Randomization and the Computational Power of Analytic and Algebraic Decision Trees

Dima Grigoriev * Marek Karpinski† Roman Smolensky ‡

Abstract

We introduce a new powerful method for proving *lower bounds* on *randomized* and *deterministic* analytic decision trees, and give direct applications of our results towards some concrete geometric problems. We design also *randomized* algebraic decision trees for recognizing the *positive octant* in $\mathbb{R}^n$ or computing MAX in $\mathbb{R}^{n+1}$ in depth $\log^{0(1)} n$. Both problems are known to have linear lower lower bounds for the depth of any deterministic analytic decision tree recognizing them. The main *new* (and *unifying*) proof idea of the paper is in the reduction technique of the signs of *testing* functions in a decision tree to the signs of their *leading terms* at the specially chosen points. This allows us to reduce the complexity of a *decision* tree to the complexity of a certain *boolean* circuit.

*Departments of Computer Science and Mathematics, Penn State University, University Park, PA 16802. Supported in part by NSF grant CCR-9424358. E-mail: `dima@cse.psu.edu`

†Department of Computer Science, University of Bonn, 53117 Bonn, and International Computer Science Institute, Berkeley. Supported in part by DFG Grant KA673/4-1, by the ESPRIT BR Grants 7097 and EC-US 030, and by the Volkswagen-Stiftung. E-mail: `marek@cs.bonn.edu`

‡Department of Computer Science, University of Bonn, 53117 Bonn. Roman Smolensky died on 19 October, 1995 in New York. The main results of the paper were obtained in the Summer of 1995. This paper is also dedicated to the memory of Roman by the other co-authors.

1 Introduction

The problem of obtaining complexity lower bounds on algebraic decision trees has a long history (a recent overview of the known methods can be found, e.g., in [GKV95]; see also [R72], [M85], [MPR94], [Y92], [Y94], [GV96], [GKMS96]). However almost all known results (with the exception of [R72], [MPR94], and [GV96]) concern *algebraic* decision trees, i.e. decision trees with the gate functions being polynomials.

In this paper we introduce a new method for proving lower bounds for a stronger computational model of (*deterministic* and *randomized*) analytical decision trees, i.e. the trees with the gate functions being *analytic* (cf. also [R72]).

Let us briefly mention the main results of the paper.

In subsection 3.1, after describing the general method we give a short proof for Rabin's ([R72]) lower bound n (closing also a gap in his original proof for the case of *analytic* functions; cf. [R72], [MPR94]) on the depth of testing membership to an octant $\mathbb{R}^n_+ = \{(x_1, \cdots, x_n) \in \mathbb{R}^n : x_1 \geq 0, \cdots, x_n \geq 0\}$ by a *deterministic analytic* decision tree. In subsection 3.2 we design a *randomized algebraic* decision tree (with the gates being polynomials of degrees at most n) which recognizes $\mathbb{R}^n_+$ with the depth $0(\log^2 n)$. Furthermore, we design another randomized tree of the same type computing $\max\{x_1, \cdots, x_n\}$ with the depth $0(\log^5 n)$. This extends the result of [TY94] which was for the case of $x_1, \cdots, x_n$ being pairwise distinct.

In Section 4 we study the size of *analytic* decision trees (which is a stronger complexity measure than the commonly considered depth, since a lower bound on the size *implies* immediately a lower bound on the depth). In particular, as a corollary we prove an exponential lower bound $2^{\Omega(n)}$ on the size of analytic decision trees testing membership to the set of the points $(x_1, \cdots, x_{2n}) \in \mathbb{R}^{2n}$ with exactly n negative coordinates. Notice, that the only known so far exponential lower bound on the size of decision trees was obtained in [GKY95] for testing the octant $\mathbb{R}^n_+$ under the assumption that the tree is *ternary* (i.e., branching according to the inequalities $<, =, >$) rather than a usual *binary* one (which branches according to $\leq, >$) and besides, the decision tree is algebraic if the gate functions are polynomials of a fixed degree.

Finally, in Section 5 we obtain a lower bound $\Omega(\sqrt{n})$ on the depth of randomized analytic decision trees which recognize a set of the type $\{(x_1, \cdots, x_n)$: the number of negative elements among $x_1, \cdots, x_n$ is a multiple of $q\}$ for a fixed q being not a power of 2. Notice that this is the first nontrivial lower bound for randomized analytic decision trees (for randomized algebraic decision trees the nonlinear lower bounds were proved in [GKMS96]).

A method for obtaining nonlinear lower bounds on the depth of Pfaffian computation trees (which are the trees with the gates being Pfaffian functions and thus, lying between the algebraic and analytic decision trees) for the problem of testing membership to a polyhedron, was developed in [GV96]. This result is however independent from the present paper since relying on the methods introduced below, one could get only linear lower bounds on the depth.

2 Preliminaries

Similarly as in [GKMS96], for a given polynomial $g \in \mathbb{R}[X_1, \cdots, X_n]$, we define its leading term $lm(g)$ as follows. First we take the terms of g with the least degree in X_n, then among them with the least degree in X_{n-1} and so on, till X_1. One can describe $lm(g)$ by means of infinitesimals (cf., e.g., [GKMS96]).

Namely for a real closed field F (see e.g. ([L65], [GV88]) we say that an element ε transcendental over F is an infinitesimal (with respect to F) if $0 < \varepsilon < a$ for any element $0 < a \in F$. This uniquely induces the order on the field $F(\varepsilon)$ of rational functions and further on the real closure $\widetilde{F(\varepsilon)}$ (see [L65]). Now let $\varepsilon_1 > \cdots > \varepsilon_n > 0$ be the elements such that $\varepsilon_{\ell+1}$ is infinitesimal with respect to the real closed field $\widetilde{\mathbb{R}(\varepsilon)}$ for $\varepsilon = (\varepsilon_1, \cdots, \varepsilon_\ell)$, $0 \leq \ell < n$. Then the sign $sgn(g(\varepsilon_1, \cdots, \varepsilon_n)) = sgn(lm(g)(\varepsilon_1, \cdots, \varepsilon_n))$ and on the other hand this property uniquely determines the term $lm(g)$. Actually, one could stick in the arguing below with the real numbers $1 = \varepsilon_0^{(0)} > \varepsilon_1^{(0)} > \cdots > \varepsilon_n^{(0)} > 0$ instead of $\varepsilon_1, \cdots, \varepsilon_n$ where $\varepsilon_{\ell+1}^{(0)}$ is "considerably smaller" than $\varepsilon_\ell^{(0)}$, $0 \leq l \leq n-1$. But then one should specify, what does it mean "considerably smaller", and it is more convenient to use infinitesimals.

As computational models we deal with the decision trees (DTs) (see e.g. [R72], [MPR94], [Y94], [GKV95], [GKY95]). We consider two kinds of gates of DTs: either polynomials of degrees at most d, then we denote the corresponding algebraic decision trees by d-DT, or the functions, being *real analytic* (cf. [C48]) in a certain vicinity of the origin, then we denote the corresponding analytic decision trees by A-DT. We denote by d-RDT or A-RDT, respectively, their *randomized* counterparts, called *randomized decision* trees, which are the sets $\{T_\alpha\}$ (see e.g. [MT82], [M85], [GKMS96]), with T_α being a *deterministic* d-DT or A-DT, respectively, chosen with a probability $p_\alpha \geq 0$, $\sum p_\alpha = 1$.

Observe that for a function $\mathfrak{a}$ in n variables $X_1, \cdots, X_n$, being real analytic at the origin, one can literally extend the notion of the leading term $lm(\mathfrak{a})$ as above, treating $\mathfrak{a}$ as a power series in $X_1, \cdots, X_n$. Also $sgn(\mathfrak{a}(\varepsilon_1, \cdots, \varepsilon_n)) = sgn((lm(\mathfrak{a}))(\varepsilon_1, \cdots, \varepsilon_n))$ holds, herewith the power series $\mathfrak{a}(\varepsilon_1, \cdots, \varepsilon_n)$ could be naturally considered as an element of the real closed field $\mathbb{R}_n$, where $\mathbb{R}_0 = \mathbb{R}$ and for each $0 \leq i \leq n-1$ $\mathbb{R}_{i+1}$ is the field of Puiseux series $\sum_{j\geq 0} p_j \varepsilon_{i+1}^{\mu_j/\partial}$, $p_j \in \mathbb{R}_i$, $1 \leq \partial \in \mathbb{Z}$, integers $\mu_0 < \mu_1 < \cdots$ increase (see e.g. [GV88]). Since $\mathbb{R}_n$ is a real closed field, due to Tarski's transfer principle [T51], the sign $sgn((lm(\mathfrak{a}))(\varepsilon_1, \cdots, \varepsilon_n))$ does not depend on, whether we regard $(lm(\mathfrak{a}))(\varepsilon_1, \cdots, \varepsilon_n)$ as an element of the real closure $\widetilde{\mathbb{R}(\varepsilon_1, \cdots, \varepsilon_n)}$ or of its extension $\mathbb{R}_n$.

3 Testing octant: deterministic vs. randomized decision trees

Testing membership to the nonnegative octant $\mathbb{R}_+^n$ was firstly studied by M. Rabin in [R72], where a (sharp) lower bound n was formulated for the depth of A-DT (a gap in the proof was filled in [MPR94] for algebraic or Nash gate functions). In the next subsection we give a short proof of the bound for the case of analytic functions, closing for the first time a gap in Rabin's original proof [R72] for this case.

3.1 Deterministic decision trees

Let an A-DT T test membership to $\mathbb{R}_+^n$. For any vector $\sigma = (\sigma_1, \cdots, \sigma_n) \in \{-1, 1\}^n$ consider a point $E_\sigma = (\sigma_1\varepsilon_1, \cdots, \sigma_n\varepsilon_n) \in (\mathbb{R}_n)^n$. Consider any gate $\mathfrak{a}$ of T, being a real analytic function. For any point $(\varepsilon_1^{(0)}, \cdots, \varepsilon_n^{(0)}) \in \mathbb{R}^n$ where $\varepsilon_1^{(0)} > \cdots > \varepsilon_n^{(0)} > 0$ and $\varepsilon_{i+1}^{(0)}$ is sufficiently less than $\epsilon_i^{(0)}$, $0 \leq i \leq n-1$, we have $sgn(\mathfrak{a}(\sigma_1\varepsilon_1^{(0)}, \cdots, \sigma_n\varepsilon_n^{(0)})) = sgn((lm(\mathfrak{a}))(\sigma_1\varepsilon_1^{(0)}, \cdots, \sigma_n\varepsilon_n^{(0)}))$. Also $sgn(\mathfrak{a}(E_\sigma)) = sgn((lm(\mathfrak{a}))(E_\sigma))$ (cf. section 2 above), obviously $sgn((lm(\mathfrak{a}))(\sigma_1\varepsilon_1^{(0)}, \cdots, \sigma_n\varepsilon_n^{(0)})) = sgn((lm(\mathfrak{a}))(E_\sigma))$. Thus, $sgn(\mathfrak{a}(E_\sigma)) = sgn(\mathfrak{a}(\sigma_1\varepsilon_1^{(0)}, \cdots, \sigma_n\varepsilon_n^{(0)}))$ and thereby runs correctly for an input point E_σ.

Notice that the above argument was necessary since we deal with A-DTs. If we would consider d-DT rather than A-DT, we could immediately apply Tarski's transfer principle [T51] to ensure that d-DT runs correctly for any input point from $(\mathbb{R}_n)^n$. For the purpose of this paper the restriction on the input points E_σ for A-DT suffices.

Take the path in T along which T runs for the point $E_{(1,\cdots,1)} = (\varepsilon_1, \cdots, \varepsilon_n)$ (and therefore, outputs "yes"). Let $g_1, \cdots, g_t$ be the testing (real analytical) functions along this path.

Lemma 1. $t \geq n$

Proof. Denote $lm(g_j) = c_j X_1^{s_{1,j}} \cdots X_n^{s_{n,j}}, c_j \in \mathbb{R}$, $1 \leq j \leq t$. The sign $sgn(lm(g_j))$ is determined by the vector $S_j = (s_{1,j}, \cdots, s_{n,j}) \pmod 2 \in (\mathbb{F}_2)^n$, $1 \leq j \leq t$. Suppose that $t < n$. Then there exists a nonzero vector $(s_1, \cdots, s_n) \in \mathbb{F}_2^n$ such that the inner products $((s_1, \cdots, s_n), S_j) = 0 \pmod 2$, $1 \leq j \leq t$. Denote $\sigma = ((-1)^{s_1}, \cdots, (-1)^{s_n})$. Then $lm(g_j(E_{(1,\cdots,1)}) = lm(g_j(E_\sigma))$,

$1 \leq j \leq t$; i.e. E_σ satisfies all the tests along the path under consideration, and thereby the output of T for the input E_σ is "yes", but E_σ does not belong to the nonnegative octant, the obtained contradiction proves the lemma.

Corollary 1. ([R72])

Any A-DT testing membership to $\mathbb{R}^n_+$ has the depth at least n.

3.2 Randomized decision trees

In [TY94] it was shown that testing membership to the octant $\mathbb{R}^n_+$ can be performed by a n-RDT with the depth $(\log n)^{0(1)}$ under the assumption that all the coordinates of an input vector $(x_1, \cdots, x_n) \in \mathbb{R}^n$ are nonzeros. In this subsection we design an n-RDT testing membership to $\mathbb{R}^n_+$ for arbitrary input vectors.

Thus, RDT (in particular n-RDT and A-RDT) could have much less depth than any DT solving the same problem, cf. corollary 1. On the other hand, in [GKMS96] it was proved the lower bound $\frac{n}{2d}$ on the depth of d-RDT testing membership to $\mathbb{R}^n_+$. This shows that there is a *noncollapsing* hierarchy on the computational power of d-RDTs with respect to d.

Let $(x_1, \cdots, x_n) \in \mathbb{R}^n$ be an input vector. Denote by $P \subset \{1, \cdots, n\}$ the subset of j such that $x_j < 0$. Treating $\{1, \cdots, n\}$ as a subset of $V = (\mathbb{F}_2)^{\lceil \log_2 n \rceil}$ (in an arbitrary way), we apply to P theorem 2.4 [VV86]. It states that for a random choice of vectors $w_1, \cdots, w_{\lceil \log_2 n \rceil} \in V$ the probability that one of the truncated sets $P_\ell = P \cap \{v \in V; (v, w_i) = 0, \ 1 \leq i \leq \ell\}$, $0 \leq \ell \leq \lceil \log_2 n \rceil$ consists of a single element is at least $1/4$ (provided that $P \neq \phi$). For any $1 > \delta > 0$ making $0(\log 1/\delta)$ rounds of choosing the vectors $w_1, \cdots, w_{\lceil \log_2 n \rceil}$, we could achieve the latter probability to be greater than $1 - \delta$ (for at least one of the rounds).

For the next step we need to be able to pick out randomly a homogeneous multilinear polynomial h_k from $\mathbb{R}[Y_1, \cdots, Y_m]$ of degree k (for $0 \leq k \leq m$) and with all the coefficients in the interval $[0, 1]$. In fact, one could pick out randomly from a suitable finite set of such polynomials, or one could use the general statement from [M85] which enables us for a randomized decision tree with a continuous random parameter to replace it by a discrete one. For the reason of simplicity we will use a continuous random parameter.

Thus, fix for the time being a chosen randomly truncated set $\{v \in V; (v, w_i) = 0, \ 1 \leq i \leq \ell\} = \{j_1, \cdots, j_m\}$. Denote $\{Y_1, \cdots, Y_m\} = \{X_{j_1}, \cdots, X_{j_m}\}$. Observe that a random homogeneous multilinear polynomial $h_k \in \mathbb{R}[Y_1, \cdots, Y_m]$ vanishes (with the probability 1) at the point $(y_1, \cdots, y_m) = (x_{j_1}, \cdots, x_{j_m})$ if and only if the number of zeroes among $y_1, \cdots, y_m$ is greater than $m - k$ (if the latter is not fulfilled it vanishes with the probability zero).

We construct an n-RDT T, which using binary search is testing $h_{\lceil m/2 \rceil}(y_1, \cdots, y_m)$, then testing $h_{\lceil m/4 \rceil}(y_1, \cdots, y_m)$ if the first test returns zero, or else testing $h_{3\lceil m/4 \rceil}(y_1, \cdots, y_m)$ and so on, finds the minimal k_0 for which $h_{k_0}(y_1, \cdots, y_m)$ vanishes. Then $m - k_0 + 1$ equals (with the probability 1) to the number of zeroes among $y_1, \cdots, y_m$. Test also $h_{k_0 - 1}(y_1, \cdots, y_m)$, unless $k_0 = 1$ and in this case $(y_1, \cdots, y_m) = (0, \cdots, 0)$ and we agree $1 \geq h_0 \geq 0$. If all $y_1, \cdots, y_m$ were nonnegative (in particular, if $(x_1, \cdots, x_n) \in \mathbb{R}^n_+$) then the latter test would be positive. If among $y_1, \cdots, y_m$ was exactly one negative element then the latter test would be negative (with the probability 1).

Summarizing, T makes $0(\log 1/\delta)$ rounds, choosing at every round some vectors

$w_1, \cdots, w_{\lceil \log_2 n \rceil}$, then for each truncated set $(y_1, \cdots, y_m)$ finds k_0 as described above and tests $h_{k_0-1}(y_1, \cdots, y_m)$. If all these tests are positive, then T returns $(x_1, \cdots, x_n) \in \mathbb{R}^n_+$, else if at least one of the tests is negative, T returns $(x_1, \cdots, x_n) \notin \mathbb{R}^n_+$.

It is not difficult to see the correctness of T in testing membership to $\mathbb{R}^n_+$. Indeed, if $(x_1, \cdots, x_n) \in \mathbb{R}^n_+$ then all the described tests $h_{k_0-1}(y_1, \cdots, y_m)$ are positive. Else, if $(x_1, \cdots, x_n) \notin \mathbb{R}^n_+$ then with the probability greater than $1 - \delta$ one of the truncated sets $(y_1, \cdots, y_m)$ contains a single negative element. Then for this truncated set the test $h_{k_0-1}(y_1, \cdots, y_m)$ would be negative.

Now complete the depth analysis of T. There are $0(\log 1/\delta)$ rounds choosing vectors $w_1, \cdots, w_{\lceil \log_2 n \rceil}$, each of these vectors yields a truncated set $\{y_1, \cdots, y_m\} \subset \{x_1, \cdots, x_n\}$. For every of these truncated sets T finds k_0 by binary search, which in its turn also requires $0(\log n)$ steps. Thus, the depth of n-RDT T can be bounded by $0(\log^2 n \log 1/\delta)$.

As an application of the described n-RDT one could design an n-RDT with a similar depth $0(\log^2 n \log 1/\delta)$ and the probability greater than $1 - \delta$ for the problem $MAX =$ (cf. [TY94], [GKY95]), namely, whether $x_1 = \max\{x_1, \cdots, x_n\}$ for an input vector $(x_1, \cdots, x_n)$. It suffices to apply T to the vector $(x_1 - x_2, \cdots, x_1 - x_n) \in \mathbb{R}^{n-1}$.

If one would like to solve the MAX problem (i.e. computing $\max\{x_1, \cdots, x_n\}$), then similarly as in [TY94] it is necessary to have a subroutine which increases a candidate for $\max\{x_1, \cdots, x_n\}$, in other words, which finds an element x_j greater than x_1 (provided that such x_j does exist). It corresponds to detecting negative coordinate among $x_1 - x_2, \cdots, x_1 - x_n$ (provided, it does exist).

Namely, when a truncated set $(y_1, \cdots, y_m)$ with the negative test $h_{k_0-1}(y_1, \cdots, y_m)$ is found, we use the binary search to test as above, whether for the set $(y_1, \cdots, y_{\lceil m/2 \rceil})$ for the maximal k_1 for which $h_{k_1-1}(y_1, \cdots, y_{\lceil m/2 \rceil})$ does not vanish, the inequality $h_{k_1-1}(y_1, \cdots, y_{\lceil m/2 \rceil}) < 0$ holds. If this is the case, then proceed to the half $(y_1, \cdots, y_{\lceil m/2 \rceil})$, else if $h_{k_1-1}(y_1, \cdots, y_{\lceil m/2 \rceil}) > 0$, then proceed to the half $(y_{\lceil m/2 \rceil}, \cdots, y_m)$, and so on. If $(y_1, \cdots, y_m)$ contained a single negative element after $\lceil \log_2 m \rceil$ steps, the described subroutine would find it. Thus, the depth of n-RDT for the described subroutine which finds a negative element among $x_1, \cdots, x_n$ (or returns that $(x_1, \cdots, x_n) \in \mathbb{R}^n_+$) is bounded by $0(\log^3 n \cdot \log 1/\delta)$. The probability of the correct output is greater than $1 - \delta$.

Finally, in [TY94] it is shown that the result of applying the procedure of finding a greater element among $x_1, \cdots, x_n$, successively $0(\log n)$ times, taking $\delta = 0(1/n)$ equals to $\max\{x_1, \cdots, x_n\}$ with the probability close to 1. Thus, one can compute $\max(\{x_1, \cdots, x_n\}$ by n-RDT with the depth $0(\log^5 n)$.

Let us summarize what we have proved in this subsection in the following theorem.

Theorem 1. For each of the following problems there is an n-RDT which for any input vector $(x_1, \cdots, x_n) \in \mathbb{R}^n$

a) tests membership to $\mathbb{R}^n_+$ or tests whether $x_1 = \max\{x_1, \cdots, x_n\}$ in the depth $0(\log^2 n)$;

b) finds a negative x_i (or returns that $(x_1, \cdots, x_n) \in \mathbb{R}^n_+$) in the depth $0(\log^3 n)$

c) computes i such that $x_i = \max\{x_1, \cdots, x_n\}$ in the depth $0(\log^5 n)$.

4 Exponential lower bound on the size of deterministic analytic decision trees

In this section we study the size of a decision tree as its complexity measure rather than its depth. Evidently, a lower bound on the size immediately implies a (logarithmic) lower bound on the depth, so it is a more difficult problem, and the known methods for obtaining lower bounds on the depth (see e.g. [GKV95] and the references there) do not give any lower bound on the size. Besides, as a counterpart to Rabin's linear lower bound on the depth for testing membership to $\mathbb{R}_+^n$ (see subsection 3.1) an upper linear bound on the size is obvious. The point is that we deal usually with the *binary* decision trees (i.e. branching at $\leq$ or $>$). In [GKY95] *ternary* decision trees were studied (i.e. branching goes according to $<$, $=$, $>$) and an exponential lower bound on the size for testing membership to $\mathbb{R}_+^n$ was obtained for algebraic d-DT where $d = const$. However, the result of [GKY95] cannot be deduced from the methods of this section since these methods work for binary decision trees, and on the other hand for binary trees there is already mentioned above obvious linear upper bound on the size for testing $\mathbb{R}_+^n$. Thus, the lower bounds on the size for binary and ternary trees are independent.

In this section we design a method for obtaining the first exponential lower bounds on the size of binary analytic decision trees, and we provide some concrete examples of the problems for which the sizes of A-DTs are exponential.

Consider an A-DT T. As in the subsection 3.1 we restrict T to the inputs E_σ. In this setting we attach to T a function $b : \{-1,1\}^n \rightarrow \{-1,1\}$ which maps σ to 1 if and only if E_σ is *accepted* by T (to each *accepting* (resp. *rejecting*) leaf of T 1 (resp. -1) is attached). One could treat b as a boolean function (cf. [BS90], [KM91]) and also as an element of a bigger set $\mathcal{B}_n$ of functions $\{-1,1\}^n \rightarrow \mathbb{R}$ which is isomorphic to $\mathbb{R}^{2^n}$. Then $\mathcal{B}_n$ is $\mathbb{R}$-space with the basis of all multilinear monomials $\{X^I = X_1^{i_1} \cdots X_n^{i_n}\}$; $i_1, \cdots, i_n \in \{0,1\}$.

Thus, for a boolean function b we have an expansion $b = \sum_I \beta_I X^I$, herewith the norm $L_2((\beta_I)_I) = \sum_I \beta_I^2 = 1$ (since the vector $(\beta_i)_I$ is an image of the vector $\frac{1}{(\sqrt{2})^n}(b(x))_x$ with L_2-norm equal to 1 under the unitary Fourier transform being n-th tensor power of the matrix $\frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$. The important feature of b studied in [BS90], [KM91] is its L_1-norm $L_1(b) = \sum_I |\beta_I|$. We use the following lemma from [KM91] for which we give here also a short proof.

Lemma 2. ([KM91])*If the tree T has m leaves then $L_1(b) \leq m$.*

Proof. As we restrict T to the inputs E_σ we could replace each gate g of T by $lm(g)$ (see subsection 3.1). Thereby, to any subtree T' of T we could assign a (boolean) function $b_{T'} : \{-1,1\}^n \rightarrow \{-1,1\}$, then b is assigned to the whole tree T.

We prove lemma by induction on the size of the tree. In case of the base of induction the tree consists of a single leaf with constant 1 or -1 boolean function attached. For the inductive step consider a term $lm(g) = c\,X^I$, $c \in \mathbb{R}$ in the root v of T and let the boolean functions $b^{(1)}, b^{(2)}$ are attached to two subtrees $T^{(1)}, T^{(2)}$ of T with the roots being the sons of v. Then $b = \frac{1}{2}(1 - X^I)b^{(1)} + \frac{1}{2}(1 + X^I)b^{(2)}$ and hence $L_1(b) \leq L_1(b^{(1)}) + L_1(b^{(2)})$. Then applying inductive hypothesis to the subtrees $T^{(1)}, T^{(2)}$ completes the proof of the lemma.

To exhibit an example of a set, for which the membership requires an exponential size for any A-DT T, denote by $C_\sigma = \{\sigma_1 X_1 > 0, \cdots, \sigma_n X_n > 0\}$, $\sigma_1, \cdots, \sigma_n \in \{-1, 1\}$ an octant. Assume that T recognizes membership to a set $\mathcal{M}$ such that $\cup_{\sigma \in M} C_\sigma \subset \mathcal{M} \subset \cup_{\sigma \in \mathcal{M}} \mathcal{C}_\sigma \cup \{\mathcal{X}_1 \cdots X_n = 0\}$ for a certain set $M \subset \{-1, 1\}^n$, i.e. the inner part of $\mathcal{M}$ coincides with $\cup_{\sigma \in M} C_\sigma$. Denote by $b_{\mathcal{M}} : \{-1, 1\}^n \to \{-1, 1\}$ the boolean function such that $b_{\mathcal{M}}(\sigma) = -1$ if and only if $\sigma \in M$.

Lemma 2 provides the lower bound $L_1(b_{\mathcal{M}})$ on the size of a decision tree T testing the set $\mathcal{M}$. Now we give two examples of sets $\mathcal{M}$ with a big norm $L_1(b_{\mathcal{M}})$ taken from [BS90].

Let $n = 2k$ and define $\mathcal{M}_{\text{EXACT}} \subset \mathbb{R}^n$ to be the set of points $(x_1, \cdots, x_n)$ with exactly k negative coordinates among $x_1, \cdots, x_n$.

Now let $n = 4k$ and define $\mathcal{M}_4 \subset \mathbb{R}^n$ to be the set of all the points $(x_1, \cdots, x_n)$ such that for each $0 \le i \le k-1$ either x_{4i+1}, x_{4i+2} are both negative or x_{4i+3}, x_{4i+4} are both negative.

Using the bounds $L_1(b_{\mathcal{M}_{\text{EXACT}}}) \ge 2^k/k$ (observe that this bound is close to the possible largest bound due to the Cauchy inequality $L_1(b) \le 2^{n/2}$ for any boolean function $b \in \mathcal{B}_n$), $L_1(b_{\mathcal{M}4}) \ge (1.25)^k$ [BS90] and Lemma 2 we get the following corollary.

Corollary 2. Any analytic decision tree testing membership to a) $\mathcal{M}_{\text{EXACT}}$ or to b) $\mathcal{M}_4$ has the size greater than $2^{\Omega(n)}$.

5 Lower bound on the depth of randomized analytic decision trees

We have shown in Section 3 that *randomization* can enhance dramatically the efficiency of analytic decision trees. In this section we a prove lower bound $\Omega(\sqrt{n})$ for randomized analytic decision trees recognizing sets like $\mathcal{L}_{i,q} = \cup_\sigma \pmod q \equiv i C_\sigma$, where the union is taken over $\sigma \in \{-1, 1\}^n$ such that the number of -1 in σ has a residue $i \pmod q$, and q is not a power of 2.

Thus, assume A-RDT $T^{(i,q)} = \{T_\alpha\}$ with the depth t recognizes $\mathcal{L}_{i,q}$. Assuming that q is small (say, a constant), one can suppose q to be an odd prime, taking into account that the complexities of recognizing $\mathcal{L}_{i,q}$ for diverse i (and fixed q) coincide. Indeed, in order to reduce recognizing $\mathcal{L}_{i,q}$ to recognizing $\mathcal{L}_{j,q}$ one replaces the input $(x_1, \cdots, x_n)$ by $(-x_1, \cdots, -x_{j-i}, x_{j-i+1}, \cdots, x_n)$.

Again as in the previous section we restrict $T^{(i,q)}$ to the set of 2^n points E_σ and take A-DT T_α which makes at most $\frac{1}{3}2^n$ errors on the points $E \subset \{E_\sigma\}_\sigma$, i.e. $|E| \le \frac{1}{3}2^n$. Again as in Section 4 we associate with T_α a boolean function b_α, but unlike Section 4 in a more standard setting, namely $b_\alpha : \{0, 1\}^n \to \{0, 1\} = \mathbb{F}_2$. For each gate g (being an analytic function) of T_α consider $lm(g) = cX_1^{i_1} \cdots X_n^{i_n}$ and replace g by a linear form $L_g(y_1, \cdots, y_n) = i_1 y_1 + \cdots + i_n y_n \pmod 2 : \mathbb{F}_2^n \to \mathbb{F}_2$. To every path of T_α with the gate functions $g_1, \cdots, g_k$, we attach the product of linear functions $(L_{g_1} + \tau_1) \cdots (L_{g_k} + \tau_k)$ where $\tau_i \in \{0, 1\}$, $1 \le i \le k$ is the corresponding sign of the branch at the path with the gate function g_i. Then b_α coincides with the sum of the products $(L_{g_1} + \tau_1) \cdots (L_{g_k} + \tau_k)$ attached to all the paths with the outputs 1. Similar to Section 4 we can give an inductive description of b_α. For the base of induction consider a tree consisting of a single leaf and b_α equals to the output of this leaf. For the inductive step let the gate g be assigned at the root v of T_α and the boolean functions $b^{(1)}, b^{(2)}$ are attached to the left and right subtrees, respectively, with the roots being the sons of v. Then $b_\alpha = L_g \cdot b^{(1)} + (L_g + 1) b^{(2)}$.

Therefore, $\deg b_\alpha \le t$. Thus, b_α coincides with the boolean function $\mathrm{MOD}_{i,q}$ at more than $\frac{2}{3}2^n$ points, and hence the Corollary and Lemma 4 [S87] imply that $\deg b_\alpha \ge \Omega(\sqrt{n})$ for a certain $0 \le i \le q-1$, see above (to apply directly Corollary [S87] one has to imbed the functions b_α, $\mathrm{MOD}_{i,q}$ in the set of functions $\{0,1\}^n \to \mathbb{F}_{2^\ell}$ for a suitable extension $\mathbb{F}_{2^\ell}$ of $\mathbb{F}_2$, cf. lemma 5 [S87]). Thus, we get the following theorem.

Theorem 2. *Any A-RDT which recognizes the union of octants $\cup_{\sigma \pmod q)\equiv i} C_\sigma$ has the depth greater than $\Omega(\sqrt{n})$ (for a fixed q being not a power of 2).*

□

6 Open Problems and Further Research

There remain important open problems on *randomized* decision complexity of many concrete problems which are expressible by *simultaneous* positivity of small degree polynomials, like quadratic or cubic ones. The interesting examples include *Element Distinctness* in algebraic *computation tree* model or for n-RDTs (cf. a *randomized* lower bound $\Omega(n \log n)$ [GKMS96] for n^δ-RDTs with sufficiently small $\delta > 0$), *Finite Union of Balls* in $\mathbb{R}^n$, or algebraic version of 3SAT being the existentional problem of simultaneous positivity of cubic polynomials.

Acknowledgement

We thank Andy Yao for many stimulating discussions on the subject of the paper.

References

[BLY92] A. Björner, L. Lovasz, A. Yao, *Linear Decision Trees: Volume Estimates and Topological Bounds*, Proc. ACM STOC, 1992, pp. 170–177.

[BS90] J. Bruck, R. Smolensky, *Polynomial threshold functions, AC^0 functions and spectral norms.*, Proc. IEEE FOCS, 1990, pp. 632–641.

[BKL93] P. Buergisser, M. Karpinski, T. Lickteig, *On Randomized Algebraic Test Complexity*, Computational Complexity **9**, 1993, pp. 231–251.

[C48] R. Courant, *Theory of Functions of a Complex Variable*, New York University Press, 1948.

[GK93] D. Grigoriev, M. Karpinski, *Lower Bounds on Complexity of Testing Membership to a Polygon for Algebraic and Randomized Computation Trees*, Technical Report TR-93-042, International Computer Science Institute, Berkeley, 1993.

[GKMS96] D. Grigoriev, M. Karpinski, F. Meyer auf der Heide, R. Smolensky, *A lower bound for randomized algebraic decision trees*, to appear in Proc. ACM STOC, 1996.

[GKV95] D. Grigoriev, M. Karpinski, N. Vorobjov, *Improved lower bound on testing membership to a polyhedron by algebraic decision trees*, Proc. IEEE FOCS, 1995, pp. 258–265.

[GKY95] D. Grigoriev, M. Karpinski, A. Yao, *An exponential lower bound on the size of algebraic decision trees for MAX*, Technical Report TR-95-066, International Computer Science

Institute, Berkeley, 1995; submitted to Computational Complexity.

[GV88] D. Grigoriev, N. Vorobjov, *Solving systems of polynomial inequalities in subexponential time*, J. Symb. Comput. **5**, 1988, pp. 37–64.

[GV96] D. Grigoriev, N. Vorobjov, *Complexity lower bounds for computation trees with elementary transcendental function gates*, Proc. IEEE FOCS, 1994, pp. 548–552.

[KM91] E. Kushilevitz, Y. Mansour, *Learning decision trees using the Fourier spectrum*, Proc. ACM STOC, 1991, pp. 455–464.

[L65] S. Lang, *Algebra*, Addison-Wesley, 1965.

[MT82] U. Manber, M. Tompa, *Propabilistic, Nondeterministic and Alternating Decision Trees*, Proc. ACM STOC, 1982, pp. 234–244.

[M85] F. Meyer auf der Heide, *Simulating Probabilistic by Deterministic Algebraic Computation Trees*, Theor. Comput. Sci. **41**, 1985, pp. 325-330.

[MPR94] L. Montana, L. Pardo, T. Recio, *A note on Rabin's width of a complete proof*, Computational Complexity **4**, 1994, pp. 12–36.

[R72] M. Rabin, *Proving simultaneous positivity of linear forms*, J. Comput. Syst. Sci. **6**, 1972, pp. 639–650.

[S87] R. Smolensky, *Algebraic methods in the theory of lower bounds for boolean circuit complexity*, Proc. ACM STOC, 1987, pp. 77–82.

[T51] A. Tarski, *A decision method for elementary algebra and geometry*, University of California Press, 1951.

[TY94] H. Ting, A. Yao, *A randomized algorithm for finding maximum with* $0(\log^2 n)$ *polynomial tests*, Inform. Processing Letters **49**, 1994, pp. 39–43.

[VV86] L. Valiant, V. Vazirani, *NP is as easy as detecting unique solutions*, Theor. Comp. Sci. **47**, 1986, pp. 85–93.

[Y92] A. Yao, *Algebraic decision trees and Euler characteristics*, Proc. IEEE FOCS, 1992, pp. 268–277

[Y94] A. Yao, *Decision Tree Complexity and Betti Numbers*, Proc. ACM STOC, 1994, pp. 615–624.