ESANN'1995 proceedings - European Symposium on Artificial Neural Networks
Brussels (Belgium), 19-20-21 April 1995, D-Facto public., ISBN 2-9600049-3-0, pp. 211-216

On Threshold Circuit Depth *

Andreas Albrecht

BerCom GmbH
Bruno-Taut-Str. 4 - 6
D-12524 Berlin, Germany

Abstract. We consider boolean circuits as a discrete model of learning devices such as neural nets. The circuits are of polynomial size with threshold gates of unbounded fan-in. The gates are restricted to realize threshold functions with weights from $\{-1, 0, +1\}$. We prove that concept classes representing n-ary boolean functions $f(\tilde{x})$ with a hyperpolynomial set of quasi-singular tuples $\tilde{\sigma}$ where $f(\tilde{\sigma}) = 1$ cannot be realized by circuits of depth two, that means, at least three layers are necessary for the corresponding learning devices. Among these concept classes are e.g. functions representing simple arithmetical properties. Furthermore, the presented approach is extended to circuits of depth four.

1. Introduction

The interest in constant depth circuits results from the small depth of real nets of neurons (which contain, however, also feedback edges). By *Maass*, *Schnitger*, and *Sontag* [8] it was shown that constant depth sigmoid circuits of polynomial size with polynomially bounded weights compute the same class of boolean functions as the corresponding threshold circuits; there is only a difference for constant size circuits. In order to characterize boolean functions which are realized by constant depth circuits of polynomial size the relation to other complexity classes is investigated, e.g. to $\mathcal{NC}^1$ -the family of boolean circuits with bounded fan-in, gate number $n^{O(1)}$, and depth $O(\log n)$. For a deeper analysis of $\mathcal{NC}^1$ the class $\mathcal{AC}$ was introduced, representing languages accepted by polynomial size, constant depth circuits consisting of NOT gates and unbounded fan-in AND and OR gates (that means gates with fixed thresholds $\vartheta_\vee = 1$ and $\vartheta_\wedge = n$). The subclass of circuits with constant depth O(1) is denoted by $\mathcal{AC}^0$. As a first result *Furst/Saxe/Sipser* [4] proved that parity cannot be computed by circuits from $\mathcal{AC}^0$. The consequence is that depth-bounded circuits for multiplying integers or taking the transitive closure of graphs require more than a polynomial number of AND and OR gates. *J. Håstad* [7] obtained a lower bound $\Omega(\log(n)/\log\log(n))$ for the depth of circuits from $\mathcal{AC}$ realizing the parity function $\oplus_{i=1}^{n} x_i$. *A. Hajnal et al.* [6] separated subclasses from $\mathcal{TC}^0$ of small depth.

*This research has been in part supported by the German Ministry for Research and Technology (BMFT) within the joint project GOSLER under grant no. 01 IW 101.

ESANN'1995 proceedings - European Symposium on Artificial Neural Networks
Brussels (Belgium), 19-20-21 April 1995, D-Facto public., ISBN 2-9600049-3-0, pp. 211-216

In particular, they showed that the inner product mod 2 of two binary vectors of length n, the function $f(\vec{x},\vec{y}) = x_1y_1 \oplus x_2y_2 \oplus ... \oplus x_ny_n$, belongs to depth-three circuits from $\mathcal{TC}^0$, but not to depth-two circuits. In [10] *A.C. Yao* has shown that languages accepted by monotone depth k threshold circuits require exponential size for depth $2 \cdot k$ circuits from $\mathcal{AC}^0$. In contrast to lower bounds *E. Allender* [3] proved that any language accepted by depth k $\mathcal{AC}^0$ circuits is accepted by depth-three threshold circuits of size $n^{O(\log^k n)}$.
We consider boolean circuits of polynomial size with threshold gates of unbounded fan-in, where the gates are restricted to realize threshold functions with weights from $\{-1,+1\}$. That type of circuits is an important class in the theory of neural network synthesis, see e.g. [9]. In [5] *Goldmann/Håstad/Razborov* proved that the depth d class with arbitrary weights is contained in depth $d+1$ with small weights. The present paper is based on the methods and results from [6], [1], and [2].

2. Local Properties of Depth-two Circuits

The set of all n-ary boolean functions is denoted by $\mathcal{B}_n$. We consider threshold circuits whose basic functions are from the set

$$\mathrm{TF}^n_{bin} := \{h : \exists[\tilde{w},\vartheta](\tilde{w} \in \{-1,0,+1\}^n \wedge h(\tilde{\sigma}) = 1 \Leftrightarrow \sum_{i=1}^{n} w^i \cdot \sigma_i \geq \vartheta)\},$$

where $\vartheta \in \mathrm{N}[-(n+1),+(n+1)]$. The inputs to the circuits are from $X_n := \{x_0,...,x_{n-1}\}$ because of the transformation $-1\cdot x_i = -1+1\cdot(1-x_i) = -1+1\cdot\bar{x}_i$. The corresponding class of sequences $\mathrm{F} = \{f_n\}_{n=1}^{\infty}$ of boolean functions which are realized by threshold circuits consisting of a polynomial number $n^{O(1)}$ of gates (functions from TF^n_{bin}, n =1, 2,....) with unbounded fan-in is denoted by $\mathcal{TC}_{bin}$. If for $\mathrm{F} = \{f_n\}_{n=1}^{\infty} \in \mathcal{TC}_{bin}$ exists a constant k as a universal upper bound for the depth of all circuits $C_1, C_2, ..., C_n, ...$, we say that F is realized by bounded depth threshold circuits. The class of sequences F having bounded depth circuits is denoted by $\mathcal{TC}^o_{bin}$.
Given the subsets $\mathrm{W} \subseteq \{-1,0,+1\}^n$ and $\mathrm{T} \subseteq \mathrm{W} \times \mathrm{N}[-(n+1),+(n+1)]$, we denote $\tau := |\mathrm{T}|$, and $\mathrm{TF(T)} \subseteq \mathrm{TF}^{bin}_n$ is the corresponding subset consisting of $g_1,...,g_\tau$. We consider at first the case of depth-two circuits $C_n \in \mathcal{TC}^o_{bin}$, where the input gates of $\{C_n\}_{n=1}^{\infty}$ are from TF(T) and the output gate is of fan-in τ; C_n realizes f_n. For $\tilde{\sigma} \in \{0,1\}^n$ we denote $\|\tilde{\sigma}\| := \sum_{i=1}^{n} \sigma_i$.

Let C^2_n from a sequence of $\mathcal{TC}^o_{bin}$ denote a circuit of depth 2. The output gate G_o has p inputs with positive weights, q inputs with negative weights, and the threshold is ϑ_o, $\mathrm{p}+\mathrm{q} = \tau$. We transform the output gate G_o to a threshold gate G_1 with $\vartheta_1 = \vartheta_o + \mathrm{q}$ and the q input gates with a negative weight are transformed to the opposite inequalities where the thresholds ϑ_i are changed to $-(\vartheta_i - 1)$. These transformations can be performed also for circuits with a depth larger than two. Therefore, one obtains

ESANN'1995 proceedings - European Symposium on Artificial Neural Networks
Brussels (Belgium), 19-20-21 April 1995, D-Facto public., ISBN 2-9600049-3-0, pp. 211-216

Lemma 1 *Circuits from $\mathcal{TC}^0_{bin}$ having $n^{O(1)}$ gates can be transformed to polynomial size threshold circuits of the same depth where only the input gates may have negative weights.*

In the following we consider only circuits having the canonical structure of lemma 1. By $\mathcal{G} := \{ g_i , i = 1, ..., \tau \}$ we denote the set of input gates of C^2_n. Let $\chi^{\tilde{\alpha}}(g)$ be the characteristic function on $\mathcal{G}$ with respect to $\tilde{\alpha}$: $\chi^{\tilde{\alpha}}(g) = 1 \iff g(\tilde{\alpha}) = 1$. Furthermore, we use the notations $\mathcal{V}^m_n := \{ \tilde{\alpha} \mid \| \tilde{\alpha} \| = m \}$, $Q := \{\tilde{\alpha} \mid \tilde{\alpha} \in \mathcal{V}^m_n$ and $f(\tilde{\alpha}) = 1 \}$, and we set $P := \mathcal{V}^m_n \setminus Q$ and $q := | Q |$, $p := | P |$. Let ϑ_1 be the threshold of the output gate of C^2_n. As in [6] we consider the following inequalities :

$$\sum_{\tilde{\alpha} \in Q} \sum_{i=1}^{\tau} \chi^{\tilde{\alpha}}(g_i) \geq q \cdot \vartheta_1, \quad \sum_{\tilde{\beta} \in P} \sum_{i=1}^{\tau} \chi^{\tilde{\beta}}(g_i) \leq p \cdot (\vartheta_1 - 1).$$

From both inequalities we obtain

$$(1) \qquad \sum_{i=1}^{\tau} \Big[\frac{1}{q} \cdot \sum_{\tilde{\alpha} \in Q} \chi^{\tilde{\alpha}}(g_i) - \frac{1}{p} \cdot \sum_{\tilde{\beta} \in P} \chi^{\tilde{\beta}}(g_i) \Big] \geq 1.$$

We are especially interested in sets Q having a distribution in the whole set $\{0,1\}^n$ which induces that the negative part in (1) dominates the part defined by Q. For this reason we consider partitions of $X_n = \{x_0, x_2, ..., x_{n-1}\}$ into three sets X^1, X^2, and X^3, where $| X^1 | = l$, $| X^2 | = r$, and $| X^3 | = k := n - l - r$. The corresponding to X^i parts of $\tilde{\alpha} \in \{0,1\}^n$ are denoted by $\tilde{\alpha}^i$, $i = 1, 2, 3$; furthermore, we set $\| \tilde{\alpha}^1 \| = a$, $\| \tilde{\alpha}^2 \| = b$, and $\| \tilde{\alpha}^3 \| = c$. Given a gate $g(\vartheta) \in \mathcal{G}$ with the threshold ϑ, we have $\sum_{a+b+c=m} \sum_{a \geq \vartheta + b} \binom{l}{a} \cdot \binom{r}{b} \cdot \binom{n-l-r}{m-a-b}$ tuples $\tilde{\alpha}$ from $\mathcal{V}^m_n$ with $\chi^{\tilde{\alpha}}(g) = 1$; the set of $\tilde{\alpha}$ satisfying $a \geq \vartheta + b$ for fixed a and b is denoted by $R(a,b)$. Finally, we denote for a given partition of X_n

$$(2) \qquad M(a,b) := R(a,b) \cap Q.$$

We introduce the following notation :

Definition 1 *The sequence $\mathcal{Q} = \{Q_n\}^{\infty}_{n=1}$, where $Q_n \subseteq \mathcal{V}^m_n$, is called a homogeneous distributed sequence, iff for any $\varphi(n) \longrightarrow \infty$ there exists a function $\psi(n) \longrightarrow \infty$ such that for any partition $[X^1, X^2, X^3]$ and arbitrary conditions $a \geq \vartheta + b$ satisfying*

$$(3) \qquad | R(a,b) | \geq n^{\varphi(n)}$$

it holds

$$\left| \frac{| R(a,b) | - | M(a,b) |}{| M(a,b) |} - \frac{p}{q} \right| \leq \frac{1}{n^{\psi(n)}} \cdot \frac{| R(a,b) | - | M(a,b) |}{| M(a,b) |}.$$

ESANN'1995 proceedings - European Symposium on Artificial Neural Networks
Brussels (Belgium), 19-20-21 April 1995, D-Facto public., ISBN 2-9600049-3-0, pp. 211-216

In fact, we consider sequences of threshold gates g_n, sequences of partitions $[X_n^1, X_n^2, X_n^3]$ e.t.c. But for simplicity we will speak about only a particular gate, partition, and so on. Furthermore, we consider definition 1 w.l.o.g. mainly for particular values of m; that means, we will use definition 1 for arbitrary $m \in [\gamma(n)\,,\, n - \gamma(n)]$ only if it is necessary.
Now, we illustrate definition 1 by a positive and a negative example :
At first we consider as a negative example the inner product $f(\vec{x}, \vec{y}) = x_1y_1 \oplus x_2y_2 \oplus ... \oplus x_ny_n$. Based on standard methods for estimating $\binom{k}{l}$ one can prove that the sequence $\{f_n(\vec{x}, \vec{y})\}_n$ does not satisfy definition 1.
The positive example for homogeneous distributed sequences is based on the set $Q_{f_n^u}$ of natural numbers which are divisible by an odd number $u = 2 \cdot t + 1$, $t \geq 1$:

$$(4) \qquad f_n^u(\tilde{\sigma}) = 1 \iff \sum_{\substack{\sigma_i = 1 \\ \|\tilde{\sigma}_N\| = n/2}} 2^{i-1} = D \cdot u.$$

We take only the simple special case $u = 3$. An arbitrary number 2^j can be represented by $2^j = V \cdot 3 + (-1)^j$. That means, we have the residue 1 for an even j and the residue 2 for an odd j. Given a partition $[X^1, X^2, X^3]$, one has to distinguish only between the odd and even variables in X^i. By an inductive method one can prove

Lemma 2 *The sequence* $Q_3 = \{Q_{f_n^3}(m)\}_{n=\lceil \log 3 \rceil}^{\infty}$ *represents a homogeneous distributed sequence of boolean functions.*

Now, we consider arbitrary homogeneous sequences :

Theorem 1 *If* $F = \{f_n\}_{n=1}^{\infty}$ *is defined by a homogeneous sequence* Q *of hyper-polynomial sets* $Q \subseteq \mathcal{V}_n^m$*, then* F *cannot be realized by depth-two circuits from* $\mathcal{TC}_{bin}^0$.

The proof is based on the inequality (1). As a consequence one obtains

Corollary 1 *Sequences* $Q_3 = \{Q_{f_n^3}\}_{n=\lceil \log 3 \rceil}^{\infty}$ *cannot be realized by circuits from* $\mathcal{TC}_{bin}^0$ *of depth two.*

We note, that the sequence $Q_3(m)$ can be realized by threshold circuits of depth three.

3. Threshold Circuits of Depth Three

We consider threshold circuits C_n^3 of depth three realizing functions $f(x_1, ..., x_n)$. Because of lemma 1 the circuits C_n^3 have starting from the second level only gates with positive weights. Let g_0 denote the output gate of C_n^3. The number of inputs of g_0 is $\tau(g_0) = n^{O(1)}$. Let $h_1, ..., h_{\tau(g_0)}$ denote the gates that are

ESANN'1995 proceedings - European Symposium on Artificial Neural Networks
Brussels (Belgium), 19-20-21 April 1995, D-Facto public., ISBN 2-9600049-3-0, pp. 211-216

connected to g_0; H_i^σ is the set of input tuples producing the output σ at h_i. As before we set $q := \mid Q \mid$ and $p := \mid P \mid = \mid \mathcal{V}_n^m \setminus Q \mid$. We make the following observation :

Lemma 3 *If C_n^3 realizes f_n from Q, then there exists a gate h_{j_0} such that $\mid H_{j_0}^1 \mid \geq O(\frac{q}{n^{O(1)}})$ and*

$$\frac{\mid H_{j_0}^1 \cap Q \mid}{\mid H_{j_0}^1 \cap P \mid} \geq \frac{q}{p} \cdot \left(1 + \frac{p}{\mid H_{j_0}^1 \cap P \mid \cdot n^{O(1)}}\right). \tag{5}$$

One can show that (5) cannot be satisfied if f_n represents a homogeneous distributed sequences Q with hyperpolynomial sets of tuples : We consider a single input gate $g := g_{i_0}$ of h representing a hyperpolynomial number of tuples from H^1. We assume that g has the properties required in definition 1; otherwise the input gates of h could not realize $O(\frac{q}{n^{O(1)}})$ elements of Q. The main goal is to show that a hyperpolynomial number of elements $\tilde{\sigma} \in H^1 \cap Q$, which are equal to 1 on g, forces a defined number of tuples $\tilde{\eta} \in H^1 \cap P$ also to produce the output 1 for the threshold gate g. We use the notations $G^1[Q] := \{\tilde{\alpha} : g(\tilde{\alpha}) = 1 \wedge f_n(\tilde{\alpha}) = 1\}$ and $G^1[P] := \{\tilde{\beta} : g(\tilde{\beta}) = 1 \wedge f_n(\tilde{\beta}) = 0\}$.

Lemma 4 *If g is an input gate representing a hyperpolynomial number of tuples from $H^1 \cap Q$, then*

$$\frac{\mid G^1[Q] \cap H^1 \mid}{\mid G^1[P] \cap H^1 \mid} \leq \frac{q}{p} \cdot \left(1 + \frac{1}{n^{\xi(n)}}\right) \tag{6}$$

is satisfied, where $\xi(n) \rightarrow \infty$ is a (slowly) growing function.

From lemma 3 and lemma 4 follows

Theorem 2 *Sequences of functions $F = \{f_n\}_{n=1}^{\infty}$ which are defined by homogeneous distributed hyperpolynomial sets Q cannot be realized by polynomial threshold circuits of depth three with weights from $\{-1, 0, +1\}$.*

That means, if Q can be realized by depth three circuits, there exist subsets of variables such that a relatively large number of tuples from $R(a, b)$ satisfies defined threshold equalities. These properties can be used e.g. for learning procedures of functions having depth three circuits.

4. Concluding Remarks

For learning circuits with binary weights we described a method for obtaining lower bounds for circuit depth which is based on local properties of the function. These properties can be used for the design of learning algorithms, e.g. in the case of discrete *pac* algorithms. As candidates for homogeneous distributed sequences of depth at least four we see the elementary operation $div_n(\tilde{\alpha}, \tilde{\beta})$ of division and the function f_n^{prim} representing prime numbers. It was shown by

ESANN'1995 proceedings - European Symposium on Artificial Neural Networks
Brussels (Belgium), 19-20-21 April 1995, D-Facto public., ISBN 2-9600049-3-0, pp. 211-216

L. Adleman in in 1978 that f_n^{prim} can be realized by circuits of polynomial size $O(n^c)$, where c is a small number not exceeding 3. The number $\pi(N)$ of prime numbers $p \leq N$ and the distribution of prime numbers within natural numbers are well studied.

The presented approach cannot be extended immediately to circuits of depth five or a larger depth, because after the first decomposition (from depth four to depth three) it seems to be difficult to ensure the relation $\frac{p}{q} \cdot (1 \pm \frac{1}{n^{\varphi(n)}})$ for the decomposition from depth three to depth two.

We think that lower bounds for threshold circuits provide a better insight into the computational power of neural nets from organic structures.

References

[1] A. Albrecht. On the complexity of learning circuits. *IIR*, 5(12):29 – 35, 1989.

[2] A. Albrecht. On bounded-depth threshold circuits for pattern functions. In Proc. of The International Conference on Artificial Neural Networks; *ICANN'92*, pages 135 – 138, Brighton, 1992.

[3] E. Allender. A note on the power of threshold circuits. Technical Report TR-5, Univ. of Würzburg, Inst. of Informatics, July 1989.

[4] M. Furst, J.B. Saxe, and M. Sipser. Parity, circuits, and the polynomial-time hierarchy. In *Proc. 22nd IEEE Symp. on Foundations of Computer Science*, pages 260 – 270, 1981.

[5] M. Goldmann, J. Håstad, and A. Razborov. Majority gates vs. general weighted threshold gates. In *Proc. 7th IEEE Structure in Complexity Theory Conf.*, 1992.

[6] A. Hajnal, W. Maass, P. Pudlak, M. Szegedy, and G. Turan. Threshold functions of bounded depth. In *Proc. 28 IEEE Symp. on Foundations of Computer Science*, pages 99 – 110, 1987.

[7] J. Håstad. Almost optimal lower bounds for small depth circuits. In *Proc. of the 18th ACM Symp. on Theory of Computing*, pages 6 – 20, 1986.

[8] W. Maass, G. Schnitger, and E. Sontag. On the computational power of sigmoid versus boolean threshold circuits. In *Proc. of the 32nd IEEE Symp. on Foundations of Computer Science*, pages 767 – 776, 1991.

[9] Ph. Treleaven. Neurocomputers. Technical Report CS-RN-89-8, University College London, 1989.

[10] A.C. Yao. Circuits and local computation. In *Proc. of the 21st ACM Symp. on Theory of Computing*, pages 186 – 196, 1989.