On linear decision trees computing Boolean functions

BY

HANS DIETMAR GRÖGER AND GYÖRGY TURÁN
Department of Mathematics, Statistics, and Computer Science
University of Illinois at Chicago

RESEARCH REPORT IN COMPUTER SCIENCE
NO. 44, SEPTEMBER 1990

On linear decision trees computing Boolean functions

HANS DIETMAR GRÖGER[1] and GYÖRGY TURÁN[2,3]

September 1990

Abstract

An $\Omega(n)$ lower bound is proved for the complexity of computing the inner product mod 2 of two n-bit vectors by linear decision trees, even if randomization and any fixed error probability $\epsilon < \frac{1}{2}$ is allowed.

[1] Department of Applied Computer Sciences, Attila József University, Szeged

[2] Department of Mathematics, Statistics and Computer Science, University of Illinois at Chicago, Chicago

[3] Automata Theory Research Group of the Hungarian Academy of Sciences, Szeged

1. Introduction

We consider the complexity of computing Boolean functions by deterministic and randomized linear decision trees, i.e. by decision trees where each test decides if a linear inequality holds for the input. Let us give first a brief overview of the relationship between this model and several similar ones investigated in complexity theory.

This model is a generalization of the standard decision tree model, where each test determines whether an input variable is 0 or 1. The decision tree model was studied in great detail (see the survey in Bollobás [5], and Kahn, Saks and Sturtevant [18]). Recently the randomized version also received attention (Yao [34], [36], Saks and Wigderson [28], King [19], Hajnal [17], Gröger [13], Nisan [26]).

A generalization of the standard model allowing restricted linear tests was considered in Hajnal, Maass and Turán [16], where the problem of proving lower bounds for linear decision trees was also raised. Another related generalization is discussed in Aggarwal, Coppersmith and Kleitman [1].

Linear decision trees were intensively studied in the case when the inputs are reals, rationals or integers, for problems such as sorting (Yao [33], Snir [30]) and the knapsack problem (Dobkin and Lipton [10]). The lower bound proofs in the real case typically use geometric arguments such as component counting or face counting (Dobkin and Lipton [11], Yao and Rivest [37], Snir [29]). In the integer case one can also make use of similar arguments (Klein and Meyer auf der Heide [20], Meyer auf der Heide [23]). Here it is important that the inputs can be large integers as well. We note that the lower bounds in these results are quadratic (Dobkin and Lipton [10], Klein and Meyer auf der Heide [20], Meyer auf der Heide [23]) and in some cases even exponential (Dietzfelbinger and Maass [9]), while in the model considered in this paper n is always an upper bound. Results about the probabilistic and nondeterministic versions are given in Manber and Tompa [21], Snir [31], Meyer auf der Heide [24], [25].

Linear decision trees computing Boolean functions are also somewhat related to threshold circuits (a linear decision tree of depth ℓ may be viewed as a size $O(2^{\ell})$, depth 3 threshold circuit with special structure). This connection is illustrated by the fact that a very simple proof, analogous to the proof of Theorem 1., gives a linear lower bound for the size of unrestricted depth threshold circuits (Gröger and Turán [14]).

In this paper we consider the function

$$\text{INNER PRODUCT MOD2}_n(x_1,\dots,x_n,y_1,\dots,y_n) := \mathbf{x}\cdot\mathbf{y} := (x_1 \wedge y_1)\oplus\dots\oplus(x_n \wedge y_n).$$

There are several lower bound results for this function in different models, using the fact that its table is an Hadamard matrix (Tarján [32], Chor and Goldreich [7], Babai, Frankl and Simon [2], Hajnal, Maass, Pudlák, Szegedy and Turán [15], Bruck [6]).

We show that the complexity of computing INNER PRODUCT MOD2_n by linear decision trees is $\Omega(n)$, even if randomization and an arbitrary fixed error probability $\epsilon < \frac{1}{2}$ for every input is allowed. Previously sublinear lower bounds for the complexity of graph connectivity, resp. parity in the deterministic model were obtained by Dietzfelbinger [8] and Gröger [12].

Sections 2, 3 and 4 present the lower bounds for the deterministic, Las Vegas and 2-way error randomized models, respectively. Although the linear lower bound for the 2-way error randomized model implies linear lower bounds for the simpler models, we give a proof for each case, partly because better constants are obtained for the simpler models and partly for reasons for exposition. Section 5 contains some remarks and open problems.

2. Deterministic linear decision tree complexity

A linear decision tree (LDT) over the variables $x_1, \ldots, x_m$ is a binary tree T with each inner node labelled by a linear test L of the form $\sum\limits_{i=1}^{m} \alpha_i x_i : \gamma$, where $\alpha_1, \ldots \alpha_m, \gamma$ are real numbers. The edges of T are labelled by 0 or 1, corresponding to the outcomes $<$, resp. $\geq$ of the tests. The leaves of T are labelled by 0 or 1. The output $T(\mathbf{x})$ of T for $\mathbf{x} \in \{0,1\}^m$ is the label of the leaf where $\mathbf{x}$ arrives, $t_T(\mathbf{x})$ is the number of tests along the path followed by $\mathbf{x}$. The complexity of T is $C(T) := \max\{t_T(\mathbf{x}) : \mathbf{x} \in \{0,1\}^m\}$. The LDT's having no inner nodes just a single leaf labelled by 0 or 1 are called trivial, their complexity is 0. We always assume that the trees contain no redundant tests, i.e. every leaf is reachable by some input. Two LDT's T and T' are equivalent if there is an isomorphism between the two trees such that every $\mathbf{x}$ follows paths corresponding to each other in T and T'. Note that although there are infinitely many linear tests, the number of pairwise nonequivalent LDT's is finite, as for a linear test L what matters is the dichotomy of $\{0,1\}^m$ realized by L. Let $\mathcal{T}_m$ denote a fixed set of LDT's containing exactly one representative of each equivalence class.

An LDT T computes a Boolean function $f : \{0,1\}^m \to \{0,1\}$ if $T(\mathbf{x}) = f(\mathbf{x})$ for every $\mathbf{x} \in \{0,1\}^m$. Let $\mathcal{T}_m(f)$ denote a fixed set of LDT's containing exactly one representative of each equivalence class of trees computing f.

The deterministic linear decision tree complexity of f is

$$D(f) := \min\{C(T) : T \in \mathcal{T}_m(f)\}.$$

Theorem 1. D(INNER PRODUCT MOD2_n) $\geq \frac{n}{2}$.

Proof. A linear test L given by $\sum\limits_{i=1}^{n} \alpha_i x_i + \sum\limits_{i=1}^{n} \beta_i y_i : \gamma$ is 0 (resp. 1) over a set $Z \subseteq \{0,1\}^{2n}$ if for every $(\mathbf{x}, \mathbf{y}) \in Z$ it holds that $\boldsymbol{\alpha}\mathbf{x} + \boldsymbol{\beta}\mathbf{y} < \gamma$ (resp. $\boldsymbol{\alpha}\mathbf{x} + \boldsymbol{\beta}\mathbf{y} \geq \gamma$). L is nontrivial over Z if it is neither 0, nor 1 over Z.

Lemma 1. Let $X, Y \subseteq \{0,1\}^n$, $|X| = |Y| = k > 1$, and assume that L is nontrivial on $X \times Y$. Then there are partitions (X_L, X_R) of X, and (Y_L, Y_R) of Y such that $0 < |X_L| = |Y_L| < k$, L is 0 on $X_L \times Y_L$, and L is 1 on $X_R \times Y_R$.

Proof. Order the elements of X (resp. Y) according to the value of $\boldsymbol{\alpha}\mathbf{x}$ (resp. $\boldsymbol{\beta}\mathbf{y}$), resolving ties arbitrarily. Thus $X = \{\mathbf{x}_1, \dots, \mathbf{x}_k\}$, $Y = \{\mathbf{y}_1, \dots, \mathbf{y}_k\}$, where $\boldsymbol{\alpha}\mathbf{x}_1 \leq \dots \leq \boldsymbol{\alpha}\mathbf{x}_k$ and $\boldsymbol{\beta}\mathbf{y}_1 \leq \dots \leq \boldsymbol{\beta}\mathbf{y}_k$. Let ℓ be the largest index such that $\boldsymbol{\alpha}\mathbf{x}_\ell + \boldsymbol{\beta}\mathbf{y}_\ell < \gamma$. By our assumptions $0 < \ell < k$. Then $X_L := \{\mathbf{x}_1, \dots, \mathbf{x}_\ell\}$, $X_R := X \backslash X_L$, $Y_L := \{\mathbf{y}_1, \dots, \mathbf{y}_\ell\}$, $Y_R := Y \backslash Y_L$ satisfy the requirements of the lemma. □

Now consider an LDT T computing INNER PRODUCT MOD2_n. We define a sequence of sets $X_i, Y_i \subseteq \{0,1\}^n$ $(i = 0, 1, \dots)$ such that all inputs in $X_i \times Y_i$ follow the same path of length i in T. Let $X_0 = Y_0 := \{0,1\}^n$. Assume that X_{i-1} and Y_{i-1} are defined and let v_{i-1} be the node of T where the inputs in $X_{i-1} \times Y_{i-1}$ arrive after $i-1$ tests are evaluated (thus v_0 is the root). Assume that v_{i-1} is not a leaf. Apply Lemma 1. to X_{i-1}, Y_{i-1} and the test L_{i-1} in v_{i-1} to get $(X_{i-1,L}, X_{i-1,R})$ and $(Y_{i-1,L}, Y_{i-1,R})$. If $|X_{i-1,L}| \geq |X_{i-1,R}|$ then let $X_i := X_{i-1,L}$, $Y_i := Y_{i-1,L}$. Otherwise let $X_i := X_{i-1,R}$, $Y_i := Y_{i-1,R}$. Clearly all inputs in $X_i \times Y_i$ follow the same path of length i in T and it holds that $|X_i| = |Y_i| \geq 2^{n-i}$.

Theorem 1. is proved by showing that v_i is not a leaf if $i < \frac{n}{2}$.

Lemma 2. (Lindsey, see e.g. [2]). For every $X, Y \subseteq \{0,1\}^n$

$$\left| |\{(\mathbf{x},\mathbf{y}) \in X \times Y : \mathbf{x}\cdot\mathbf{y} = 1\}| - |\{(\mathbf{x},\mathbf{y}) \in X \times Y : \mathbf{x}\cdot\mathbf{y} = 0\}| \right| \leq \sqrt{|X||Y| \cdot 2^n}.$$

□

Hence if $|X| = |Y| = k$ and INNER PRODUCT MOD2_n is constant on $X \times Y$ then $k^2 \leq \sqrt{k \cdot k \cdot 2^n}$, thus $k \leq 2^{n/2}$. If v_i is a leaf then INNER PRODUCT MOD2_n is constant on $X_i \times Y_i$. Thus $2^{n-i} \leq 2^{n/2}$ and so $i \geq \frac{n}{2}$. □

3. Las Vegas linear decision tree complexity

A Las Vegas, or 0-error randomized LDT algorithm computing a Boolean function $f : \{0,1\}^m \to \{0,1\}$ is probability distribution P over $\mathcal{T}_m(f)$. $P(T)$ denotes the probability of T. The complexity of P on $\mathbf{x}$ is $E(P, \mathbf{x}) := \sum_{\mathcal{T}_m(f)} t_T(\mathbf{x}) \cdot P(T)$, i.e. the expected number of tests asked for $\mathbf{x}$. The complexity of P is $C(P) := \max\{E(P,\mathbf{x}) : \mathbf{x} \in \{0,1\}^m\}$. The Las Vegas linear decision tree complexity of f is

$$LV(f) := \inf\{C(P) : P \text{ is a Las Vegas } LDT \text{ algorithm for } f\}.$$

Theorem 2. LV(INNER PRODUCT MOD2_n) $\geq c_0 n$, where $c_0(\geq 0.21)$ is the minimum of the function $\frac{1}{4}\left(-x^2\log x-(1-x^2)\log(1-x)\right)^{-1}$ on (0, 1).

Proof. We use Yao's method [34] to translate the problem into one about distributional complexity.

Let T be an LDT computing $f:\{0,1\}^m\to\{0,1\}$ and let Q be a distribution on $\{0,1\}^m$. $Q(\mathbf{x})$ denotes the probability of $\mathbf{x}$. The complexity of T wrto Q is $A(T,Q):=\sum_{\{0,1\}^m} t_T(\mathbf{x})\cdot Q(\mathbf{x})$. The complexity of f wrto Q is $A(f,Q):=\min\{A(T,Q): T\in\mathcal{T}_m(f)\}$. The distributional linear decision tree complexity of f is

$$A(f):=\sup\{A(f,Q): Q \text{ is a distribution over } \{0,1\}^m\}.$$

Lemma 3. (Yao [34]). $LV(f)=A(f)$. □

We note that Yao proved this result and Lemma 6. (to be given in Section 4.) for decision trees, but both apply with the same proofs to LDT's as well.

Therefore to prove Theorem 2. it is sufficient to show that if Q is the uniform distribution over $\{0,1\}^{2n}$ then

$$A(\text{INNER PRODUCT MOD2}_n,\ Q)\ \geq\ c_0 n. \tag{1}$$

Let $Z\subseteq\{0,1\}^{2n}$. An LDT T is correct on Z if $T(\mathbf{x},\mathbf{y})=\mathbf{x}\cdot\mathbf{y}$ for every $(\mathbf{x},\mathbf{y})\in Z$. Let

$$A_T(X,Y):=\frac{1}{|X||Y|}\sum_{X\times Y} t_T(\mathbf{x},\mathbf{y})$$

be the average complexity of T on $X\times Y$. We define

$$f_n(k):=\min\{A_T(X,Y): X,Y\subseteq\{0,1\}^n,\ |X|=|Y|=k,\ T \text{ is correct on } X\times Y\},$$

i.e. $f_n(k)$ is the best average complexity achievable on a $k\times k$ square. With this notation (1) becomes

$$f_n(2^n)\geq c_0 n. \tag{2}$$

Now (2), and thus Theorem 2. follow from the next lemma.

Lemma 4. For every k $(1\leq k\leq 2^n)$ it holds that $f_n(k)\geq 2c_0(\log k-\frac{n}{2})$.

Proof. We argue by induction on k, if $k \leq 2^{n/2}$ then there is nothing to prove. For the induction step consider $k > 2^{n/2}$. We show that if $X, Y \subseteq \{0,1\}^n$, $|X| = |Y| = k$ and T is a correct LDT on $X \times Y$ then $A_T(X,Y) \geq 2c_0(\log k - \frac{n}{2})$.

The remark following Lemma 2. implies that T is nontrivial and hence from Lemma 1. we get partitions (X_L, X_R) and (Y_L, Y_R) such that $0 < |X_L| = |Y_L| < k$, and the test in the root of T is 0 on $X_L \times Y_L$, and 1 on $X_R \times Y_R$. Assume *w.ℓ.o.g.* that $|X_L| = \ell \geq \frac{k}{2}$.

We estimate A_T for the four rectangles obtainable from these sets. As the test in the root of T is 0 on $X_L \times Y_L$ (resp. 1 on $X_R \times Y_R$), it must be the case that the left subtree T_L of T (resp. the right subtree T_R of T) is correct on $X_L \times Y_L$ (resp. on $X_R \times Y_R$). Therefore by induction

$$A_T(X_L, Y_L) = A_{T_L}(X_L, Y_L) + 1 \geq f_n(\ell) + 1 \tag{3}$$

and

$$A_T(X_R, Y_R) = A_{T_R}(X_R, Y_R) + 1 \geq f_n(k-\ell) + 1. \tag{4}$$

For the other two rectangles we claim

$$A_T(X_R, Y_L) \geq f_n(k-\ell) \quad \text{and} \quad A_T(X_L, Y_R) \geq fn(k-\ell). \tag{5}$$

This follows from the following observation.

Lemma 5. Let $U, V \subseteq \{0,1\}^n$, $|U| = u$, $|V| = v$ and let T be a correct LDT on $U \times V$. Then $A_T(U,V) \geq f_n(\min\{u,v\})$.

Proof. Assume *w.ℓ.o.g.* $u \leq v$. If $V' \subseteq V$, $|V| = u$ then by the definition of f_n, $A_T(U, V') \geq f_n(u)$. Averaging over all such sets V'

$$f_n(u) \leq \frac{1}{\binom{v}{u}} \sum_{\substack{V' \subseteq V \\ |V'|=u}} A_T(U, V') = A_T(U, V)$$

as every $(\mathbf{x}, \mathbf{y}) \in U \times V$ is contained in the same number of rectangles $U \times V'$. □

From (3), (4) and (5), using $\ell^2 + (k-\ell)^2 \geq \frac{k^2}{2}$,

$$A_T(X,Y) \geq \frac{1}{k^2}\left(\ell^2(f_n(\ell)+1) + (k-\ell)^2(f_n(k-\ell)+1) + 2\ell(k-\ell)f_n(k-\ell)\right) =$$

$$= \frac{\ell^2 + (k-\ell)^2}{k^2} + \left(\frac{\ell}{k}\right)^2 f_n(\ell) + \left(1 - \left(\frac{\ell}{k}\right)^2\right) f_n(k-\ell) \geq$$

$$\geq \frac{1}{2} + \left(\frac{\ell}{k}\right)^2 f_n(\ell) + \left(1 - \left(\frac{\ell}{k}\right)^2\right) f_n(k-\ell).$$

From the induction hypothesis

$$\begin{aligned}
&\frac{1}{2} + (\frac{\ell}{k})^2 f_n(\ell) + \left(1 - (\frac{\ell}{k})^2\right) f_n(k-\ell) \geq \\
&\geq \frac{1}{2} + 2c_0 \left((\frac{\ell}{k})^2 \left(\log \ell - \frac{n}{2}\right) + \left(1 - (\frac{\ell}{k})^2\right) \left(\log(k-\ell) - \frac{n}{2}\right) \right) = \\
&= 2c_0 \left(\log k - \frac{n}{2}\right) + \left(\frac{1}{2} + 2c_0 \left((\frac{\ell}{k})^2 \log \frac{\ell}{k} + (1 - (\frac{\ell}{k})^2) \log(1 - \frac{\ell}{k}) \right) \right) \\
&\geq 2c_0 \left(\log k - \frac{n}{2}\right),
\end{aligned}$$

where in the last step we used the definition of c_0. Hence

$A_T(X,Y) \geq 2c_0 \left(\log k - \frac{n}{2}\right)$. □

This also completes the proof of Theorem 2. □

4. Randomized 2-way error linear decision tree complexity

A randomized 2-way error LDT algorithm computing a Boolean function $f : \{0,1\}^m \to \{0,1\}$ is a distribution P over $\mathcal{T}_m$. (As errors are permitted we do not restrict ourselves to $\mathcal{T}_m(f)$.) An LDT T makes an error on $\mathbf{x}$ if $T(\mathbf{x}) \neq f(\mathbf{x})$. Let $e(T, \mathbf{x}) = 1$ if T makes an error on $\mathbf{x}$, and $e(T, \mathbf{x}) = 0$ otherwise. The error probability of P on $\mathbf{x}$ is $e(P, \mathbf{x}) := \sum_{\mathcal{T}_m} e(T, \mathbf{x}) \cdot P(T)$. P computes f with error $\leq \epsilon$ if $e(P, \mathbf{x}) \leq \epsilon$ for every $\mathbf{x} \in \{0,1\}^m$. The complexity of P on $\mathbf{x}$ is $E(P, \mathbf{x}) := \sum_{\mathcal{T}_m} t_T(\mathbf{x}) \cdot P(T)$. The complexity of P is $C(P) := \max\{E(P, \mathbf{x}) : \mathbf{x} \in \{0,1\}^m\}$. The randomized 2-way ϵ-error linear decision tree complexity of f is

$$\begin{aligned}
R^\epsilon(f) := \inf\{C(P) : \quad & P \text{ is a randomized 2-way error LDT algorithm} \\
& \text{computing } f \text{ with error } \leq \epsilon\}.
\end{aligned}$$

Theorem 3. For every ϵ such that $0 \leq \epsilon < \frac{1}{4}$

$$R^\epsilon(\text{INNER PRODUCT MOD2}_n) \geq c_1(1-4\epsilon)\left(\frac{n}{2} + \log(1-4\epsilon)\right)$$

where $c_1 = \dfrac{1}{4 \log 3} = 0.157\ldots$.

Proof. Here again we use Yao's method [34] to reduce the problem to distributional complexity with error.

Let T be an LDT and Q be a distribution over $\{0,1\}^m$. Then T computes $f : \{0,1\}^m \to \{0,1\}$ with error $\leq \epsilon$ under Q if $\sum_{\{0,1\}^m} e(T, \mathbf{x}) \cdot Q(\mathbf{x}) \leq \epsilon$. The complexity of T wrto Q is $A(T,Q) := \sum\limits_{\{0,1\}^m} t_T(\mathbf{x}) \cdot Q(\mathbf{x})$. The ϵ-error complexity of f wrto Q is $A^\epsilon(f,Q) := \min\{A(T,Q) : T$ is an LDT computing f with error $\leq \epsilon$ under $Q\}$. The ϵ-error distributional linear decision tree complexity of f is

$$A^\epsilon(f) := \sup\{A^\epsilon(f,Q) : Q \text{ is a distribution over } \{0,1\}^m\}.$$

Lemma 6. (Yao [34]). $R^\epsilon(f) \geq \frac{1}{2}A^{2\epsilon}(f)$ for $0 \leq \epsilon \leq \frac{1}{2}$. □

Hence it is sufficient to prove that if Q is the uniform distribution over $\{0,1\}^{2n}$ and $0 \leq \epsilon < \frac{1}{2}$ then

$$(6) \qquad A^\epsilon(\text{INNER PRODUCT MOD}2_n, Q) \geq \frac{1}{2\log 3}(1-2\epsilon)\left(\frac{n}{2} + \log(1-2\epsilon)\right).$$

Let $Z \subseteq \{0,1\}^{2n}$. An LDT T is ϵ-correct on Z if $\left|\{(\mathbf{x},\mathbf{y}) \in Z : T(\mathbf{x},\mathbf{y}) \neq \mathbf{x}\cdot\mathbf{y}\}\right| = \epsilon \cdot |Z|$. T is $\leq \epsilon$-correct on Z if it is ϵ'-correct for some $\epsilon' \leq \epsilon$. We define

$$f_n(k,\epsilon) := \min\{A_T(X,Y) : X,Y \subseteq \{0,1\}^n,\ |X| = |Y| = k,\ T \text{ is } \leq \epsilon\text{−correct on } X \times Y\},$$

where $A_T(X,Y)$ is as defined in section 3. Note that $f_n(k,\epsilon)$ generalizes $f_n(k)$ as $f_n(k,0) = f_n(k)$. With this notation (6) can be expressed as

$$(7) \qquad f_n(2^n,\epsilon) \geq \frac{1}{2\log 3}(1-2\epsilon)\left(\frac{n}{2} + \log(1-2\epsilon)\right).$$

Consider the following function for $k > 0$ and $0 \leq \epsilon \leq 1$:

$$g_n(k,\epsilon) = \begin{cases} \frac{1}{2\log 3}k(1-2\epsilon)\left(\log(k(1-2\epsilon)) - \frac{n}{2}\right)^+ & \text{if } \epsilon < \frac{1}{2}, \\ 0 & \text{otherwise.} \end{cases}$$

For later reference we mention that

a) $g_n(k,\epsilon)$ is convex in ϵ for every fixed k,

b) $g_n(k,\epsilon)$ is convex in k for every fixed ϵ,

c) $\frac{1}{k}g_n(k,\epsilon)$ is monotone increasing in k for every fixed ϵ,

d) $\frac{1}{k}g_n(k,\epsilon)$ is monotone decreasing in ϵ for every fixed k,

e) $g_n(k,\epsilon) = g_n\left(\frac{1-2\epsilon}{1-2\epsilon'}k, \epsilon'\right)$ if $\epsilon, \epsilon' < \frac{1}{2}$.

Here a) and b) follow from the convexity of $x \log x$ and the fact that f^+ is convex if f is convex; c) follows from the fact that f^+ is monotone increasing if f is monotone increasing; e) follows from the definitions. Property d) follows as $(1-2\epsilon)\big(\log\big(k(1-2\epsilon)\big)-\frac{n}{2}\big)$ is convex and tends to 0 if $\epsilon \nearrow \frac{1}{2}$.

Now (7), and hence Theorem 3. are implied by the following lemma.

Lemma 7. For every k and ϵ such that $1 \leq k \leq 2^n$ and $0 \leq \epsilon \leq 1$ $f_n(k,\epsilon) \geq \dfrac{1}{k} g_n(k,\epsilon)$.

Proof. We proceed by induction on k, and for a given k we show that the lemma holds for every ϵ. If $k \leq 2^{n/2}$, the claim is obvious.

For the induction step consider $k > 2^{n/2}$. It has to be shown that if $X, Y \subseteq \{0,1\}^n$, $|X| = |Y| = k$ and T is an $\leq \epsilon$-correct LDT for $X \times Y$ then $A_T(X,Y) \geq \dfrac{1}{k} g_n(k,\epsilon)$. From property d) above we may assume *w.ℓ.o.g.* that T is ϵ-correct.

If $\epsilon \geq \frac{1}{2}$, or $k \leq \frac{2^{n/2}}{1-2\epsilon}$, the claim is again obvious. Otherwise $k > \frac{2^{n/2}}{1-2\epsilon}$. Lemma 2. implies that in this case T has to be nontrivial. Indeed, as

$$\Big|\big|\{(\mathbf{x},\mathbf{y}) \in X \times Y : \mathbf{x}\cdot\mathbf{y} = 1\}\big| - \big|\{(\mathbf{x},\mathbf{y}) \in X \times Y : \mathbf{x}\cdot\mathbf{y} = 0\}\big|\Big| \leq k \cdot 2^{n/2}$$

we get that both sets have size $\geq \frac{1}{2}k^2 - \frac{1}{2}k \cdot 2^{n/2}$, thus the trivial LDT's have error $\geq \frac{1}{2} - \frac{1}{2} \cdot \frac{2^{n/2}}{k} > \epsilon$.

Applying Lemma 1. to the linear test at the root of T we find partitions (X_L, X_R) and (Y_L, Y_R) such that the test is constant on $X_L \times Y_L$ and $X_R \times Y_R$ and *w.ℓ.o.g.* $|X_L| = \ell$, $\frac{k}{2} \leq \ell < k$.

Assume that T is ϵ_1-correct on $X_L \times Y_L$, $\epsilon_{2,1}$-correct on $X_R \times Y_R$, $\epsilon_{2,2}$-correct on $X_R \times Y_L$, $\epsilon_{2,3}$-correct on $X_L \times Y_R$ and ϵ_2-correct on $(X \times Y)\backslash(X_L \times Y_L)$. We claim

$$(8) \qquad A_T(X_L,Y_L) \geq \frac{1}{\ell} g_n(\ell,\epsilon_1) + 1 \quad \text{and} \quad A_T(X_R,Y_R) \geq \frac{1}{k-\ell} g_n(k-\ell,\epsilon_{2,1}) + 1,$$

$$(9) \qquad A_T(X_R,Y_L) \geq \frac{1}{k-\ell} g_n(k-\ell,\epsilon_{2,2}) \quad \text{and} \quad A_T(X_L,Y_R) \geq \frac{1}{k-\ell} g_n(k-\ell,\epsilon_{2,3}).$$

Here (8) follows by applying the induction hypothesis to the left and right subtrees of T; (9) follows from the following variant of Lemma 5.

Lemma 8. Let $U, V \subseteq \{0,1\}$, $|U| = u$, $|V| = v$, $w := \min\{u,v\}$, and let T be an ϵ'-correct LDT on $U \times V$. Assume that $f_n(w,\epsilon) \geq \frac{1}{w} g_n(w,\epsilon)$ holds for every ϵ $(0 \leq \epsilon \leq 1)$. Then $A_T(U,V) \geq \dfrac{1}{w} g_n(w,\epsilon')$.

Proof. Assume *w.ℓ.o.g.* $u \leq v$. If $V' \subseteq V$, $|V'| = u$ and T is $\epsilon(V')$-correct on $U \times V'$ then by the assumption $A_T(U,V') \geq \frac{1}{u} g_n\big(u, \epsilon(V')\big)$. Averaging over all sets V' and using the convexity of $g_n(k,\epsilon)$ in ϵ

$$A_T(U,V) = \frac{1}{\binom{v}{u}} \sum_{\substack{V' \subseteq V \\ |V'|=u}} A_T(U,V') \geq \frac{1}{\binom{v}{u}} \sum_{\substack{V' \subseteq V \\ |V'|=u}} \frac{1}{u} g_n\big(u, \epsilon(V')\big) \geq \frac{1}{u} g_n\Bigg(u, \frac{1}{\binom{v}{u}} \sum_{\substack{V' \subseteq V \\ |V'|=u}} \epsilon(V')\Bigg)$$

$$= \frac{1}{u} g_n(u, \epsilon').$$

□

From (8) and (9)

$$A_T(X,Y) = \frac{1}{k^2}\big(\ell^2 A_T(X_L,Y_L) + (k-\ell)^2 A_T(X_R,Y_R)$$

$$+\ell(k-\ell)\big(A_T(X_R,Y_L) + A_T(X_L,Y_R)\big)\big) \geq$$

$$\geq \frac{1}{2} + \frac{1}{k^2}\big(\ell g_n(\ell,\epsilon_1) + (k-\ell) g_n(k-\ell,\epsilon_{2,1}) + \ell\big(g_n(k-\ell,\epsilon_{2,2}) + g_n(k-\ell,\epsilon_{2,3})\big)\big)$$

where again we used $\ell^2 + (k-\ell)^2 \geq \frac{k^2}{2}$. To simplify this expression note

$$\epsilon_2(k^2 - \ell^2) = \epsilon_{2,1}(k-\ell)^2 + (\epsilon_{2,2} + \epsilon_{2,3})\ell(k-\ell),$$

so the convexity of $g_n(k,\epsilon)$ in ϵ implies

$$(k-\ell) g_n(k-\ell,\epsilon_{2,1}) + \ell\big(g_n(k-\ell,\epsilon_{2,2}) + g_n(k-\ell,\epsilon_{2,3})\big) \geq$$

$$\geq (k+\ell) g_n\left(k-\ell, \frac{k-\ell}{k+\ell}\epsilon_{2,1} + \frac{\ell}{k+\ell}(\epsilon_{2,2} + \epsilon_{2,3})\right) = (k+\ell) g_n(k-\ell,\epsilon_2).$$

Hence

(10) $$A_T(X,Y) \geq \frac{1}{2} + \frac{1}{k^2}\big(\ell g_n(\ell,\epsilon_1) + (k+\ell) g_n(k-\ell,\epsilon_2)\big).$$

Now we claim

(11) $$A_T(X,Y) \geq \frac{1}{2} + \frac{3}{k} g_n\left(\frac{k}{3}\right).$$

The definitions imply

(12) $$(1-2\epsilon)k^2 = (1-2\epsilon_1)\ell^2 + (1-2\epsilon_2)(k^2-\ell^2).$$

To prove (11) we distinguish 3 cases.

Case 1. $\epsilon_1 < \dfrac{1}{2}$ and $\epsilon_2 < \dfrac{1}{2}$.

From (10), (12) and properties b), c) and e)

$$
\begin{aligned}
A_T(X,Y) &\geq \frac{1}{2} + \frac{1}{k^2}(\ell g_n(\ell,\epsilon_1) + (k+\ell)g_n(k-\ell,\epsilon_2)) = \\
=&\frac{1}{2} + \frac{1}{k^2}\left(\ell g_n\left(\frac{1-2\epsilon_1}{1-2\epsilon}\ell,\epsilon\right) + (k+\ell)g_n\left(\frac{1-2\epsilon_2}{1-2\epsilon}(k-\ell),\epsilon\right)\right) \geq \\
\geq&\frac{1}{2} + \frac{k+2\ell}{k^2}g_n\left(\frac{1-2\epsilon_1}{1-2\epsilon}\frac{\ell^2}{k+2\ell} + \frac{1-2\epsilon_2}{1-2\epsilon}\frac{k^2-\ell^2}{k+2\ell},\epsilon\right) = \\
=&\frac{1}{2} + \frac{k+2\ell}{k^2}g_n\left(\frac{k^2}{k+2\ell},\epsilon\right) \geq \frac{1}{2} + \frac{3}{k}g_n\left(\frac{k}{3},\epsilon\right).
\end{aligned}
$$

Case 2. $\epsilon_1 < \dfrac{1}{2}$ and $\epsilon_2 \geq \dfrac{1}{2}$.

We use (10), properties c), e) and $\frac{1-2\epsilon_1}{1-2\epsilon}\ell \geq \frac{k^2}{\ell}$, which in turn follows from (12) and $\epsilon_2 \geq \frac{1}{2}$ to get

$$
\begin{aligned}
A_T(X,Y) &\geq \frac{1}{2} + \frac{1}{k^2}\ell g_n(\ell,\epsilon_1) = \frac{1}{2} + \frac{\ell}{k^2}g_n\left(\frac{1-2\epsilon_1}{1-2\epsilon}\ell,\epsilon\right) \\
&\geq \frac{1}{2} + \frac{1-2\epsilon}{1-2\epsilon_1}\frac{1}{\ell}g_n\left(\frac{1-2\epsilon_1}{1-2\epsilon}\ell,\epsilon\right) \geq \\
&\geq \frac{1}{2} + \frac{\ell}{k^2}g_n\left(\frac{k^2}{\ell},\epsilon\right) \geq \frac{1}{2} + \frac{1}{k}g_n(k,\epsilon) \geq \frac{1}{2} + \frac{3}{k}g_n\left(\frac{k}{3},\epsilon\right).
\end{aligned}
$$

Case 3. $\epsilon_1 \geq \dfrac{1}{2}$ and $\epsilon_2 < \dfrac{1}{2}$.

Here we use (10), properties c), e) and $\frac{1-2\epsilon_2}{1-2\epsilon}(k-\ell) \geq \frac{k^2}{k+\ell}$, implied by (12) and $\epsilon_1 \geq \frac{1}{2}$ to obtain

$$
\begin{aligned}
A_T(X,Y) &\geq \frac{1}{2} + \frac{1}{k^2}\cdot(k+\ell)g_n(k-\ell,\epsilon_2) = \frac{1}{2} + \frac{k+\ell}{k^2}g_n\left(\frac{1-2\epsilon_2}{1-2\epsilon}(k-\ell),\epsilon\right) \geq \\
\geq&\frac{1}{2} + \frac{1-2\epsilon}{1-2\epsilon_2}\frac{1}{k-\ell}g_n\left(\frac{1-2\epsilon_2}{1-2\epsilon}(k-\ell),\epsilon\right) \geq \frac{1}{2} + \frac{k+\ell}{k^2}g_n\left(\frac{k^2}{k+\ell},\epsilon\right) \\
\geq&\frac{1}{2} + \frac{2}{k}g_n\left(\frac{k}{2},\epsilon\right) \geq \frac{1}{2} + \frac{3}{k}g_n\left(\frac{k}{3},\epsilon\right).
\end{aligned}
$$

This proves (11). (As we assumed $\epsilon < \frac{1}{2}$, $\epsilon_1 \geq \frac{1}{2}$ and $\epsilon_2 \geq \frac{1}{2}$ is not possible.) Now to get Lemma 7. we show

$$\frac{1}{2} + \frac{3}{k}g_n\left(\frac{k}{3},\epsilon\right) \geq \frac{1}{k}g_n(k,\epsilon).$$

If $g_n\left(\frac{k}{3},\epsilon\right) > 0$ then

$$\frac{1}{2} + \frac{1}{2\log 3}(1-2\epsilon)\left(\log(\frac{k}{3}(1-2\epsilon)) - \frac{n}{2}\right) \geq \frac{1}{2\log 3}(1-2\epsilon)\left(\log\left(k(1-2\epsilon)\right) - \frac{n}{2}\right).$$

If $g_n\left(\frac{k}{3},\epsilon\right) = 0$ but $g_n(k,\epsilon) > 0$ then $\log\left(\frac{k}{3}(1-2\epsilon)\right) \leq \frac{n}{2}$ implies $\log\left(k(1-2\epsilon)\right) - \frac{n}{2} \leq \log 3$, so

$$\frac{1}{k}g_n(k,\epsilon) = \frac{1}{2\log 3}(1-2\epsilon)\left(\log(k(1-2\epsilon)) - \frac{n}{2}\right) \le \frac{1}{2}, \qquad \square$$

This completes the proof of Theorem 3. □

Finally we state a bound for every $\epsilon < \frac{1}{2}$.

Theorem 4. For every $\delta\left(0 < \delta \le \frac{1}{2}\right)$

$$R^{\frac{1}{2}-\delta}(\text{INNER PRODUCT MOD2}_n) \ge \frac{1}{20}\delta^2 \left(\frac{n}{2} - 1\right).$$

Proof. In order to decrease the error probability of a randomized algorithm one can repeat the algorithm several times and take majority vote among the outputs obtained. The standard Chernoff inequality states that if S_m is the sum of m independent random variables each taking value 1 (resp. 0) with probability p (resp. $1-p$) then $Pr(S_m \ge pm + h) \le e^{-\frac{2h^2}{m}}$. In order to decrease the error probability below $\frac{1}{8}$ it suffices to have $Pr\left(S_m \ge \frac{m}{2}\right) \le e^{-2\delta^2 m} \le \frac{1}{8}$. This holds if $m = \frac{3}{2\delta^2}$, and then Theorem 4. follows from Theorem 3. □

5. Some remarks and problems

The lower bounds of Theorems 1., 2. and 3. apply also to decision trees with separable tests, i.e. with tests of the form $f(x_1, \ldots, x_n) + g(y_1, \ldots, y_n) : \gamma$, where f and g are arbitrary functions.

On the other hand INNER PRODUCT MOD2$_n$ can be computed by a quadratic decision tree of depth $\lceil \log(n+1) \rceil$. It would be interesting to know if e.g. the function $(x_1 \wedge y_1 \wedge z_1) \oplus \ldots \oplus (x_n \wedge y_n \wedge z_n)$ is difficult for quadratic decision trees, and in general to prove lower bounds and hierarchy results for algebraic decision trees computing Boolean functions. In this context the results of Babai, Nisan and Szegedy [3] may be useful.

It would also be interesting to compare the computational power of deterministic, nondeterministic and randomized LDT's. We note that both for Boolean decision trees and for LDT's with real inputs the deterministic and the randomized versions are polynomially related (Nisan [26], Meyer auf der Heide [24]), but the reasons for this are apparently different in the two cases.

Another possibility is to consider linear branching programs computing Boolean functions, suggested by Pudlák [27]. Here again, both branching programs computing Boolean functions, and linear branching programs with real inputs (e.g. for sorting) have been studied in detail (see e.g. Barrington [4], resp. Yao [35]).

Concerning the complexity of the most difficult function in the deterministic LDT model, the standard counting method gives an $n - 2\log n$ lower bound. On the other hand every 3 variable function has complexity ≤ 2, hence every n variable function has complexity $\leq n-1$, i.e. there are no evasive functions in this model.

Finally we note that Theorem 2. implies an $\Omega(n)$ lower bound for the Las Vegas LDT complexity of determining if an n vertex graph is connected, using the reduction of [15]. The best upper bound is $O(n \log n)$, which also holds in the deterministic model.

REFERENCES

[1] A. AGGARWAL, D. COPPERSMITH, D. KLEITMAN: A generalized model for understanding evasiveness, *Inf. Proc. Lett.* **30**(1989), 205–208.

[2] L. BABAI, P. FRANKL, J. SIMON: Complexity classes in communication complexity, **27**.*FOCS* (1986), 337–347.

[3] L. BABAI, N. NISAN, M. SZEGEDY: Multiparty protocols and logspace hard pseudorandom sequences, **21**.*STOC* (1989), 1–11.

[4] D. A. BARRINGTON: Bounded-width polynomial size branching programs recognize exactly those languages in NC^1, *JCSS* 38(1989), 150–164.

[5] B. BOLLOBÁS: Extremal Graph Theory, Academic Press, 1984.

[6] J. BRUCK: Harmonic analysis of polynomial threshold functions, *SIAM J. Disc. Math.* **3**(1990), 168–177.

[7] B. CHOR, O. GOLDREICH: Unbiased bits from sources of weak randomness and probabilistic communication complexity, **26**.*FOCS* (1985), 429–442.

[8] M. DIETZFELBINGER (1987), unpublished.

[9] M. DIETZFELBINGER, W. MAASS: Lower bound arguments with "inaccessible" numbers, *JCSS* **36**(1988), 313–335.

[10] D.P. DOBKIN, R.J. LIPTON: A lower bound of $n^2/2$ on linear search programs for the knapsack problem, *JCSS* **16**(1978), 413–417.

[11] D.P. DOBKIN, R.J. LIPTON: On the complexity of computations under varying sets of primitives, *JCSS* **18**(1979), 86–91.

[12] H.D. GRÖGER (1988), unpublished.

[13] H.D. GRÖGER: On the randomized complexity of monotone graph properties (1989), submitted.

[14] H.D. GRÖGER, GY. TURÁN: A linear lower bound for the size of threshold circuits (1990), submitted.

[15] A. HAJNAL, W. MAASS, P. PUDLÁK, M. SZEGEDY, GY. TURÁN: Threshold circuits of bounded depth, **28**.*FOCS* (1987), 99–110.

[16] A. HAJNAL, W. MAASS, GY. TURÁN: On the communication complexity of graph properties, **20**.*STOC* (1988), 186–191.

[17] P. HAJNAL: The complexity of graph problems, Ph.D. Thesis, University of Chicago (1988), *TR* 88–13.

[18] J. KAHN, M. SAKS, D. STURTEVANT: A topological approach to evasiveness, *Combinatorica* 4(1984), 297–306.

[19] V. KING: Lower bounds on the complexity of graph properties, **20**.*STOC* (1988), 468–476.

[20] P. KLEIN, F. MEYER AUF DER HEIDE: A lower time bound for the knapsack problem on random access machines, *Acta Inf.* **19**(1983), 385–395.

[21] U. MANBER, M. TOMPA: The complexity of problems on probabilistic, nondeterministic and alternating decision trees, *JACM* **31**(1985), 720–732.

[22] F. MEYER AUF DER HEIDE: A polynomial linear search algorithm for the n-dimensional knapsack problem, *JACM* **31**(1984), 668–676.

[23] F. MEYER AUF DER HEIDE: Lower bounds for solving linear diophantine equations on random access machines, *JACM* **32**(1985), 929–937.

[24] F. MEYER AUF DER HEIDE: Simulating probabilistic by deterministic algebraic computation trees, *TCS* **41**(1985), 325–330.

[25] F. MEYER AUF DER HEIDE: Nondeterministic versus probabilistic linear search algorithms, **26**.*FOCS* (1985), 65–73.

[26] N. NISAN: CREW PRAMs and decision trees, *STOC* (1989), 327–335.

[27] P. PUDLÁK (1986), personal communication.

[28] M. SAKS, A. WIGDERSON: Probabilistic Boolean decision trees and the complexity of evaluating game trees, **27**.*FOCS* (1986), 29–38.

[29] M. SNIR: Proving lower bounds for linear decision trees, 8.*ICALP, LNCS* **115**(1981), 305–315.

[30] M. SNIR: Comparisons between linear functions can help, *TCS* **19**(1982), 321–330.

[31] M. SNIR: Lower bounds on probabilistic linear decision trees, *TCS* **38**(1985), 69–82.

[32] T.G. TARJÁN: Complexity of lattice–configurations, *Studia Sci. Math. Hung.* **10**(1975), 203–211.

[33] A.C. YAO: On the complexity of comparison problems using linear functions, **16**.*FOCS* (1975), 85–89.

[34] A.C. YAO: Probabilistic computations: toward a unified measure of complexity, **18**.*FOCS* (1977), 222–227.

[35] A.C. YAO: On the time-space tradeoff for sorting with linear queries, *TCS* **19**(1982), 203–218.

[36] A.C. YAO: Lower bounds to randomized algorithms for graph properties, **28**.*FOCS* (1987), 393–400.

[37] A.C. YAO, R. RIVEST: On the polyhedral decision problem, *SIAM J. Comp.* **9**(1980), 343–347.