S defined by arbitrary polynomial inequalities but faced heavy problems concerning real algebraic varieties which he could not solve adequately. Unfortunately, Rabin's proof is not correct (see Section 3). The same is probably true for Jaromczyk's generalization because Rabin's error lies in a basic definition which is also used by Jaromzyk.

Surprisingly, we can give an alternative (and hopefully correct) proof of Rabin's Theorem by using a new lower bound technique which we have developed to solve an open question raised by Yao. In [Y92], Yao showed that median tests are not really more powerful than simple comparisons between the input variables when computing the largest k elements of n given numbers. He raised the question whether this can be generalized to functions which are arbitrary products of linear functions (the median test can be written as the product of only two linear functions).

We show in this paper that it can be generalized. Our proof technique is based on a dimension argument and works only for sets S defined by linear inequalities. Let $rank(S)$ be the maximal dimension of a linear subspace contained in the closure of S. We show that, for any computation path p in the decision tree, the closure of the set of inputs x which have computation path p always contains a linear subspace of dimension $n - length(p)$. Hence $length(p) \geq n - rank(S)$.

It seems to be the first time that a dimension argument is used to derive good lower bounds for nonlinear decision trees. And it seems to be the first time that someone has looked carefully in Rabin's proof and, finding it incorrect, gives a correct proof. This is even more important because no other lower bounds for this most general model are known.

This paper is organized as follows. In Section 2 we give some geometric definitions and Lemmas. In Section 3 we define certificates and proofs, an abstraction of the decision tree model, and show why we consider Rabin's proof incorrect. The generalization of Yao's Theorem and Rabin's corrected Theorem follow in Sections 4 and 5, respectively. And we conclude with some remarks in Section 6.

2 . Geometric Preliminaries

In this Section we will give some elementary definitions. We use the notations $\overline{B} = cl\, B$ and $B^0 = int\, B$, where $B \subseteq \mathbb{R}^n$ is any set. $B_n(z, \epsilon)$ denotes the n–dimensional ball of radius ϵ centered at z.

A *linear variety (flat)* in $\mathbb{R}^n$ is a subset $G \subseteq \mathbb{R}^n$ of the form $G = v + L$ where $L \subseteq \mathbb{R}^n$ is a linear subspace and $v \in \mathbb{R}^n$. The flat has dimension $dim\, G := dim\, L$.

Let $L_n := \{\lambda_0 + \sum_{i=1}^{n} \lambda_i x_i \mid \lambda_0, \ldots, \lambda_n \in \mathbb{R}\}$ be the set of linear functions in n real variables $x = (x_1, \ldots, x_n)$. Let $L_n^{(j)} := \{l_1 \cdots l_j \mid l_i \in L_n, \forall i\}$ be the set of *multilinear functions* of degree j in n variables which is a proper subset of the set of all polynomials of degree j in n variables and $L_n^* := L_n^\infty = \bigcup_{j=1}^{\infty} L_n^{(j)}$.

Each $l \in L_n$ induces an oriented hyperplane $h = \{x \in \mathbb{R}^n \mid l(x) = 0\}$ with normal vector n. Let $\Delta := \{>, \geq, =, \neq\}$ be the set of all comparison operators and $\Delta_> := \{>, \geq\}$. For a set $L = \{l_1, \ldots, l_m\} \subset L_n$ of linear functions we usually denote the set of its corresponding hyperplanes by $H = \{h_1, \ldots, h_m\}$. $Arr(H)$ is the arrangement of the hyperplanes in H (see [Ed] for details of arrangements). If $Op = \{op_1, \ldots, op_m\}$, $op_i \in \Delta$, is a set of comparison operators then we define the set of simultaneous solutions of L with respect to Op as $S_{L,Op} := \{x \in \mathbb{R}^n \mid l_i(x)\ op_i\ 0,\ i = 1, \ldots, m\}$; if $L = \emptyset$ then $S_{L,Op} := \mathbb{R}^n$. We omit the index Op whenever it is clear from the context which Op should be used.

To measure the degree of independence of the linear functions in L we introduce the rank of L : $rank(L) := n - dim\, span(n_1, \ldots, n_m)$ where n_i is the normal vector of the hyperplane h_i defined by l_i. The following Lemma shows that $rank(L)$ is the maximal dimension of a linear variety contained in $\overline{S_L}$.

Lemma 2.1 Let $L = \{l_1, \ldots, l_m\} \subset L_n$ be a set of linear functions and let $Op \in \Delta_>^m$ be arbitrary. Let H be the set of hyperplanes defined by L. If $S_L \neq \emptyset$ then

(a) S_L contains a linear variety of dimension $rank(L)$;

(b) $\overline{S_L}$ does not contain a linear variety of dimension $rank(L) + 1$;

(c) $rank(L) = \min\{k \mid Arr(H) \text{ contains a } k\text{-face}\}$, and there is a $rank(L)$-face of $Arr(H)$ contained in $\overline{S_L}$.

Proof : Let h_i be the hyperplane corresponding to the function l_i with normal vector n_i.

(a) Let $z \in S_L$ be arbitrary and let V be a linear subspace of dimension $rank(L)$ such that $span(V, n_1, \ldots, n_m) = \mathbb{R}^n$. Then V is perpendicular to all normal vectors n_i and therefore parallel to all hyperplanes h_i; hence $z + V \subseteq S_L$.

(b) If V is any linear variety contained in $\overline{S_L}$ then V must be parallel to all hyperplanes h_i, i.e. perpendicular to all normal vectors n_i. Hence $dim\, V \leq n - dim\, span(n_1, \ldots, n_m)$.

(c) Any k-face in $Arr(H)$ is a k-dimensional subset of the intersection of $n - k$ linearly independent hyperplanes of H ([Ed]). Since there are at not more than $span(n_1, \ldots, n_m)$ linearly independent hyperplanes in H, $Arr(H)$ contains only k-faces for $k \geq n - dim\, span(n_1, \ldots, n_m) = rank(L)$. Furthermore, all cells in $Arr(H)$ are bounded by at least one $rank(L)$-face. □

Let $L \subset L_n$ be a set of linear functions and H the set of hyperplanes defined by L. Then $Arr(H)$ induces a *signature* on the points $x \in \mathbb{R}^n$: $sig(x) := (\epsilon_1, \ldots, \epsilon_m)$ where $\epsilon_i = -, 0, +$ iff x lies under, on, above h_i, respectively. We call L *sign-independent* if all possible signatures are realized in $Arr(H)$. The notion of sign-independence was introduced by Rabin ([Rab]) but he failed to observe the following simple characterization of sign-independence in terms of $rank(L)$.

Lemma 2.2 Let $L = \{l_1, \ldots, l_m\} \subset L_n$ and let h_i be the hyperplane defined by l_i. Then

$$L \text{ is sign-independent} \iff dim \bigcap_{i=1}^{m} h_i = n - m \iff rank(L) = n - m.$$

Proof : Elementary geometry, or see [Ed]. □

The definitions above will mainly be used in the next Section for defining the target set, i.e. the subset of $\mathbb{R}^n$ which is computed by the decision tree. But we also need some definitions concerning the functions which are used at internal nodes of the decision tree. Yao restricted the internal functions to products of linear functions, whereas Rabin allowed arbitrary analytic functions. We give a more general framework by defining the functions by means of the properties they should have (to make our proofs work). Theorem 2.4 shows that our model includes the models of Yao and Rabin.

Let $F = (F_n)_{n=1,2,3,\ldots}$ be a family of sets of real-valued functions in n real variables with the following properties :

Let $f_n, g_n \in F_n$ and h be a hyperplane in $\mathbb{R}^n$.

(F1) f_n is continuous.

(F2) $f_n \cdot g_n \in F_n$, i.e. F_n is closed under multiplication.

(F3) F_n is closed under translations and rotations of the coordinate system.

(F4) If there is an open set $U \subseteq \mathbb{R}^n$ with $f_n\big|_U \equiv 0$ then $f_n \equiv 0$.

(F5) $f_n\big|_{(x_n = 0)} \in F_{n-1}$.

(F6) If $f_n\big|_h \equiv 0$ then there exists an $g_n \in F_n$ such that $f_n = l \cdot g_n$, where l is the linear function defining h.

Lemma 2.3 Then the following properties are also satisfied for $f_n \in F_n$:

(F7) If h is a hyperplane in $\mathbb{R}^n$ then $f_n\big|_h \in F_{n-1}$.

(F8) Let V be a linear subspace of $\mathbb{R}^n$ of dimension $k < n$. If a relatively open set $U \subseteq V$ exists with $f_n\big|_U \equiv 0$ then also $f_n\big|_V \equiv 0$.

Proof : (F7) follows from F(3) and (F5).

(F8) V can be defined as the intersection of $n - k$ hyperplanes : $V = \bigcap_{i=1}^{n-k} h_i$.

Successive application of (F7) yields a function $f_k = f_n\big|_V \in F_k$. Now (F4) proofs the claim. $\square$

Theorem 2.4 *The following classes of functions satisfy (F1)–(F6) : The linear functions L_n, L_n^*, real polynomials and analytic functions.*

Proof : (F1)–(F5) are obviously true. Only (F6) needs a closer look if we deal with polynomials, so let f_n be a polynomial and h a hyperplane in $\mathbb{R}^n$ (with defining linear function l) such that $f_n\big|_h \equiv 0$. From the Hilbertsche Nullstellensatz (see [vdW], page 164, formula (2)) it follows that a $q \in \mathbb{N}$ and a polynomial f_n' exist such that $f_n^q = f_n' \cdot l$. But l is prime and hence divides f_n. $\square$

3 . Certificates and Proofs

In this Section we recall Yao's definition of certificates ([Y89]), give a definition of complete proofs which slightly differs from Rabin's definition ([Rab]) and show how these definitions are related to decision trees. Then we prove some simple geometric Lemmas and show why Rabin's model is not so useful when proving lower bounds for decision trees.

Let $L = \{l_1, \ldots, l_m\} \subset L_n$ be a set of linear functions, and let $Op_L \in \Delta_>^m$ be a set of comparison operators for L. Then the pair (L, Op_L) defines the *target set* $S_L = S_{L,Op_L}$. Let $F = (F_n)_{n=1,2,3,\ldots}$ be a class of functions satisfying (F1)–(F6). This is the universe of functions which can be used at internal nodes of the decision tree.

Let $G = \{g_1, \ldots, g_k\}$ be a set of functions with either $g_i \in F_n$ or $\frac{1}{g_i} \in F_n$ (in fact, since we are only interested in the sign of g_i, we can w.l.o.g. assume that $g_i \in F_n$), and let $Op_G \in \Delta^k$ be a set of comparison operators for G. Then $S_G = S_{G,Op_G}$ is defined as before (where all g_i had been linear functions). We call the pair $Z = (G, Op_G)$ a *certificate* for (L, Op_L) if $S_G \subseteq S_L$. The size of the certificate is defined to be $|Z| := k$, i.e. the number of defining functions of G. Z is called *strict* if $Op_G = \{>\}^k$. In this case we write for short $(G,>)$ or $S_{G,>}$ instead of $(G, \{>\}^k)$ or $S_{G,\{>\}^k}$, respectively. Similarly, we call the target set *strict* if $Op_L = \{>\}^m$ and we then also write $S_{L,>}$. Analogously, we write $S_{g,=}$ for the set of zeros of a function $g \in L_n^*$. Since all functions in F_n are continous, if $S_{G,>} \neq \emptyset$ then $S_{G,>}$ is an open set and hence truly n-dimensional.

Let $0 \not\equiv Q \in F_n$ be an arbitrary function and let $Z = \{Z_1, \ldots, Z_p\}$ be a set of certificates. Z is a *complete proof* for (L, Op_L) with respect to Q if

(C1) Each Z_i is a certificate for L, i.e. $\forall x \forall i : x \in S_{Z_i} \Rightarrow x \in S_L$.

(C2) S_L is covered by $S_{Q,=}$ and the S_{Z_i}, i.e. $\forall x, Q(x) \neq 0 : (x \in S_L \Rightarrow \exists i : x \in S_{Z_i})$.

The size of Z is the maximal size of one of its certificates, i.e. $|Z| := \max_i |Z_i|$. If all Z_i are strict then we call Z a *strict complete proof.*

There is a strong correspondence between certificates and complete proofs on the one hand and decision trees on the other hand. Given a decision tree which decides membership in a set S_{L,Op_L} the set of functions evaluated along any 1-path (i.e. a path which gives the answer "is member") is a certificate for (L, Op_L). And the collection of all certificates corresponding to all 1-paths is a complete proof for (L, Op_L) w.r.t. any function Q. Hence any lower bound on the size of a certificate or complete proof is also a lower bound for decision trees.

Also, each certificate can easily be transformed into an equivalent path of a decision tree. However, the transformation of an arbitrary complete proof into a whole decision tree can result in a rather complicated and nonoptimal decision tree.

The next two Lemmas show that we can restrict our lower bound proofs to strict complete proofs. Here we differ from Rabin who only allowed nonstrict inequalities (i.e. $\geq$) in the definition of S_G and S_L.

Lemma 3.1 Let $Z = (G, >)$ be a strict certificate for (L, Op), $Op \in \Delta^m_>$ arbitrary. Then Z is also a certificate for $(L, >)$.

Proof : W.l.o.g. is $S_G \neq \emptyset$. Assume that there exists a $z \in S_G$ with $l_i(z) = 0$. But since S_G is an open set there exists an $\epsilon > 0$ such that $B_n(z, \epsilon) \subseteq S_G$ which means that also $B_n(z, \epsilon) \subseteq S_L$, a contradiction to $l_i(z) = 0$. □

Lemma 3.2 Let Z be a complete proof for (L, Op_L) w.r.t. Q. Then there exists a set of certificates Z' and a $0 \not\equiv Q' \in F_n$ such that $|Z'| \leq |Z|$ and Z' is a strict complete proof for $(L, >)$ w.r.t. Q'.

Proof : Assume that $Z = \{Z_1, \ldots, Z_p\}$ with $Z_i = (G_i, Op_i)$, $G_i = \{g_{i1}, \ldots, g_{ik}\}$ and $Op_i = \{op_{i1}, \ldots, op_{ik}\}$. We first define $Q' := Q \cdot \prod_{i,j} g_{ij}$ where the product is taken over all $g_{ij} \not\equiv 0$. Then we define Z' by

(1) Replacing all predicates "$g_{ij} \neq 0$" by "$g_{ij}^2 > 0$".

(2) Throwing away all Z_i with $op_{ij} =$ '$=$' for some j.

(3) And then defining all remaining $op'_{ij} :=$ '$>$'.

Obviously, $|Z'| \leq |Z|$ and all certificates are strict. It remains to show that Z' is a complete proof for $(L, >)$. Step (1) did not change anything except the degree of some functions which is irrelevant for us. So suppose w.l.o.g. that Z did not contain any $op_{ij} =$ '$\neq$'.

(C1) Since $S_{Z'_i} \subseteq S_{Z_i}$ for all Z'_i remaining after step (2), the Z'_i are also certificates for (L, Op_L) and hence for $(L, >)$ by Lemma 3.1.

(C2) If there is an $x \in S_{L,>}$ with $Q'(x) \neq 0$ then in particular $x \in S_L$ and $Q(x) \neq 0$. By (C2) there must be an i such that $x \in S_{Z_i}$, i.e. $\forall j$: $g_{ij}(x)\ op_{ij}\ 0$, $op_{ij} \in \{>, \geq, =\}$. From $Q'(x) \neq 0$ follows $op_{ij} \in \{>, \geq\}$ and

$g_{ij}(x) > 0$ for all j. Hence Z_i was not removed in step (2) and $Z_i^!$ is a strict certificate for x, i.e. $x \in S_{Z_i^!}$. $\square$

We remark that in the Lemma above $S_{L,>}$ might be empty, but then all certificates $Z_i^!$ would also define empty sets $S_{Z_i^!}$. In Section 5 we will prove lower bounds for strict complete proofs and then use Lemma 3.2 to extend these results to arbitrary complete proofs. In [Rab] Rabin tried just the opposite. Instead of restricting the inequalitites of L to strict inequalities (which is what we have done above) he restricted them to nonstrict inequalities ($\geq$). This enabled him to give a fairly straightforward inductive proof for his lower bound, but unfortunately his definition of a complete proof can not model decision trees correctly. We give an example in $\mathbb{R}^2$: Let $S_L = \{(x,y) \in \mathbb{R}^2 \mid x, y \geq 0\}$ and consider the decision tree in Fig. 3.1 which tests membership in S_L. Then the 1-paths in the tree are a complete proof for L according to our definition, and Lemma 3.2 shows how to transform it into a strict complete proof for $(L, >)$ to which our lower bounds of Section 5 can be applied.

However, Rabin claimed that all 1-paths in the tree are a complete proof for S_L (after certain sign changes) according to his notion of a complete proof, i.e. if all comparisons are changed to '$\geq$'. But this is wrong because then the rightmost path would become the certificate $Z_1 = \{x \geq 0, x \cdot y \geq 0\}$ which is true for $(x,y) = (0,-5)$, a point far away from S_L. Hence Z_1 would not be a valid certificate for L.

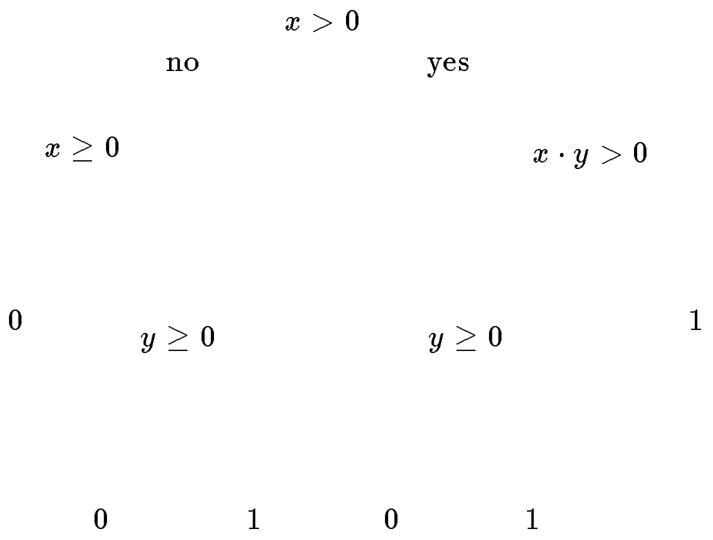

Fig. 3.1 A decision tree for $S_L = \{(x,y) \mid x, y \geq 0\}$

The problem hidden in the example above is the fact that $S_{G,\geq}$ can be a proper superset of $\overline{S_{G,Op}}$, $Op \in \Delta_{>}^{|G|}$, whereas always $S_{G,>} \subseteq S_{G,Op}^0$. It seems difficult to fix Rabin's proof only by slight changes of his definitions because the whole proof depends heavily on the fact that '$\geq$' is used everywhere instead of '$>$'. (Remark : We found this bug when we observed that the first step in his proof is to simply forget about this strange polynomial Q whereas this very polynomial Q is needed when the proof is applied to decision trees).

We close this Section with some simple geometric observations.

Lemma 3.3 Let $L = \{l_1, \ldots, l_m\}$ be a set of linear functions. Let h be a hyperplane with defining function l and let x be some point on h.

(a) Let $C \subseteq \mathbb{R}^n$ be truly n-dimensional and $g_1, \ldots, g_k$ be a set of nonzero functions from F_n. Then for all $x \in C$ and all $\epsilon > 0$ a $z \in C^0 \cap B_n(x, \epsilon)$ exists such that $g_i(z) \neq 0$ for all i, i.e. each $x \in C$ can be slightly perturbed within C^0 to avoid the zerosets of all the g_i.

(b) Let $0 \not\equiv g \in F_n$ with l not dividing g. Then for all $\epsilon > 0$ a $z \in B_n(x, \epsilon)$ exists with $l(z) \cdot g(z) > 0$.

(c) If an $\epsilon > 0$ exists such that $B_n(x, \epsilon) - h \subseteq S_{L,>}$ then $x \in S_{L,>}$.

(d) Let Z be a strict certificate for L. If an $\epsilon > 0$ exists such that $h \cap B_n(x, \epsilon) \subseteq \overline{S_Z}$ then $l_i(x) > 0$ for all $l_i \neq l$, i.e. if some small environment of x within some hyperplane does not stick outside of S_Z then x must be a point within $S_{L,>}$ or lying on its boundary if h is a bounding hyperplane of $S_{L,>}$.

Proof : (a) $C^0 \cap B_n(x, \epsilon)$ is an open set in $\mathbb{R}^n$. Then (F4) proves the claim.

(b) From (F8) and (F6) it follows that there must be a $y \in h \cap B_n(x, \frac{\epsilon}{2})$ with $g(y) \neq 0$. But then $sgn(g)$ is constant in $B_n(y, \gamma)$ for a $\frac{\epsilon}{2} > \gamma > 0$. Since h divides $B_n(y, \gamma)$ into two halves with different signs of l there exists a $z \in B_n(y, \gamma) \subseteq B_n(x, \epsilon)$ with $l(z) \cdot g(z) > 0$.

(c) Obvious.

(d) $x \in \overline{S_Z} \subseteq \overline{S_L}$ implies $l_i(x) \geq 0$ for all i. If $l_i(x) = 0$ and $l_i \neq l$ for an i then l_i divides the $(n\text{–}1)$–dimensional ball $h \cap B_n(x, \epsilon)$ into two halves ($h \cap l_i$ is a $(n\text{–}2)$–dimensional hyperplane in the $(n\text{–}1)$–dimensional space h), and l_i is positive in one of the halves and negative in the other. But this contradicts $h \cap B_n(x, \epsilon) \subseteq \overline{S_Z} \subseteq \overline{S_L}$. □

4. Yao's Theorem (Improved)

In [Y89] A.C. Yao showed that the size of a certificate which uses only linear functions (i.e. from L_n) is bounded from below by the number of linearly independent functions in the target set L. He raised the question whether this result generalizes to certificates which use multilinear functions (from L_n^*). In this Section we will prove that it does indeed generalize by extracting and exploiting the main idea behind Yao's proof (which is carefully hidden in [Y89]).

Theorem 4.1 *Let (G, Op_G), $G \subset L_n^*$, be a certificate for some (L, Op_L), $L \subset L_n$. If $S_G \neq \emptyset$ then $|Z| \geq n - rank(L)$. Hence any accepting path in a decision tree for S_{L,Op_L} with functions from L_n^* must have length at least $n - rank(L)$.*

Proof : The Theorem follows immediately from the next Theorem and Lemma 2.1. □

Theorem 4.2 *Let $G = \{g_1, \ldots, g_k\} \subset L_n^*$ and $Op \in \Delta^k$. If $S_{G,Op} \neq \emptyset$ then there exists a linear variety $V \subseteq \overline{S_{G,Op}}$ of dimension $n - k$.*

A similar Theorem (for L_n only) was stated in [Y89] but there was no Lemma 2.1 so the proofs became rather complicated. The proof of Theorem 4.2 is obtained by induction on k but one has to be very careful about some subtle difficulties (the proof is quite trivial in the case of linear functions only). The inductive step is based on the following reduction scheme.

Reduction 4.3 Let $G = \{g_1, \ldots, g_k\} \subset L_n^*$ and $Op \in \Delta^k$. Let h be a hyperplane in $\mathbb{R}^n$. Then we define $G' \subset L_{n-1}^*$ and Op_i' by

$$g_i' := \begin{cases} g_i\big|_h & \text{if } g_i\big|_h \not\equiv 0; \\ \text{not existing} & \text{if } g_i\big|_h \equiv 0; \end{cases}$$

$$op_i' := op_i \quad \text{if } g_i' \text{ exists.}$$

Then obviously $|G'| \leq |G|$ and $g_i'(x) = g_i(x)$ for all $x \in h$. But we also need $\emptyset \neq S_{G'} \subseteq \overline{S_G}$, which is not necessarily true. In fact, this may fail for three reasons. Firstly, if $h \cap \overline{S_G} = \emptyset$ then $S_{G'} = \emptyset$. Secondly, if $g_i = l^2 \cdot \hat{g}_i$ where l is the defining function of h then we discard $\underline{g_i}$ in the reduction step; but if $op_i =$ '$>$' this can add points to $S_{G'}$ which may not be in $\overline{S_G}$, namely all points $z \in h$ with $\hat{g}_i(z) < 0$ which are not excluded by other constraints. And thirdly, if l is a common factor of several of the g_i (which are all missing in G') then $S_{G'}$ can also be larger than $\overline{S_G}$. For example, if $Z = (\{x, x \cdot y, x + 5\}, >)$ and h is the hyperplane $(x = 0)$ then the reduced certificate is just $(\{x + 5\}, >)$, which is true everywhere on h whereas only the upper half of h bounds S_Z.

In the next seven Lemmas we will show how to solve these problems by transforming an arbitrary certificate into a certificate of at most the same size which does not cause any of these problems. First we will show that we can w.l.o.g. assume that the certificates do not use any '$\neq$'-comparisons.

Lemma 4.4 Let $G = \{g_1, \ldots, g_k\} \subset L_n^*$ and $Op \in \Delta^k$. If $S_{G,Op} \neq \emptyset$ then there exists an Op', '$\neq$' $\notin Op'$, such that $\emptyset \neq S_{G,Op'} \subseteq S_{G,Op}$.

Proof : Let $z \in S_{G,Op}$ be arbitrary. Now we define

$$op_i' := \begin{cases} op_i, & \text{if } op_i \in \{>, \geq, =\}; \\ >, & \text{if } op_i = \text{'}\neq\text{' and } g_i(z) > 0; \\ <, & \text{if } op_i = \text{'}\neq\text{' and } g_i(z) < 0. \end{cases}$$

Then $z \in S_{G,Op'} \subseteq S_{G,Op}$. □

So we can from now on w.l.o.g. assume that $\Delta = \{>, \geq, =\}$. The next Lemma shows that we can sometimes restrict Δ even further to $\{>\}$, i.e. we have to consider only strict certificates.

Lemma 4.5 Let $G = \{g_1, \ldots, g_k\} \subset L_n^*$ and $Op \in \Delta^k$. If $S^0_{G,Op} \neq \emptyset$ then $\emptyset \neq S_{G,>} \subseteq S_{G,Op}$.

Proof : $S^0_{G,Op} \neq \emptyset$ implies the existence of an $x \in S_{G,Op}$ and an $\epsilon > 0$ such that $B_n(x, \epsilon) \subseteq S_{G,Op}$. Then Lemma 3.3.(a) guarantees the existence of a $z \in B_n(x, \epsilon)$ with $g_i(z) \neq 0$ for all i. But this means $g_i(z) > 0$ for all i. Hence $z \in S_{G,>}$. □

The next two Lemmas show how to solve the third problem for strict certificates, i.e. if the hyperplane $l(x) = 0$ is used in the reduction and l is a common factor of several of the g_i.

Lemma 4.6 Let $F = \{l \cdot f_1, l \cdot f_2\}$ and $G = \{l \cdot f_1, f_1 \cdot f_2\}$ with $f_1, f_2 \in F_n$ and $l \in L_n$. Then $S_{F,>} = S_{G,>}$, i.e. we can replace the "bad" certificate F with two occurrences of l by the "good" certificate G with only one occurrence of l (at least if l is not a factor of f_2).

Proof :
$$\begin{aligned} x \in S_{F,>} &\iff l(x) \cdot f_1(x) > 0 \quad \wedge \quad l(x) \cdot f_2(x) > 0 \\ &\iff sgn(l(x)) = sgn(f_1(x)) = sgn(f_2(x)) \\ &\iff l(x) \cdot f_1(x) > 0 \quad \wedge \quad f_1(x) \cdot f_2(x) > 0 \\ &\iff x \in S_{G,>} \, . \end{aligned}$$
□

Lemma 4.7 Let $G = \{g_1, \ldots, g_k\} \subset L_n^*$ with $l \in L_n$ dividing g_1. Then a $G' = \{g_1', \ldots, g_k'\} \subset L_n^*$ with $S_{G,>} = S_{G',>}$ exists such that l is at most a squared factor of g_1' and does not divide any of the other g_i', $i \geq 2$.

Proof : Assume $g_i = l^{a_i} \cdot \hat{g}_i$, $a_i \in \mathbb{N}_0$, for all i. We proceed in two steps. First we show that the multiplicities of the factor l can be made small, i.e. 1 or 2 for g_1 and 0 or 1 for $g_2, \ldots, g_k$. Then we show how to eliminate the factor l from all g_i, $i \geq 2$.

(1) We define

$$g_1'' := l^{b_1} \cdot \hat{g}_1 \quad \text{with} \quad b_1 = \begin{cases} 2, & \text{if } a_1 \text{ is even;} \\ 1, & \text{if } a_1 \text{ is odd;} \end{cases}$$

and for $i = 2, \ldots, k$

$$g_i'' := l^{b_i} \cdot \hat{g}_i \quad \text{with} \quad b_i = \begin{cases} 0, & \text{if } a_i \text{ is even;} \\ 1, & \text{if } a_i \text{ is odd.} \end{cases}$$

Then $|G| = |G''|$ and $S_{G,>} = S_{G'',>}$ which can be seen as follows (observe $l^2(x) \geq 0$ for all $x \in \mathbb{R}^n$) :

$$\begin{aligned} x \in S_{G,>} &\iff \forall i \,:\, g_i(x) = l^{a_i} \cdot \hat{g}_i(x) > 0 \\ &\iff l(x) \neq 0 \quad \wedge \quad \forall i \,:\, l^{b_i}(x) \cdot \hat{g}_i(x) > 0 \\ &\iff g_1''(x) > 0 \quad \wedge \quad \forall i \geq 2 \,:\, g_i''(x) > 0 \\ &\iff x \in S_{G'',>} \end{aligned}$$

(2) If $b_i = 0$ for $i \geq 2$ then we can choose $G' := G''$. Otherwise $b_j = 1$ for a $j \geq 2$. If $b_1 = 2$ then we must exchange g_1'' and g_j'', i.e. we define $g_1'' := g_j''$ and $g_j'' := \hat{g}_1$. Now the multiplicity of the factor l in all g_i'' is at most 1 and still $S_{G,>} = S_{G'',>}$. But then we can apply Lemma 4.6 to all pairs (g_1'', g_i'') where l divides g_i'', $i \geq 2$. This gives us our G'. □

The G' constructed in the previous Lemma may still have $g_1' = l^2 \cdot \hat{g}_1$ which was our second problem. But the next Lemma shows that we can neglect squared factors without increasing the set of solutions to much.

Lemma 4.8 Let $l \in L_n$ and $G = \{g_1, \ldots, g_k\} \subset L_n^*$ with $g_1 = l^2 \cdot \hat{g}_1$. Define G' by $g_1' := \hat{g}_1$ and $g_i' := g_i$, $i \geq 2$. Then $S_{G,>} \subseteq S_{G',>} \subseteq \overline{S_{G,>}}$.

Proof : The first inclusion is trivial. If $S_{G',>} = \emptyset$ then nothing is to show. So let $x \in S_{G',>}$ be arbitrary. Since $S_{G',>}$ is truly n-dimensional, Lemma 3.3.(a) implies the existence of a sequence $(x_i)_{i=1,2,\ldots}$ of points $x_i \in S_{G',>}$ with $\lim_{i\to\infty} x_i = x$ and $l(x_i) \neq 0$ for all i. But this implies $x_i \in S_{G,>}$ for all i and hence $x \in \overline{S_{G,>}}$. □

Cor. 4.9 Let $G = \{g_1, \ldots, g_k\} \subset L_n^*$ and $l \in L_n$ dividing one of the g_i. Then a $G' \subset L_n^*$ exists such that $|G'| = |G|$, $S_{G,>} \subseteq S_{G',>} \subseteq \overline{S_{G,>}}$ and l is unique in G, i.e. divides exactly one of the g_i', and l^2 does not divide any of the g_i.

Proof : Lemmas 4.7 and 4.8. □

In this case we can prove that Reduction 4.3 works properly.

Lemma 4.10 Let h be a hyperplane with defining function l. Let $G = \{g_1, \dots, g_k\} \subset L_n^*$ such that l does not divide any of the g_i or is unique in G. If G' is constructed by Reduction 4.3, applied to G and h, then $S_{G',>} \subseteq \overline{S_{G,>}}$.

Proof : If l does not divide any of the g_i then $G' = G\Big|_{h}$. So assume $g_1 = l \cdot \hat{g}_1$, i.e. g_1' does not exist. Let $x \in S_{G',>} \subseteq h$ be arbitrary. Then an $\epsilon > 0$ exists such that $g_i(z) > 0$ for all $i \geq 2$ and $z \in B_n(x, \epsilon)$. Let ϵ_j be a sequence of numbers with $\epsilon \geq \epsilon_j > 0$ and $\lim_{j \to \infty} \epsilon_j = 0$. Then, by Lemma 3.3.(b), $\hat{x}_j \in B_n(x, \epsilon_j)$ exist with $g_1(\hat{x}_j) > 0$. Hence $x \in \overline{S_{G,>}}$. □

Proof of Theorem 4.2 :

Because of Lemma 4.4 we can assume $Op \in \{>, \geq, =\}$. The proof is obtained by induction on $k = |G|$.

<u>$k = 1$:</u> Let $G = \{g_1\}$. Now we have to consider two cases. Either g_1 is a product of squared factors, i.e. $g_1 = l^2 \cdot \hat{g}_1^2$ with $l \in L_n$ and $\hat{g}_1 \in L_n^*$. We define $V := S_{l,=}$. Then trivially $dim\, V = n - 1$ and $V \subseteq \overline{S_{G,Op}}$.

Or g_1 has a linear factor l of odd multiplicity. Then w.l.o.g. $g_1 = l \cdot \hat{g}_1$ with $\hat{g}_1 \in L_n^*$ and l does not divide $\hat{g}_1$ ($sgn(l^3 \cdot \hat{g}_1) = sgn(l \cdot \hat{g}_1)$). We define $V := S_{l,=}$. If $op_1 \in \{\geq, =\}$ then trivially $V \subseteq S_{G,Op}$. Otherwise, Lemma 3.3.(b) implies that each $x \in V$ is also in $\overline{S_{G,Op}}$.

<u>$k > 1$:</u> Let $G = \{g_1, \dots, g_k\}$, $g_i \in L_n^*$. Now we have to consider two cases.

$S^0_{G,Op} = \emptyset$: Let $z \in S_{G,Op}$ be arbitrary. Since $S^0_{G,Op} = \emptyset$ there must be an $l \in L_n$ such that $l(z) = 0$ and w.l.o.g. $g_1 = l \cdot \hat{g}_1$. Let h be the hyperplane defined by l and G' constructed by Reduction 4.3, applied to G and h. Then obviously $z \in S_{G',Op'}$ and $|G'| < k$. Hence, by induction hyptothesis, a linear variety $V \subseteq \overline{S_{G',Op'}}$ of dimension at least $(n-1) - (k-1) = n - k$ must exist.

Then for all $x \in V$ a sequence $(x_j)_{j=1,2,\dots}$ exists with $\lim_{j \to \infty} x_j = x$ and $x_j \in S_{G',Op'}$ for all j, i.e. $g_i'(x_j)\ op_i'\ 0$ for all i such that $g_i' \in G'$ exists. All g_i which were discarded in the reduction must contain the factor l; hence $g_i(x_j) = 0$ for all these i and all j. Furthermore, $g_i(z) = 0$ for all these i and, since $z \in S_{G,Op}$, all these op_i must be from $\{\geq, =\}$. But then $x_j \in S_{G,Op}$ for all j (because $g_i' = g_i\Big|_{h}$ and $op_i' = op_i$ if it exists) and hence $x \in \overline{S_{G,Op}}$.

$S^0_{G,Op} \neq \emptyset$: Then we can assume $Op = \{>\}^k$ by Lemma 4.5. $S_{G,>}$ is a subset of $\mathbb{R}^n$ which is bounded by hyperplanes whose defining functions are all among the linear factors of the g_i. Let h be such a bounding hyperplane with defining function l and assume w.l.o.g. that $g_1 = l \cdot \hat{g}_1$. By Cor. 4.9, we can assume that l is unique in G.

Let C be the $(n-1)$-face of $\overline{S_{G,>}}$ which is supported by h and let z be an arbitrary point in C^0. Then $g_1(z) = 0$ and $g_i(z) \neq 0$ for all $i \geq 2$.

Since all g_i have a constant sign in $B_n(z, \epsilon)$ for some small $\epsilon > 0$ (the g_i are continuous) and $z \in \overline{S_{G,>}}$, we even know that $g_i(z) > 0$ for all $i \geq 2$.

Reduction 4.3, applied to G and h, gives us a G' with $g_i'(z) = g_i(z) > 0$ for all $i \geq 2$ (and g_1' not existent). Hence $z \in S_{G',>}$, i.e. $S_{G',>} \neq \emptyset$. And also $S_{G',>} \subseteq \overline{S_{G,>}}$ by Lemma 4.10.

Since $|G'| < k$ we have, by induction hypothesis, a linear variety $V \subseteq \overline{S_{G',>}}$ of dimension $(n-1)-(k-1) = n-k$. Therefore we know that for all $x \in V$ a sequence $(x_j)_{j=1,2,\ldots}$ exists with $\lim_{j\to\infty} x_j = x$ and $x_j \in S_{G',>}$ for all j, i.e. $g_i(x_j) > 0$ for all $i \geq 2$. But then, for all j, $\epsilon_j > 0$ exist such that $g_i(z_j) > 0$ for all $z_j \in B_n(x_j, \epsilon_j)$ and $i \geq 2$. We can w.l.o.g. assume that $\lim_{j\to\infty} \epsilon_j = 0$. Since $x_j \in h$, by Lemma 3.3.(b) $\hat{x}_j \in B_n(x_j, \epsilon_j)$ exist with $g_1(\hat{x}_j) = l(\hat{x}_j) \cdot \hat{g}_1(\hat{x}_j) > 0$ and also $g_i(\hat{x}_j) > 0$ for $i \geq 2$. Hence $x \in \overline{S_{G,>}}$. □

We now apply Theorem 4.1 to the problem of finding the k largest elements, including their individual rankings, of n real numbers. Let $W_k(n)$ denote the worst-case complexity in the decision tree model when arbitrary functions $l \in L_n^*$ are allowed. The following Theorem generalizes Yao's Theorem ([Y89], Theorem 1) to arbitrary functions L_n^* (instead of L_n^2).

Theorem 4.11 *$W_k(n) \geq n - k + \sum_{1\leq i\leq k-1} \log(n-i+1)$ for all $n > k \geq 2$.*

Proof : (Sketch, see [Y89] for complete proof) Let $G \subset L_n^*$ be a certificate for $S_L = \{x_1 - x_i \mid 2 \leq i \leq n\}$. Then $rank(L) = n-1$ and hence $|G| \geq n-1$ by Theorem 4.1. It follows that any decision tree which finds the maximum of n numbers must have at least 2^{n-1} leaves (because each path is a certificate for maximum). Now partition the leaves of a decision tree T which finds the k largest numbers into $\prod_{1\leq i\leq k-1}(n-i+1)$ disjoint classes, each class containing the leaves which output $x_{i_1}, \ldots, x_{i_{k-1}}$ as the $k-1$ largest numbers for some fixed $i_1, \ldots, i_{k-1}$. Then each class induces a subtree of T which finds the maximum of $n-k$ numbers and hence has at least 2^{n-k} leaves. □

5. Rabin's Theorem (Corrected)

In [Rab] M. Rabin showed that the size of a complete proof for a sign-independent linear target set L is bounded from below by $|L|$ if arbitrary polynomials or even analytic functions are allowed in the complete proof. But we have seen in Section 3 that his notion of complete proofs seems to be useless in the context of decision trees. In this Section we will prove an analogous result with respect to our definition of a complete proof. In fact, our result is stronger than Rabin's because we bound the size of the complete proof by $n - rank(L)$, and more general because we allow a wider class of functions F_n.

Throughout this Section, $F = (F_n)_{n=1,2,\ldots}$ will be some set satisfying (F1)–(F6) (but it is always possible to think of F_n as real polynomials in n variables). All certificates will use functions from F_n.

Theorem 5.1 *Let Z be a complete proof for some (L, Op_L), $L \subset L_n$, w.r.t. some $Q \in F_n$. If $S^0_{L,Op_L} \neq \emptyset$ then $|Z| \geq n - rank(L)$. Hence any decision tree for S_{L,Op_L} with functions from F_n must have depth at least $n - rank(L)$.*

We remark that this bound does not necessarily hold if $S^0_{L,Op_L} = \emptyset$. For example, $l(x) \geq 0$ and $-l(x) \geq 0$ both together are equivalent to $l(x) = 0$; hence any linear subspace V of $\mathbb{R}^n$ of dimension $k < n$ can be achieved as target set using a set L of $2(n-k)$ linear functions with $rank(L) = k$. But there is a trivial complete proof Z for V which consists of only one certificate, and this certificate consists of only one quadratic polynomial (let $g = x_1^2 + \cdots + x_{n-k}^2$; then $S_{g,=}$ is equal to $\mathbb{R}^k$), i.e. $|Z| = 1$.

Since $S^0_{L,Op_L} \neq \emptyset$, we can w.l.o.g. assume that $S_{L,>} \neq \emptyset$ and Z is a strict complete proof for $(L, >)$ (Lemma 3.2). Hence we can also assume that all functions g_{ij} used in Z are $\not\equiv 0$. This makes the proofs in this Section a little bit easier than they have been in Section 4. The proof of the Theorem will be obtained by induction on $|Z|$. The inductive step is based on the following reduction scheme which is an extension of Reduction 4.3.

Reduction 5.2 Let $Z = \{Z_1, \ldots, Z_p\}$ be a strict complete proof for L w.r.t. Q, where $Z_i = (G_i, >)$ with $G_i = \{g_{i1}, \ldots, g_{ik}\} \subset F_n$. Let h be a hyperplane. Then we define a set Z' of $(n-1)$-dimensional certificates and a new $(n-1)$-dimensional target set L' by

(1) $Q' := \prod g_{ij}\big|_h$ where the product is taken over all g_{ij} which are not dividable by l.

(2) Z'_i is the result of Reduction 4.3 applied to Z_i and h; if $S_{Z'_i} = \emptyset$ then Z'_i is discarded.

(3) L' is the result of Reduction 4.3 applied to L and h.

Then obviously $|Z'| \leq |Z|$ and $g'_{ij}(x) = g_{ij}(x)$ and $l'_i(x) = l_i(x)$ for all $x \in h$ and all i, j. If h is defined by one of the linear functions $l_i \in L$ then we even know that $|Z'| \leq |Z| - 1$ because each Z_i is either shortened by Reduction 4.3 or it completely vanishes (if it does not contain the factor l) as the next Lemma shows.

Lemma 5.3 Let Z_i be a strict certificate for L and let $l \in L$ define hyperplane h. If l does not divide any of the functions of Z_i then Reduction 4.3, applied to Z_i and h, yields a certificate Z'_i with $S_{Z'_i,>} = \emptyset$.

Proof : $Z'_i = Z_i\big|_h$ because l does not divide any of the functions used in Z_i. Assume that an $x \in S_{Z'_i,>} \subseteq h$ exists. Then also $x \in S_{Z_i,>}$ and there exists an $\epsilon > 0$ such that $B_n(x,\epsilon) \subseteq S_{Z_i,>}$. But this contradicts the fact that not both sides of h can belong to S_L. □

It remains to show that Z' is a strict complete proof for L' w.r.t. Q'. Unfortunately, this is not always true as we have seen in the last Section. So, once again, we need some transformations before we can prove that Reduction 5.2 works (Lemma 5.7). Similarly to Lemma 4.8 we first show that in a strict complete proof squared linear factors are not important.

Lemma 5.4 Let $Z = \{Z_1, \ldots, Z_p\}$ be a strict complete proof for L w.r.t. Q and $g_{ij} = l^2 \cdot \hat{g}_{ij}$ a function used in certificate Z_i with $l \in L_n$. We define another set of certificates Z' which only differs in Z_i, namely we define $g'_{ij} := \hat{g}_{ij}$. Then Z' is also a strict complete proof for L w.r.t. Q, and obviously $|Z'| \leq |Z|$.

Proof : We must show (C1) and (C2) for Z'.

(C1) We must show that Z'_i is still a certificate for L. Obviously $S_{Z_i} \subseteq S_{Z'_i}$. Let h be the hyperplane defined by l. If $x \in h \cap S_{Z'_i}$ then an $\epsilon > 0$ exists such that $B_n(x,\epsilon) \subseteq S_{Z'_i}$. But then $B_n(x,\epsilon) - h \subseteq S_{Z_i} \subseteq S_L$ and hence $x \in S_L$ by Lemma 3.3.(c).

(C2) $x \in S_{Z_i}$ implies $x \in S_{Z'_i}$. Therefore S_L is still covered by the certificates of Z'. □

The next Lemma is quite fundamental for our inductive proof because it shows that each defining function of a bounding hyperplane of S_L must divide at least one of the functions used in any strict complete proof for L.

Lemma 5.5 Let Z be a strict complete proof for L w.r.t. Q and $S_L \neq \emptyset$. Let h be a bounding hyperplane of S_L with defining function l. Then there is a function g used in Z such that l divides g.

Proof : Let $0 \not\equiv f := \prod_{g \in Z} g$ and assume $f\big|_h \not\equiv 0$. Then there exists an $x \in h \cap \overline{S_L}$ with $f(x) \neq 0$ (by (F7) and (F4) and because h bounds S_L). But then an $\epsilon > 0$ exists such that all g in Z have a constant sign in $B_n(x,\epsilon)$ which means that $B_n(x,\epsilon) \subseteq S_L$ by (C1), a contradiction. Hence $f\big|_h \equiv 0$ and l divides f by (F6). But l is prime and hence divides one of the g in Z. □

Cor. 5.6 Let Z be a strict complete proof for L w.r.t. Q and $S_L \neq \emptyset$. Let h be a bounding hyperplane of S_L with defining function l. Then there exists a strict complete proof Z' for L w.r.t. Q with $|Z' \leq |Z|$ and l divides some function g used in Z'

but l^2 does not.

Proof : First eliminate in Z all linear factors of multiplicity 2 or more using Lemma 5.4, then apply Lemma 5.5. □

In this case we can prove that Reduction 5.2 works properly.

Lemma 5.7 Let Z be a strict complete proof for L w.r.t. Q and $S_L \neq \emptyset$. Let h be a bounding hyperplane of S_L with defining function l. If l is unique in each certificate where it appears as a linear factor then Reduction 5.2, applied with hyperplane h, yields a strict complete proof Z' for L' w.r.t. Q'. Furthermore $|Z'| \leq |Z| - 1$.

Proof : We must show (C1) and (C2) for Z'. Let $Z = \{Z_1, \ldots, Z_p\}$ with $Z_i = (G_i, >)$ and $G_i = \{g_{i1}, \ldots, g_{ik}\}$. By Lemma 5.3, l divides w.l.o.g. g_{i1}; hence $Z_i^! = (G_i^!, >)$ with $G_i^! = \{g_{i2}\big|_h, \ldots, g_{ik}\big|_h\}$ for all i. Furthermore, w.l.o.g. $l = l_1$ where $L = \{l_1, \ldots, l_m\}$, and hence $L' = \{l_2\big|_h, \ldots, l_m\big|_h\}$.

(C1) Let $x \in S_{Z_i^!}$ for an i. Then an $\epsilon > 0$ exists such that $h \cap B_n(x, \epsilon) \subseteq S_{Z_i^!} \subseteq \overline{S_{Z_i}}$ by Lemma 4.10. But then $l_i(x) \neq 0$ for $i \geq 2$ by Lemma 3.3.(d) and, since $x \in \overline{S_L}$, even $l_i(x) > 0$. Hence $x \in S_{L'}$.

(C2) Let $x \in S_{L'} \subseteq \overline{S_L}$ with $Q'(x) \neq 0$. There exists a sequence $(x_s)_{s=1,2,\ldots}$ with $x_s \in S_L$, $Q(x_s) \neq 0$ and $\lim_{s\to\infty} x_s = x$. For each x_s there is an index i_s such that $x_s \in S_{Z_{i_s}}$. Since we only have a finite number of certificates, one index must be infinitely often in the sequence $(i_s)_{s=1,2,\ldots}$. Let i be this index.

Therefore we have a subsequence $(y_s)_{s=1,2,\ldots}$ of $(x_s)_{s=1,2,\ldots}$ with $y_s \in S_{Z_i}$ and $\lim_{s\to\infty} y_s = x$. Hence, for all s, $g_{ij}(x_s) > 0$ for all j and therefore $g_{ij}(x) \geq 0$. Since $Q'(x) \neq 0$ we even have $g_{ij}(x) > 0$ for $j \geq 2$.

Assume $g_{i1}(x) > 0$. Then l does not divide any of the g_{ij} and hence, by Lemma 5.3, $S_{Z_i^!} = \emptyset$, a contradiction. Therefore $g_{i1}(x) = 0$ and l divides g_{i1} (otherwise, g_{i1} would be a factor of Q' and then $Q'(x) = 0$, a contradiction). Therefore $Z_i^!$ exists and $x \in S_{Z_i^!}$.

$|Z'| \leq |Z| - 1$ follows directly from Lemma 5.3. □

Proof of Theorem 5.1 :

As mentioned before, we can assume that Z is a strict complete proof for $(L, >)$ and $S_{L,>} \neq \emptyset$ (Lemma 3.2). By Lemma 2.1.(c) we know that there is a hyperplane h in L which bounds S_L and which contributes to a $rank(L)$-face in $\overline{S_L}$. Let l be the defining function of h. By Cor. 5.6 we can assume that l is a linear factor of some of the functions used in Z but l^2 is not. We can even further assume that l is unique in each certificate of Z where it appears as a linear factor (Lemma 4.7).

Now we can apply Reduction 5.2 to Z, L and h and know by Lemma 5.7 that Z' is a strict complete proof for L' w.r.t. Q' with $|Z'| \leq |Z|-1$. Furthermore, $rank(L') = rank(L)$. By induction hypothesis, $|Z'| \geq (n-1) - rank(L')$ and therefore $|Z| \geq 1 + |Z'| \geq n - rank(L)$. □

6 . Conclusions

The proofs in this paper are mainly based on two techniques. One technique is to examine the number of free dimensions of the target set and the set of solutions for a certificate (Theorem 4.2); this seems to be a new approach to derive lower bounds for nonlinear decision trees. The other technique is not to stick to the given decision tree but to transform it into another decision tree with nicer properties and of at most the same depth. Although this technique is not new, it is not widely used. There is one paper by Ramanan ([Ram]) proposing a similar approach by introducing artificial components thus increasing the quality of classical lower bounds which are based on counting the number of connected components.

The proof of the generalization of Yao's Theorem is quite straightforward when our dimension-technique is used, even if some details turn out to be rather tricky. The Theorem can not be further generalized by allowing arbitrary polynomials in the certificates because two quadratic polynomials can have an arbitrarily small solution set which can be contained in any target set. But [Y89] mentions a few more interesting problems. One of them is the question whether lower bound proofs for "simple" combinatorial problems (such as finding the k largest numbers) can always be carried out purely combinatorially, i.e. without the detour of geometric arguments.

At the first glance it seems to be surprising that the methods used in the proof of Yao's Theorem can also be used to prove Rabin's Theorem. But a closer look at our proof of Rabin's Theorem shows that in fact we inductively prove the existence of a certificate with a set of solutions which is bordered by a $rank(L)$-face of $Arr(H)$ and which has the hyperplanes defining this $rank(L)$-face as linear factors (unfortunately, things are more complicated; what we really prove is that there *could* be such a certificate). But this is very similar to Yao's Theorem.

Our proof of Rabin's Theorem is actually also a generalization. Firstly, because our notion of complete proofs allows arbitrary comparison operators instead of only '$\geq$' which makes it equivalent to decision trees (see Section 3). And secondly, because our proof applies to a wider class of functions than only polynomials or analytic functions. In fact, Rabin claimed that his Theorem may be not true if the condition "analytic functions" is relaxed and he gave an example of a decision tree using nonanalytic functions which computes the maximum of n numbers with only $\log n$ comparisons. But our proof (and actually also Rabin's original proof) shows that the reason why the Theorem can not be applied to these functions is not that they are nonanalytic but mainly the fact that these functions are not defined on the hyperplanes bounding the target set S_L. And in this case the inductive step which restricts everything to one of these bounding hyperplanes can not work. This leads us to the following observation : We do not really require the functions used in a decision tree to satisfy (F1)–(F6) everywhere in $\mathbb{R}^n$; our proof only requires them to satisfy these properties in a small environment around the $rank(L)$-face of S_L which is used in the inductive step. This is similar to the method of focussing on some convex set $C \subseteq \mathbb{R}^n$ in [Rab].

The main open problem now seems to be to generalize Rabin's Theorem to arbitrary target sets, i.e. which are not defined by linear inequalities but by arbitrary polynomial inequalities. This would not be too difficult if we considered complex polynomials instead of real polynomials. Unfortunately, real algebraic varieties are much more difficult to handle than complex varieties. A first step in this direction was taken by Jaromczyk ([Ja]) who solved all difficulties arising from real algebraic varieties by defining them away, i.e. imposing heavy restrictions onto the functions used. Even worse, it is nontrivial to verify for a given problem that his Theorems can be applied.

References

[Be] M. Ben-Or
"Lower bounds for algebraic computation trees"
Proc. 15th ACM STOC 1983, 80–86

[BLY] A. Björner, L. Lovász, A.C. Yao
"Linear decision trees : volume estimates and topological bounds"
Proc. 24th ACM STOC 1992, 170–177

[DL] D.P. Dobkin, R.J Lipton
"On the complexity of computations under varying sets of primitives"
Journal of Computer System Sciences **18** (1979), 86–91
and Automata Theory and Formal Languages, *Lecture Notes in Computer Science*, Vol. 33, Springer 1975, 110–117

[Ed] H. Edelsbrunner
"Algorithms in Combinatorial Geometry"
Springer Verlag, Heidelberg, 1987

[FG] F. Fussenegger, H.N Gabow
"A counting approach to lower bounds for selection problems"
Journal of the ACM **26** (1979), 227–238
and *Proc. 17th IEEE FOCS* 1976, 178–182

[Ja] J.W. Jaromczyk
"Lower bounds for problems defined by polynomial inequalities"
Symposium on Foundations of Computing Theory, *Lecture Notes in Computer Science*, Vol. 117, Springer 1981, 165–172

[MadH] F. Meyer auf der Heide
"A polynomial linear search algorithm for the n-dimensional knapsack problem"
Journal of the ACM **31/3** (1984), 668–667

[MMS] U. Manber, S. Moran, M. Snir
"Applications of Ramsey's theorem to decision tree complexity"
Journal of the ACM **32/4** (1985), 938–949

[MST] Y. Mansour, B. Schieber, P. Tiwari
"Lower bounds for integer greatest common divisor computations"
Proc. 29th IEEE FOCS 1988, 54–63

[MT] U. Manber, M. Tompa
"Probabilistic, nondeterministic and alternating decision trees"
Journal of the ACM **32** (1985), 720–732
and *Proc. 14th ACM STOC* 1982, 234–244

[OD] C. O'Dúnlaing
"A tight lower bound for he complexity of path-planning for a disc"
Information Processing Letters **28/4** (1988), 165–170

[PS] W.J. Paul, J. Simon
"Decision trees and random access machines"
Symposium über Logik und Algorithmik, Zürich, 1980
Monographic 30, L'Enseignement Mathematique, Logic et Algorithmic, Genf 1982, 331–340

[Rab] M. Rabin
"Proving simultaneous positivity of linear forms"
Journal on Computer System Sciences **6** (1972), 639–650

[Ram] P. Ramanan
"Obtaining lower bounds using artificial components"
Information Processing Letters **24/4** (1987), 243–246

[RY] R. Rivest, A.C. Yao
"On the polyhedral decision problem"
SIAM Journal on Computing **9** (1980), 343–347

[Sn82] M. Snir
"Comparisons between linear functions can help"
Journal on Theoretical Computer Science **19** (1982), 321–330

[Sn85] M. Snir
"Lower bounds on probabilistic linear decision trees"
Journal on Theoretical Computer Science **38** (1985), 69–82

[St] V. Strassen
"The computational complexity of continued fractions"
SIAM Journal on Computing **12** (1983), 1–27

[vdW] B.L. van der Waerden
"Algebra II"
Springer Verlag, Heidelberg, 5. Auflage, 1967

[Y89] A.C. Yao
"On selecting the k largest with median tests"
Algorithmica **4/2** (1989), 293–300

[Y92] A.C. Yao
"Algebraic decision trees and Euler characteristics"
Proc. 33th IEEE FOCS 1992, 268–277