

On the size of weights for threshold gates

Johan Håstad
Royal Institute of Technology
100 44 Stockholm, SWEDEN

Abstract

We prove that if n is a power of 2 then there is a threshold function that on n inputs that requires weights of size around $2^{(n \log n)/2-n}$. This almost matches the known upper bounds.

keywords: Threshold function, computational complexity, neural nets.

AMS classification: 68R05

Warning: Essentially this paper has been published in SIAM Journal on Discrete Mathematics and is hence subject to copyright restrictions. It is for personal use only.

1 Introduction

One very interesting computational element is that of a threshold gate. A threshold gate of n inputs is specified by a set of *weights* $w_1, w_2 \dots w_n$ and a *threshold* t . On input $x_1, x_2 \dots x_n$ it outputs $\text{sign}(\sum_{i=1}^n w_i x_i - t)$. In this notation we assume that the gate computes a function $\{-1, 1\}^n \mapsto \{-1, 1\}$, but the set $\{-1, 1\}$ could be replaced by any two element set (e.g. by $\{0, 1\}$).

Threshold circuits, i.e. circuits that contain threshold gates have been studied extensively. On the more theoretical side one has established many upper and lower bounds on the power of small depth threshold circuits. For a discussion of these results we refer to [1] and its references. It is striking to note that there are no known strong lower bounds for general depth 2 threshold circuits. On the more applied side we note that threshold circuits have many similarities with neural networks. We refer to [2, 5, 6] for more information on these connections and the area in general.

For both type of investigations mentioned above it is important to understand what conditions can be put on the weights w_i . Since some finite amount of precision is always sufficient it is easy to see that we can assume that the weights are integers. Furthermore, it has been proved many times (one early source is [4]) that if we have a function with n inputs then $|w_i| \leq 2^{-n}(n+1)^{(n+1)/2}$ is sufficient. Since there are at least $2^{n^2/2}$ different threshold functions ([4],[8],[9]), there are some functions that require $\max |w_i|$ on the order of $2^{n/2}$. On the other hand there are at most 2^{n^2} threshold functions [3], [7] and hence these are the essentially the best bounds that can be proved by this type of simple counting.

There are also known explicit functions which require weights of size at least c^n for some constant $c > 1$. In particular, if we let the input encode two numbers in binary and ask which number is the greater, then a lower bound of essentially $2^{n/2}$ holds. There are other, slightly more complicated, explicit functions giving a value of c up to $\frac{1+\sqrt{5}}{2}$ [6].

The above mentioned bounds imply that $\Omega(n)$ bits are sometimes needed and $O(n \log n)$ bits are always sufficient to specify the individual weights. The gap between these two bounds are not substantial enough to matter greatly in arguments in complexity theory. The reason for this is that one in general is only interested whether the quantities are polynomial in size. However the gap is rather large and the goal of the this paper is to bring the two bounds closer together. We do this by improving the lower bound for an explicit function F_n . We prove that when $n \geq 8$ is a power of 2, this function requires $|w_i| \geq \frac{1}{2n} e^{-4n^\beta} 2^{(n \log n)/2-n}$ for all i , where $\beta = \log_2 \frac{3}{2}$. Comparing to the known upper bounds we see that this is essentially tight.

The outline of the paper is as follows. In Section 2 we define our function and prove the lower bound on the size of the weights needed to realize this function. In Section 3 we recall the proof for the upper bound on the weights and we end by some final comments in Section 4.

2 A function requiring large weights

Let us assume that n is a power of two and that $n = 2^m$. We will use $\{-1, 1\}$ notation throughout and we will think of vectors in $\{-1, 1\}^n$ as functions from $\{-1, 1\}^m$ to $\{-1, 1\}$. This convention makes us use two types of functions. Those on m variables and those on n variables. We will use the former as inputs to the latter. To decrease the possibility of confusion we reserve capital letters for functions on n inputs.

For $\alpha \subseteq [m] = \{1, 2, \dots, m\}$ let φ_α be the character function i.e. $\varphi_\alpha(x) = \prod_{i \in \alpha} x_i$ where we let φ_\emptyset be the function which is identically 1. Choose an ordering of $\alpha_0, \alpha_1 \dots \alpha_{n-1}$ of the sets such that:

1. $|\alpha_i| \leq |\alpha_j|$ for $i \leq j$.
2. $|\alpha_i \Delta \alpha_{i+1}| \leq 2$ for all i , where $|\alpha_i \Delta \alpha_{i+1}|$ is the symmetric difference of the two sets α_i and α_{i+1} .

This implies that $\alpha_0 = \emptyset$ and $\alpha_1, \alpha_2 \dots \alpha_m$ are the singletons and that $|\alpha_i \Delta \alpha_{i+1}| = 1$ when $|\alpha_{i+1}| = |\alpha_i| + 1$ while $|\alpha_i \Delta \alpha_{i+1}| = 2$ otherwise.

Lemma 2.1 *There is an ordering that satisfies conditions 1 and 2 above.*

Proof: Assume that we have an ordering of all sets containing at most d elements satisfying conditions 1 and 2, and also an ordering of the set with $d + 1$ elements which satisfies condition 2. If we think of these orderings as lists, we can concatenate them. This might create an illegal ordering in that condition 2 might not be satisfied when $|\alpha_i| = d$ and $|\alpha_{i+1}| = d + 1$. However, if we just permute the names of the elements in the sets of size $d + 1$ we can take care of this condition and the concatenated list will satisfy both conditions.

The above reasoning implies that the only problem is to find an ordering of the d -element subsets of $[m]$ for any $d \leq m$. We prove this by an induction over d and m . The base cases $d = 1$ and $d = m$ are obvious. In the general case we first list all the sets of size d containing m and then all the other sets. This first part of the list is essentially a list of all $(d - 1)$ -element subsets from $[m - 1]$, while the second part consists of all d -element subsets of the same set. Both can be given appropriate orderings by induction and thus the only problem is the connection between the two sublists. However, as above, by possibly permuting the names of the elements in the second list, this connection can be made to satisfy condition 2. This completes the induction step and hence the lemma follows. ■

Let (f, g) denote the inner product of the functions f and g . Define $F(f) : \{-1, 1\}^n \mapsto \{-1, 1\}$ as $sign((f, \varphi_{\alpha_i}))$ where i is the largest index such that $(f, \varphi_{\alpha_i}) \neq 0$. This function is a threshold function since

$$F(f) = sign \left(\sum_{i=0}^{n-1} (n+1)^i (f, \varphi_{\alpha_i}) \right)$$

and (f, φ_{α_i}) is a linear function in the values $f(j)$. This expression is correct since $|(f, \varphi_{\alpha_i})| \leq n$ for all i .

We want to prove that if

$$F(f) = \text{sign} \left(\sum_{j=0}^{n-1} w_j f(j) - t \right)$$

then one of w_j is large. First let us observe something easy:

Lemma 2.2 *We can assume that $t = 0$.*

Proof: Note that $F(f) = -F(-f)$. Hence $|\sum_{j=0}^{n-1} w_j f(j)| > |t|$ for any f and we can set $t = 0$ without changing the function. ■

It will be easier to work with expressions of the form

$$\text{sign} \left(\sum_{i=0}^{n-1} w'_i(f, \varphi_{\alpha_i}) \right).$$

This can be done by the following lemma:

Lemma 2.3 *We have*

$$\sum_{j=0}^{n-1} w_j f(j) = \sum_{i=0}^{n-1} w'_i(f, \varphi_{\alpha_i})$$

for all f iff

$$w'_i = \frac{1}{n} \sum_{j=0}^{n-1} w_j \varphi_{\alpha_i}(j)$$

and

$$w_j = \sum_{i=0}^{n-1} w'_i \varphi_{\alpha_i}(j).$$

Proof: The second statement follows from rearranging the terms. To see the first note that by Fourier inversion

$$f(j) = \frac{1}{n} \sum_{i=0}^{n-1} \varphi_{\alpha_i}(j) (f, \varphi_{\alpha_i}).$$

This implies

$$\sum_{j=0}^{n-1} w_j f(j) = \sum_{j=0}^{n-1} w_j \frac{1}{n} \sum_{i=0}^{n-1} \varphi_{\alpha_i}(j) (f, \varphi_{\alpha_i}) = \frac{1}{n} \sum_{i=0}^{n-1} (f, \varphi_{\alpha_i}) \sum_{j=0}^{n-1} w_j \varphi_{\alpha_i}(j)$$

and the lemma follows. ■

We will establish that F requires large weights and in particular we have.

Theorem 2.4 *Assume that n is a power of 2 and*

$$F(f) = \text{sign} \left(\sum_{i=0}^{n-1} w_i(f, \varphi_{\alpha_i}) \right)$$

where w_i are integers, then $w_{n-1} \geq e^{-4n^\beta} 2^{(n \log n)/2-n}$, where $\beta = \log(3/2)$.

Before we prove Theorem 2.4 let us first deduce the corresponding result when using the normal representation of threshold gates.

Theorem 2.5 *Assume that n is a power of 2 and*

$$F(f) = \text{sign} \left(\sum_{j=0}^{n-1} w_j f(j) \right)$$

where w_j are integers, then for some j , we have $|w_j| \geq \frac{1}{n} e^{-4n^\beta} 2^{(n \log n)/2-n}$, where $\beta = \log(3/2)$.

Proof: (Of Theorem 2.5) If we have an expansion of the kind given in the theorem then by Lemma 2.3 setting

$$w'_i = \frac{1}{n} \sum_{j=0}^{n-1} w_j \varphi_{\alpha_i}(j)$$

we convert it to an expansion of the form considered in Theorem 2.4. The w'_i might not be integers, but nw'_i are integers for all i . Multiplying every weight by the same integer does not change the function and hence by Theorem 2.4 $nw_{n-1} \geq e^{-4n^\beta} 2^{(n \log n)/2-n}$. This is equivalent to saying that

$$\sum_{j=0}^{n-1} w_j \varphi_{\alpha_i}(j) \geq e^{-4n^\beta} 2^{(n \log n)/2-n}$$

and the theorem follows. ■

Let us now prove Theorem 2.4.

Proof: (Of Theorem 2.4) Let us start by an easy observation.

Lemma 2.6 *For any i , $w_i > 0$.*

Proof: This follows from setting $f = \varphi_{\alpha_i}$ and noting that $F(f) = 1$ for such f . ■

By choosing a suitable sequence of test functions f we will prove that w_i have to grow exponentially. Since the ordering of the α_i is not explicit we will sometimes use the notation w_α which should be read as w_i where i is chosen such that $\alpha_i = \alpha$.

Lemma 2.7 *Suppose $|\alpha_{i+1}| = |\alpha_i| = k$ where $2 \leq k \leq n - 1$ and that $\alpha_i \Delta \alpha_{i+1} = \{a, b\}$. Let $v \in \{-1, 1\}^m$ be any point with $v_a = v_b$, then*

$$w_{i+1} > (2^{k-1} - 1)w_i - \sum' \varphi_{\alpha_i}(v)\varphi_\alpha(v)w_\alpha$$

where the sum extends over all α such that $\alpha \subset \alpha_i \cup \alpha_{i+1}$ and α contains exactly one of a and b and α is not equal to α_i or α_{i+1} .

Proof: Let us assume that $\alpha_i = \{1, 2 \dots k\}$ and $\alpha_{i+1} = \{1, 2 \dots k-1, k+1\}$. Let v^1 be a vector of length $k+1$ which satisfies $v_j^1 = v_j$ for $1 \leq j \leq k+1$. Furthermore let v^2 be a similar vector such that $v_j^2 = v_j$ for $j < k$ and $v_j^2 = -v_j$ for $j = k$ and $j = k+1$. Define the following function on the first $k+1$ variables:

$$f(w) = \begin{cases} \varphi_{\alpha_i}(w) & \text{if } w = v^1 \text{ or } w = v^2 \\ -\varphi_{\alpha_i}(w) & \text{otherwise} \end{cases}$$

We extend f to a function of m variables by ignoring the rest of the variables. First note that $(f, \varphi_\alpha) = 0$ for any α that contains an element larger than $k+1$. We have that $(f, \varphi_{\alpha_i}) = 2^{m-k-1}(4 - 2^{k+1})$ while for other $\alpha \subseteq \{1, 2 \dots k+1\}$ we have

$$\begin{aligned} (f, \varphi_\alpha) &= (-\varphi_{\alpha_i}, \varphi_\alpha) + 2^{m-k} \varphi_{\alpha_i}(v^1) \varphi_\alpha(v^1) + 2^{m-k} \varphi_{\alpha_i}(v^2) \varphi_\alpha(v^2) \\ &= 2^{m-k} \varphi_{\alpha_i}(v^1) \varphi_\alpha(v^1) (1 + \varphi_{\alpha_i}(v^1 v^2) \varphi_\alpha(v^1 v^2)) \end{aligned}$$

where $v^1 v^2$ is the pointwise product of v^1 and v^2 . Since this vector is 1 except for coordinates k and $k+1$ we get a nonzero inner product iff $\alpha \subset \{1, 2, \dots k+1\}$ and α contains exactly one of the elements k and $k+1$. The only two sets of size k with these properties are α_{i+1} and α_i while all other sets with these properties have cardinality at most $k-1$. Note that by property 1 of our ordering all these sets appear before α_i and this implies that $F(f) = \text{sign}((\varphi_{\alpha_{i+1}}, f)) = 1$. Writing this statement as an inequality of the weights yields the inequality of the lemma. \blacksquare

Next we have

Lemma 2.8 *Suppose $|\alpha_{i+1}| = 1 + |\alpha_i| = k$ where $2 \leq k \leq n - 1$ and $\alpha_{i+1} = \alpha_i \cup \{a\}$, then for any vector v with $v_a = 1$ we have*

$$w_{i+1} > (2^{k-1} - 1)w_i - \sum_{\alpha \subset \alpha_{i+1}, \alpha \neq \alpha_i} \varphi_{\alpha_i}(v)\varphi_{\alpha}(v)w_{\alpha}$$

Proof: Let us assume that $\alpha_i = \{1, 2 \dots k - 1\}$ and $\alpha_{i+1} = \{1, 2 \dots, k\}$ and let v^1 be the vector of length k which satisfies $v_j^1 = v_j$ for $1 \leq j \leq k$. Define the following function on the first k variables:

$$f(w) = \begin{cases} \varphi_{\alpha_i}(w) & \text{if } w = v^1 \\ -\varphi_{\alpha_i}(w) & \text{otherwise} \end{cases}$$

Extend f to a function of m variables by ignoring the rest of the variables. Again $(f, \varphi_{\alpha}) = 0$ for any α that contains an element larger than k . Clearly $(f, \varphi_{\alpha_i}) = 2^{m-k}(2 - 2^k)$ while for other $\alpha \subset \{1, 2 \dots k\}$ we have

$$(f, \varphi_{\alpha}) = (-\varphi_{\alpha_i}, \varphi_{\alpha}) + 2^{m+1-k}\varphi_{\alpha_i}(v^1)\varphi_{\alpha}(v^1) = 2^{m+1-k}\varphi_{\alpha_i}(v^1)\varphi_{\alpha}(v^1)$$

Again $F(f) = \text{sign}((\varphi_{\alpha_{i+1}}, f)) = 1$ and writing out the corresponding inequality for the weights gives the lemma. ■

Using the above two lemmas we will establish the main lemma for the proof of Theorem 2.4.

Lemma 2.9 *For each i such that $|\alpha_{i+1}| \geq 2$ we have $w_{i+1} > (2^{|\alpha_{i+1}|-1} - 1)w_i$. Furthermore if $\alpha = \{a, b\}$ then*

$$w_{\alpha} > w_{\{a\}} + w_{\{b\}} + w_0.$$

Proof: We establish the lemma by induction over i and we need to handle the cases with $|\alpha_{i+1}|$ small separately.

Let us first establish the lower bounds for w_{α} when $|\alpha| = 2$. Suppose $a = 1$ and $b = 2$ and construct the function of the proof of Lemma 2.8 when $\alpha_{i+1} = \alpha$, $\alpha_i = \{1\}$ and $v^1 = (-1, 1)$. This shows that

$$w_{\alpha} > w_{\{1\}} + w_{\{2\}} + w_0.$$

Let us now establish the lemma when $|\alpha_{i+1}| = |\alpha_i| = 2$. Suppose $\alpha_{i+1} = \{1, 3\}$ and $\alpha_i = \{1, 2\}$. Then apply Lemma 2.7 with $v^1 = (-1, 1, 1)$. This gives

$$w_{i+1} > w_i + w_{\{2\}} + w_{\{3\}}.$$

Next suppose $|\alpha_{i+1}| = 3$ and $|\alpha_i| = 2$. We might assume that $\alpha_{i+1} = \{1, 2, 3\}$ and $\alpha_i = \{1, 2\}$. Apply Lemma 2.8 with $v^1 = (-1, -1, 1)$. This gives

$$w_{i+1} > 3w_i + w_{\{1,3\}} + w_{\{2,3\}} + w_{\{1\}} + w_{\{2\}} - w_{\{3\}} - w_0 > 3w_i$$

where the last inequality follows from the already established bounds.

Let us now take the case $|\alpha_{i+1}| = |\alpha_i| = 3$, where we assume that $\alpha_{i+1} = \{1, 2, 4\}$ and $\alpha_i = \{1, 2, 3\}$. Apply Lemma 2.7 with $v^1 = (-1, -1, 1, 1)$. This gives

$$w_{i+1} > 3w_i + w_{\{1,3\}} + w_{\{2,3\}} + w_{\{1,4\}} + w_{\{2,4\}} - w_{\{3\}} - w_{\{4\}} > 3w_i.$$

Now consider the case when $|\alpha_{i+1}| = k$ and $|\alpha_i| = k - 1$ where $k \geq 4$. We can assume that $\alpha_{i+1} = \{1, 2, \dots, k\}$ and $\alpha_i = \{1, \dots, k - 1\}$. Suppose that α_j is the proper subset of α_{i+1} other than α_i that has the highest index. Choose a vector v with $v_k = 1$ such that $\varphi_{\alpha_i}(v)\varphi_{\alpha_j}(v) = -1$. Now apply Lemma 2.8 with this v . We have

$$w_{i+1} > (2^{k-1} - 1)w_i + w_j - \sum_{\alpha \subset \alpha_{i+1}, \alpha \neq \alpha_j, \alpha_i} \varphi_{\alpha_i}(v)\varphi_{\alpha}(v)w_{\alpha}$$

Thus the lemma will follow from establishing

$$\sum_{\alpha \subset \alpha_{i+1}, \alpha \neq \alpha_j, \alpha_i} |w_{\alpha}| \leq w_j.$$

We divide the sum into those α of size at least 3 and those of size at most 2. To bound the first sum we note that since $w_{l+1} > 3w_l$ when $|\alpha_{l+1}| \geq 3$ and $l < j$, even the sum over all sets of size at least 3 and index less than j is bounded by $w_j/2$.

To bound the second sum we observe that there are at most $k(k-1)/2 + k + 1$ terms and each is bounded the maximal weight of a set of size 2. Now there are at least $k-2$ sets of size 3 before α_j in the enumeration (this bound is tight for $k = 4$ but very weak otherwise) which implies by induction that w_j is at least 3^{k-1} times the maximal weight of any set of size 2. The inequality $3^{k-1} > k(k-1) + 2k + 2$ valid for $k \geq 4$ concludes this case.

Finally suppose $|\alpha_{i+1}| = |\alpha_i| = k$ and $k \geq 4$. We assume that $\alpha_{i+1} = \{1, 2, \dots, k-1, k+1\}$ and $\alpha_i = \{1, \dots, k\}$. Suppose that α_j is the set of highest index that appears in the sum of Lemma 2.7. Choose a vector v with $v_k = v_{k+1} = 1$ such that $\varphi_{\alpha_i}(v)\varphi_{\alpha_j}(v) = -1$. Now apply Lemma 2.7 with this v . The analysis is similar to the last case. This finishes the proof of lemma. \blacksquare

All that remains to prove Theorem 2.4 is a simple calculation. Let α_{i_0} be the last set of size 2 in our ordering. Then

$$w_{n-1} \geq \prod_{|\alpha_i| > 2} (2^{|\alpha_i|-1} - 1) w_{i_0} > 2^{\sum_{i=1}^{n-1} |\alpha_i|-1} \prod_{|\alpha_i| \geq 2} (1 - 2^{1-|\alpha_i|}),$$

since $w_{i_0} > 1$ and the two factors introduced when $|\alpha_i| = 2$ cancel each other.

The first factor equals $2^{nm/2+1-n}$. This follows since the average size of a subset of $[m]$ is $m/2$ and that there are n such sets. The extra “ $1 - n$ ” is the result of -1 inside the summation. To estimate the second factor we use that when $0 < x < \frac{1}{2}$ then $(1 - x) > e^{-2x}$ and hence

$$\prod_{|\alpha_i| \geq 2} (1 - 2^{1-|\alpha_i|}) > e^{-\sum_{k=2}^n \binom{m}{k} 2^{2-k}} \geq e^{-4(1+\frac{1}{2})^m} = e^{-4n^\beta}.$$

and the theorem follows. \blacksquare

Please note that by slightly extra work, the constant in front of n^β can be reduced to any value greater than 2. In fact if we were willing to get an extra term we could in fact get the value 2. This would be achieved by using an inequality of the type $1 - x \geq e^{-x-cx^2}$ for an appropriate constant c . We do not think this is of great concern unless the upper bound is improved.

If we use the full strength of Lemma 2.9 then we can actually strengthen Theorem 2.5 to apply to all weights.

Theorem 2.10 *Assume that n is a power of 2 and*

$$F(f) = \text{sign} \left(\sum_{j=0}^{n-1} w_j f(j) \right)$$

where w_j are integers, then for $n \geq 8$ and all j , we have $|w_j| \geq \frac{1}{2n} e^{-4n^\beta} 2^{(n \log n)/2-n}$, where $\beta = \log(3/2)$.

Proof: Use Lemma 2.3 to get a corresponding expansion

$$\sum_{i=0}^{n-1} w'_i(f, \varphi_{\alpha_i}),$$

where

$$w'_i = \frac{1}{n} \sum_{j=0}^{n-1} w_j \varphi_{\alpha_i}(j)$$

We know by Lemma 2.9 that $w'_{n-1} \geq (2^{m-1} - 1)w'_{n-2} \geq (2^{m-1} - 1)(2^{m-2} - 1)w'_{n-3} \dots$, where $n = 2^m$. By Lemma 2.3

$$w_j = \sum_{i=0}^{n-1} w'_i \rho_{\alpha_i}(j)$$

and for $n \geq 8$,

$$\left| \sum_{i=0}^{n-2} w'_i \rho_{\alpha_i}(j) \right| \leq \frac{1}{2} w'_{n-1}$$

and since

$$w'_{n-1} \geq \frac{1}{n} e^{-4n^\beta} 2^{(n \log n)/2-n}$$

the theorem follows. ■

3 Recalling the upper bound

For completeness, let us recall the proof for the upper bound. Let F be any threshold function and let H_0 be a linear function with the following properties.

1. $\text{sign}(H_0(x)) = F(x)$ for $x \in \{-1, 1\}^n$
2. $|H_0(x)| \geq 1$ for $x \in \{-1, 1\}^n$
3. Among H satisfying the above conditions it maximizes the number of $x \in \{-1, 1\}^n$ such that $|H_0(x)| = 1$. If there are several H giving the same number of such points, choose one arbitrarily.

Since F can be represented by a linear threshold, there is some linear function satisfying the first two conditions and thus there will exist such a linear function H_0 .

Let $x^{(i)}$, $i = 1, 2, \dots, r$ be the set of points in $\{-1, 1\}^n$ with $|H_0(x^{(i)})| = 1$. We claim that H_0 is uniquely determined by the equations

$$H_0(x^{(i)}) = F(x^{(i)}), i = 1, 2, \dots, r.$$

Suppose this was not the case. The set of solutions to these equations would contain linear functions $H_0 + tH_1$ for all rational t and nonzero H_1 . There is some point $x^{(0)} \in \{-1, 1\}^n$ such that $H_1(x^{(0)}) \neq 0$. Suppose

for concreteness that $H_1(x^{(0)}) > 0$ and that $H_0(x^{(0)}) < -1$ (note that we cannot have $|H_0(x^{(0)})| = 1$ since we must have $H_1(x^{(i)}) = 0$ for $1 \leq i \leq r$). Now let t_0 be the minimal $t > 0$ such that $|H_0(x) + tH_1(x)| = 1$ for some $x \notin \{x^{(1)}, x^{(2)}, \dots, x^{(r)}\}$. There must be such a t_0 , since $t = \frac{-1-H_0(x^{(0)})}{H_1(x^{(0)})}$ will give $|H_0(x^{(0)}) + tH_1(x^{(0)})| = 1$. Since $|H_0(x) + tH_1(x)| \geq 1$ for all $t \in [0, t_0]$ and all $x \in \{-1, 1\}^n$ it follows that the two first conditions are satisfied by $H_0 + t_0H_1$. This implies that we violate the maximality condition used to define H_0 and we have reached a contradiction.

By the claim, the coefficients of H_0 can be obtained by solving a linear system of equations where the coefficients and the right hand side belong to $\{-1, 1\}$. By Cramer's rule this means that each coefficient of H_0 is given by the ratio of two $(n+1) \times (n+1)$ (remember that H_0 has a constant coefficient) determinants with entries in $\{-1, 1\}$. Every coefficient has the same denominator and hence we can clear it. By Hadamard's inequality each absolute value of a determinant of the above type is bounded by $(n+1)^{(n+1)/2}$. Thus F can be realized with integer weights of absolute value at most $(n+1)^{(n+1)/2}$. To get a better bound we need the following lemma:

Lemma 3.1 *The determinant of an $m \times m$ matrix which has entries from the set $\{-1, 1\}$ is divisible by 2^{m-1} .*

Proof: Add the first row to each other row. Now these rows will consist of elements from $\{-2, 0, 2\}$. The determinant of this matrix is clearly divisible by 2^{m-1} . ■

Using the lemma and clearing the common factor 2^n gives:

Theorem 3.2 *A threshold function of n variables can be realized with integer weights of size at most $2^{-n}(n+1)^{(n+1)/2}$.*

4 Final discussion

When n is a power of two we have established upper and lower bounds that are only a subexponential factor apart. It is interesting to note that we do not know how to establish as sharp bounds when n is not a power of 2. It is not clear that this is an important problem to determining the true bounds for every n . After all, taking the function F for the largest power of 2 less than n will give fairly good lower bounds. However it is one of

these problems where we do get much better bounds for special values of the parameter.

One natural way to try to prove the lower bounds given by Theorem 2.4 is to try to establish that a random threshold function require large weights. In view of the fact that there are only 2^{n^2} threshold functions it is not clear that this could succeed. One natural question that arises is how to define a random threshold function.

One definition is to pick a random point $(w_1, w_2 \dots w_n)$ uniformly from the real n -dimensional sphere ($\sum_{i=1}^n w_i^2 = 1$) and then define the random function to be $sign(\sum_{i=1}^n w_i x_i)$.

It is not hard to see that with very high probability we can replace w_i by integers with $O(n)$ bits and keep the same function. It is not clear to me that this is due to a deficiency in the definition or that this is the typical behavior of threshold functions.

Let us finally note that our results extend to the case when the inputs are from the set $\{0, 1, \dots a\}$ for $a \geq 2$ (or the more symmetric range $\{-a, 2-a, 4-a, \dots a\}$). We can use the same definition of the function and the proof extends essentially word by word. The only part that does not seem to have a counterpart is Lemma 3.1. We thus get a lower bounds which is roughly a factor a^n stronger, and an upper bound which is a factor $2^n a^{n+1}$ worse.

Acknowledgment: I thank Pekka Orponen for a stimulating conversation on this topic. I am also grateful to Ilan Newman, Ian Parberry, and Sasha Razborov for helpful comments.

References

- [1] M. Goldmann, J. Håstad, and A. Razborov. Majority gates vs. general weighted threshold gates. In *Proceedings of the 1992 Structures conference*, pages 2–13, 1992.
- [2] J. Hertz, R. Krogh, and A. Palmer. *An Introduction to the Theory of Neural Computation*. Addison-Wesley, 1991.
- [3] P. M. Lewis and C. L. Coates. *Threshold logic*. John Wiley & Sons, Inc, 1967.
- [4] S. Muroga. *Threshold logic and its applications*. Wiley-Interscience, 1971.

- [5] P. Orponen. Neural networks and complexity theory (invited talk). In *Proceedings of the 17th International Symposium on Mathematical Foundations of Computer Science*, pages 50–61, 1992.
- [6] I. Parberry. *The computational and learning complexity of neural networks*. MIT Press, In Preparation.
- [7] V. Roychowdhury, K-Y Siu, A Orlisky, and Kailath T. A geometric approach to threshold circuit complexity. In *Proceedings of 4th annual workshop on computational learning*, pages 97–111, 1991.
- [8] D. R. Smith. Bounds on the number of threshold functions. *IEEE Trans. Electronic Computers*, EC-15:368–369, 1966.
- [9] S. Yajima and T. Ibaraki. A lower bound on the number of threshold functions. *IEEE Trans. Electronic Computers*, EC-14:926–929, 1965.