INFORMATIQUE THÉORIQUE ET APPLICATIONS

STASYS JUKNA

A note on read-k times branching programs

Informatique théorique et applications, tome 29, n° 1 (1995), p. 75-83
<http://www.numdam.org/item?id=ITA_1995__29_1_75_0>

Informatique théorique et Applications/Theoretical Informatics and Applications
(vol. 29, n° 1, 1995, p. 75 à 83)

A NOTE ON READ-k TIMES BRANCHING PROGRAMS (*)

by Stasys JUKNA (1)

Communicated by I. WEGENER

Abstract. – *A syntactic read-k times branching program has the restriction that no variable occurs more than k times on any path (whether or not consistent). We exhibit an explicit Boolean function f, which cannot be computed by nondeterministic syntactic read-k times branching programs of size less than* $\exp\left(\Omega\left(\frac{\sqrt{n}}{k^{2k}}\right)\right)$, *although its complement* $\neg f$ *has a nondeterministic syntactic read-once branching program of polynomial size. This, in particular, means that the nonuniform analogue of* $NLOGSPACE = co - NLOGSPACE$ *fails for syntactic read-k times networks with* $k = o(\log n)$. *We also show that (even for* $k = 1$*) the syntactic model is exponentially weaker then more realistic "nonsyntactic" one.*

Résumé. – *Un programme syntaxique arborescent à k lectures, est défini par la restriction qu'aucune variable n'apparaisse plus de k fois le long d'un chemin (consistant ou non). Nous exhibons une fonction booléenne explicite f, qui ne peut pas être calculée par un programme syntaxique arborescent non déterministe de taille inférieure à* $\exp\left(\Omega\left(\frac{\sqrt{n}}{k^{2k}}\right)\right)$, *bien que son complémentaire* $\neg f$ *admette un programme syntaxique non-déterministe arborescent à une unique lecture de taille polynomiale. Ceci signifie en particulier que l'analogue non-uniforme de* $NLOGSPACE = co - NLOGSPACE$ *ne vaut plus pour les réseaux syntaxiques à k lectures où* $k = o(\log n)$. *Nous montrons aussi que (même pour* $k = 1$*), le modèle syntaxique est exponentiellement plus faible que le modèle plus réaliste « non-syntaxique ».*

1. INTRODUCTION

We will consider the classical model of switching-and-rectifier networks together with two its restrictive versions-deterministic and non-deterministic branching programs. Let us briefly recall their definitions. (Basic relationships between these models one can find in the survey [8].)

(*) Received October 21, 1992, revised August, 1, 1993, accepted August 10, 1993.

(1) Present adress: Universität Trier, FB Informatik, 54286 Trier, Germany. Email: jukna©uni-trier.de.

Research done partly while visiting Fachbereich Informatik, Universität Dortmund under the support of Alexander von Humboldt Foundation.

Informatique théorique et Applications/Theoretical Informatics and Applications
0988-3754/95/01/$ 4.00/© AFCET-Gauthier-Villars

A *switching-and-rectifier network* is a directed acyclic multigraph G with a distinguished source node s and a distinguished sink node t. For each non-sink node, each edge directed out of the node is either unlabeled or labeled by some variable or its negation. The size (G) is the number of *labeled* edges in G. The network G computes a Boolean function $f : \{0, 1\}^n \to \{0, 1\}$ in the obvious way: for each $u \in \{0, 1\}^n$ we let $f(u) = 1$ iff there exists at least one (directed) s-t path starting in the source node and leading to the accepting node and such that all labels along this path are consistent with u. Following [8] we denote the minimal possible size of a switching-and-rectifier network computing a Boolean function f by $RS(f)$.

There are several ways to restrict the power of switching-and-rectifier networks. The most restrictive version is the well known model of branching programs. Namely, a *deterministic branching program* is a switching-and-rectifier network in which the outdegree of each non-sink node is exactly 2 and the two outgoing edges are labeled by x_i and $\neg x_i$ for some variable associated with the node. The branching program becomes *nondeterministic* if we allow "guessing nodes" that is nodes with both two outgoing edges being *unlabeled.* The measures corresponding to the size of these devices are denoted by $BP(f)$ and $NBP(f)$.

A network is *syntactic read-k times* if each variable occurs at most k times along *each* path going from s. We denote the corresponding complexity measures by $BP_k(f)$, $NBP_k(f)$ and $RS_k(f)$. We adopt the following notation: for a complexity measure $M(f)$ let $\mathbf{M}$ denote the class of all sequences of Boolean functions $\langle f_n | n \geq 0 \rangle$ for which $M(f_n) = n^{O(1)}$. Then $\mathbf{BP}_k \subseteq \mathbf{NBP}_k \subseteq \mathbf{RS}_k$ and $\mathbf{NBP}_k = \mathbf{RS}_k$, *i.e.* in the case of syntactic restriction the read-k times models of non-deterministic branching programs and switching-and-rectifier networks are equivalent. (We will briefly discuss the "nonsyntactic" case in the last section.)

The model of syntactic read-k times networks was intensively investigated in the last ten years. For small values of k the following separation results were proved (throughout, $\subset$ means strong inclusion): $\mathbf{BP}_1 \subset \mathbf{BP}_\infty$[10, 11]; $\mathbf{BP}_1 \subset \mathbf{BP}_2$ [10]; $\mathbf{NBP}_1 \subset \mathbf{NBP}_\infty$ [6]; $\mathbf{BP}_1 \subset \mathbf{NBP}_1 \subset \mathbf{NBP}_2$ and $\mathbf{NBP}_1 \neq co\text{-}\mathbf{NBP}_1$ [4, 5]. This last inequality was established in [4, 5] by proving that the "Exact-Perfect-Matching" function in not in $\mathbf{NBP}_1$ while its complement is obviously in this class. Another proof of this inequality was recently given in [1] using "Exact-Half-Clique" function.

The progress in the field was made recently by Borodin, Razborov and Smolensky in [1] by proving that $\mathbf{NBP}_k \subset \mathbf{NBP}_n$ for $k \approx \log_2 n$.

This was done using functions $g_n : \mathbf{F}_q^{2n} \to \{0, 1\}$ $(q \geq 3)$ given by: $g_n(x_1, ..., x_n, y_1, ..., y_n) = 1$ iff $\sum_{i,j=1}^{n} a_{i,j} x_i y_j = 0$, where $A = \{a_{i,j}\}$ is an $n \times n$ matrix over the field $\mathbf{F}_q$, and proving that the (Boolean version of) g_n is not in $\mathbf{NBP}_k$ if all sufficiently large minors of A have large rank. The next step was to prove that, so called, Generalized Fourier Transform matrices and, in particular–Sylvester matrices, have this property.

A similar result but for the weaker class of *deterministic* branching programs, namely, the separation $\mathbf{BP}_k \subset \mathbf{BP}_{\sqrt{n}}$ for $k \approx \ln n / \ln \ln n$, was obtained independently by Okolnishnikovà [7]. This was done by proving that the characteristic function of well-known Bose-Chaudhuri codes requires deterministic reak-k times branching program of size $\exp\left(\Omega\left(\frac{\sqrt{n}}{k^k}\right)\right)$. This function is defined by

$$f_{n,d}(x_1, ..., x_n) = \bigwedge_{i=1}^{m} \left(1 \oplus \left(\bigoplus_{j=1}^{n} a_{ij} x_j\right)\right) \tag{1}$$

where $A = \{a_{i,j}\}$ is an $m \times n$ 0, 1-matrix with $m \leq d \log(n+1)$ rows and such that every $2d$ columns of A are linearly independent over $\mathbf{F}_2$. Such matrices are explicitly described in [2].

The function $f_{n,d}$ have one nice property: the complement $\neg f_{n,d}$ is the OR of m parity functions, and hence, is clearly in the class $\mathbf{NBP}_1$.

The goal of this note is to extend the results of [1] and [7] by proving that the function $f_{n,d}$ requires also *nondeterministic* read-k times branching programs of size $\exp\left(\Omega\left(\frac{\sqrt{n}}{k^{2k}}\right)\right)$. Thus, although $\neg f_{n,d} \in \mathbf{NBP}_1$, the function $f_{n,d}$ itself does not belong to $\mathbf{NBP}_k$ if $k \leq k_0 = (1/2 - \varepsilon) \ln n / \ln \ln n$. This fact means that

$$co - \mathbf{NBP}_1 \setminus \mathbf{NBP}_{k_0} \neq \varnothing,$$

and hence, for all $k \leq k_0$

$$\mathbf{NBP}_k \neq co - \mathbf{NBP}_k.$$

In particular, this shows that the Immerman-Szelepcsènyi [3, 9] constructions, yielding the equality $\mathbf{NBP} = co - \mathbf{NBP}$, necessarily require at least logarithmic multiplicity of reading.

Let us also mention that we derive our lower bound for $f_{n,d}$ using only the fact that:

(i) this function accepts sufficiently many vectors, namely, at least $2^n (n+1)^{-d}$, and

(ii) the Hamming distance between any two accepted vectors is also sufficienlty large, namely, at least $2d+1$.

2. THE THEOREM

For a Boolean functin f, let $|f|$ denote the number of vectors in $f^{-1}(1)$ and $H(f)$ denote the minimal Hamming distance between any two vectors in $f^{-1}(1)$.

THEOREM: *Let a, k, d be positive integers, $a \geq k+1$, and let f be a Boolean function in n variables with $H(f) \geq 2d+1$. Then*

$$NBP_k(f) \geq \frac{1}{2}\left(\Delta_{a,k}(f).\frac{|f|}{2^n}\right)^{\frac{1}{2ka}}. \tag{2}$$

where

$$\Delta_{a,k}(f) = \left(\frac{n^2}{d^2 a^k e^{k+1}}\right)^d.$$

We postpone the proof of the theorem to the next section.

The theorem yields large lower bounds for any Boolean function which accepts many vectors with large Hamming distance between them. Thus code functions are good candidates for large lower bounds.

To illustrate this, let us take the characteristic function $f_{n,d}$ of Bose-Chaudhuri code defined by (1). It is well known that for this function we have (*see* [2]): $|f_{n,d}| \geq 2^n (n+1)^{-d}$ and $H(f_{n,d}) \geq 2d+1$. Thus, taking $a = k+1$ in (2) after simple computations we obtain the following

COROLLARY 1: *If $d \leq \sqrt{(n-1)/(2(k+1)^k e^{k+1})}$ then*

$$\mathbf{NBP}_k(f_{n,d}) \geq \exp\left(\Omega\left(\frac{d}{k^2}\right)\right). \tag{3}$$

In particular, for the maximal possible d,

$$NBP_k(f_{n,d}) \geq \exp\left(\Omega\left(\frac{\sqrt{n}}{k^{2k}}\right)\right). \tag{4}$$

COROLLARY 2: *Let* $k_0 = (1/2 - \varepsilon)\ln n/\ln\ln n$, $\varepsilon > 0$. *Then for any* $k = k(n) \leq k_0$ *we have that* $co - \mathbf{NBP}_1 \setminus \mathbf{NBP}_{k_0} \neq \varnothing$ *and hence* $\mathbf{NBP}_k \neq co - \mathbf{NBP}_k$.

Proof: By (4) we have (for appropriate values of $d = d(n)$) that $NPB_{k_0}(f_{n,d}) = \exp(\Omega(n^{\varepsilon}))$. On the other hand, $\neg f_{n,d}$ is the OR of $m \leq d\log(n+1)$ parity functions, and hence, $NBP_1(\neg f_{n,d}) = O(n^2)$. □

3. THE PROOF

First we recall from [1] the following result stating that functions computed by read-k times programs can be represented in some special form. Say that a Boolean function $g(x_1, ..., x_n)$ is a (k, a)-*rectangle* if g can be represented in the form

$$g = \bigwedge_{i=1}^{ka} g_i(X_i)$$

where g_i is a Boolean function depending only on variables from $X_i \subseteq \{x_1, ..., x_n\}$, $|X_i| \leq [n/a]$ and each variable belongs to at most k of the sets $\{X_1, ..., X_{ka}\}$.

LEMMA 3 ([1]): *Let f be a Boolean function and k, a be positive integers. Let* $T = (2NBP_k(f))^{2ka}$. *Then f is an OR of at most* T (k, a)-*rectangles.*

Thus, in order to prove the lower bound (2), it is enough to prove that each (k, a)-rectangle $g \leq f$ can accept at most $2^n/\Delta_{a,k}(f)$ vectors from $f^{-1}(1)$, *i.e.* that $|g| \leq 2^n/\Delta_{a,k}(f)$. We split the proof of this fact into two simple lemmas.

Convention: Throughout this section, let $\alpha = \binom{ak}{k}^{-1}$ and $\beta = 1 - \frac{k}{a}$.

LEMMA 4: *Let* $g(x_1, ..., x_n)$ *be a* (k, a)-*rectangle. Then g can be represented in the form*

$$g = g^0(X^0) \wedge g^1(X^1) \tag{$\star$}$$

where $|X^0 \setminus X^1| \geq \alpha n$ *and* $|X^1 \setminus X^0| \geq \beta n$.

Proof: Let $g = g_1(X_1) \wedge ... \wedge g_m(X_m)$ be a (k, a)-rectangle, $m = ka$. We consider a random subset $\mathbf{I} \subseteq \{1, ..., m\}$ with $|\mathbf{I}| = k$, and associate with it the following two sets of variables: $X^0 = \bigcup_{i \in \mathbf{I}} X_i$ and $X^1 = \bigcup_{j \notin \mathbf{I}} X_j$. For a variable $x \in \{x_1, ..., x_n\}$, put $J_x = \{i | x \in X_i\}$. Since each variable x belongs to at most k of the sets $\{X_1, ..., X_m\}$, we have that $|J_x| \leq k$, and hence, $\Pr[x \in X^0 \setminus X^1] = \Pr[\mathbf{I} \supseteq J_x] \geq \binom{m}{k}^{-1} = \alpha$. This implies that the mean of $|X^0 \setminus X^1|$ is at least $\alpha\eta$. Fix any set I in $\binom{[m]}{k}$ for which $|X^0 \setminus X^1| \geq \alpha n$. Since $|X_i| \leq \lceil n/a \rceil$ for all $i = 1, ..., m$, we have that $|X^0| \leq \lceil n/a \rceil |I|$. Hence $|X^1 \setminus X^0| = n - |X_0| \geq n - \lceil n/a \rceil \cdot k \geq (1 - k/a) n = \beta n$ which completes the proof of the lemma. □

LEMMA 5: *Let* g *be a Boolean function in* n *variables. If* $g \leq f$ *and* g *can be represented in the form* $(\star)$ *then*

$$|g| \leq \frac{2^n}{\triangle_{a,k}(f)}.$$

Proof: Define the r-th degree $D_r(f)$ of a Boolean function f to be the maximal possible number of vectors in $f^{-1}(1)$ such that all of them coincide in at least i coordinates. In other words, $D_r(f)$ is the maximum of $|f_{\sigma \to Y}|$ over all $Y \subseteq X$ with $|Y| = r$ and all assignments $\sigma : Y \to \{0, 1\}$. Hence, $D_0(f) = |f|$ and $D_r(f) \to 1$ as $r \to n$.

Let g have the representation $(\star)$. Take $Y^0 \subseteq X^0 \setminus X^1$ and $Y^1 \subseteq X^1 \setminus X^0$ with $|Y^0| = \alpha n$ and $|Y^1| = \beta n$. Let $Z = X \setminus (Y^0 \cup Y^1)$.

Any assignment $\sigma \to Z$ of constants to variables in Z leads to the subfunction $g_{\sigma \to Z}$ of g which can be represented in the form $g_{\sigma \to Z} = h^0(Y^0) \wedge h^1(Y^1)$ where $Y^0 \cap Y^1 = \varnothing$. Thus, for each assignment $\sigma \to Z$ there are at most $D_{|Y^0|+|Z|}(f) \cdot D_{|Z|+|Y^1|}(f) \leq D_{(1-\beta)n}(f) \cdot D_{(1-\alpha)n}(f)$ vectors in $g^{-1}(1)$ consistent with σ. Since there are exactly $2^{|Z|} = 2^{(1-\alpha-\beta)^n}$ such assignments $\sigma \to Z$, we conclude that

$$|g| \leq 2^{(1-\alpha-\beta)n} D_{(1-\alpha)n}(f) \cdot D_{(1-\beta)n}(f). \tag{5}$$

Next, observe that either $D_r(f) = 1$ (if $r > n - H(f)$) or

$$D_r(f) \leq 2^{n-r} \binom{n-r}{d}^{-1} \tag{6}$$

Indeed, take a set $A \subseteq f^{-1}(1)$ and suppose that all the vectors in A coincide on some set of coordinates $I \subseteq \{1, ..., n\}$, $|I| = r$. Let $A' \subseteq \{0, 1\}^{n-r}$ be the projection of A onto the set of remaining indices $[n] \setminus I$. For each vector $x \in A'$ draw the Hamming ball $B_d(x) \subseteq \{0, 1\}^{n-r}$ of radius d with the center in x. Each such ball has exactly $1 + \sum_{i=1}^{d} \binom{n-r}{i} > \binom{n-r}{d}$ vectors. On the other hand, the condition $H(f) \geq 2d + 1$ means that all these balls must be pairwise disjoint. Since $|A| = |A'| \leq 2^{n-r}$, we obtain the desired upper bound $2^{n-r} \binom{n-r}{d}^{-1}$ on the number of possible balls $B_d(x)$ with $x \in A'$, and hence, the desired upper bound for the number of vectors in A.

Using (6), we have by (5) that $|g| \leq 2^n/N$ where

$$N = \binom{\alpha n}{d}\binom{\beta n}{d} > \left(\frac{\alpha\beta n^2}{d^2}\right)^d \geq \left(\frac{n^2}{d^2 a^k e^{k+k/a}}\right)^d > \Delta_{k,a}(f),$$

which completes the proof of Lemma 5, and thus, the proof of the theorem. □

4. CONCLUDING REMARK

In "syntactic" read-k times networks, each variable is allowed to be tested at most k times in *any* path (consistent or not). This restriction for inconsistent paths is somewhat artificial. In order to capture space limitations in so-called *eraser Turing machines* which erase each input cell after a fixed number k of *readings,* one has to consider "nonsyntactic" read-k times networks, *i.e.* networks in which only *consistent* paths are required to test each variable at most k times (no matter how many times variables appear on inconsistent paths). Namely, say that a switching-and-rectifier network is *read-k times* if each variable occurs at most k times along each *consistent* path going from the source s. Let $k\text{-}BP(f)$, $k\text{-}NBP(f)$ and $k\text{-}RS(f)$ denote the corresponding complexity measures.

Although we have that $1 - \mathbf{BP} = \mathbf{BP}_1$ and $1 - \mathbf{NBP} = \mathbf{NBP}_1 = \mathbf{RS}_1$, the following simple observation shows that nonsyntactic read-k times devices can be much more powerful than syntactic ones and we need new lower bound arguments for them (even for small values of k).

PROPOSITION 6: $\mathbf{BP}_1 \subset \mathbf{RS}_1 \subset 1\text{-}\mathbf{RS}$.

Proof: The first inclusion was established in [4, 5] by the "Exact-Perfect-Matching" function. This is the function $f(X)$ in n^2 variables which, given a $n \times n$ matrix $X = \{x_{i,j} | 1 \leq i, j \leq n\}$, computes 1 iff X is a permutation matrix, *i.e.* iff each row and each column of X has exactly one 1. It is known ([4, 5]) that $RS_1(f) \geq \exp(\Omega(n))$. Since for any function f, $BP_k(\neg f) = BP_k(f)$ and $BP_k(f) \geq RS_k(f)$, we have that $\neg f$ does not belong to $\mathbf{BP}_1$. On the other hand, $\neg f(X) = 1$ iff there is a line (*i.e.* a row or a column) in X which has either no ones or at least two ones. Thus, $\neg f$ belongs to $\mathbf{RS}_1$.

To prove the second inclusion, it is enough to verify that the function f itself has a read once switching-and-rectifier network of polynomial size. Define the network $G(X)$ as the AND of two networks $G_1(X)$ and $G_2(X)$ where

$$G_1(X) = \bigwedge_{i=j}^{n} \bigvee_{j=1}^{n} x_{i,j} \qquad \text{and} \qquad G_2(X) = \bigwedge_{j=1}^{n} \bigvee_{k=1}^{n} \bigwedge_{\substack{i=1 \\ i \neq k}}^{n} \neg x_{i,j}$$

Observe that $G_1(X) = 1$ iff each row of X has at least 1 one, and $G_2(X) = 1$ iff each column of X has at least $n-1$ zero. Thus, $G(X)$ computes $f(X)$ and has size $O(n^3)$. Finally, since G_1 has no edge labelled by a negated variable and all the edges in G_2 are labelled by negated variables, we have that in each consistent path, starting in s, each variable is tested at most once (otherwise the path becomes inconsistent). Hence, G is a (non-syntactic!) read-once switching-and-rectifier network. □

ACKNOWLEDGMENT

I would like to thank Ingo Wegener for helpful comments on the preliminary version of this note.

REFERENCES

1. A. BORODIN, A. RAZBOROV, R. SMOLENSKY, On lower bounds for read-k times branching programs, *Computational Complexity*, 1993, *3*, pp. 1-18.
2. R. C. BOSE, D. K. RAY-CHAUDHURI, On a class of error-correcting binary group codes, *Information and Control,* 1960, *3*, 1, pp. 68-79.

3. N. IMMERMAN, Nondeterministic space is closed under complementation, *SIAM J. Comput.*, 1988, *17*, pp. 935-938.
4. S. JUKNA, The effect of null-chains on the complexity of contact circuits, *Springer Lecture Notes in Computer Science,* 1989, *380,* pp. 246-256.
5. M. KRAUSE, Ch. MEINEL, S. WAACK, Separating the eraser Turing machine classes L_e, NL_e, $co - NL_e$ and P_e, *Theor. Comp. Sci.,* 1991, *86*, pp. 267-275.
6. S. E. KUZNETSOV, The influence of null-chains on the complexity of contact circuits, *Probabilistic Methods in Cybernetics,* 1984, *20*, in Russian.
7. E. A. OKOLNISHNIKOVA, Lower bounds on the complexity of realization of characteristic functions of binary codes by branching programs, *Diskretnii Analiz,* 1991, Novosibirsk, *51*, pp. 61-83, in Russian.
8. A. A. RAZBOROV, Lower bounds for deterministic and nondeterministic branching programs, *Springer Lecture Notes in Computer Science,* 1991, *529*, pp. 47-60.
9. R. SZELEPCÉNYI, The method of forcing for nondeterministic automata, *Bull. European Assoc. Theoret. Comput. Sci.,* 1987, *33*, pp. 96-100.
10. I. WEGENER, *On the complexity of branching programs and decision trees for clique functions*, Internal Rept. 5/84, FB Informatik, Univ. of Frankfurt, 1984 [*see* also: *Journal of the ACM*, 1988, *35*, pp. 461-471].
11. S. ŽAK, An exponential lower bound for one-time-only branching programs, *Springer Lecture Notes in Computer Science,* 1984, *176*, pp. 562-566.