Probabilistic Construction of Monotone Formulae for

Positive Linear Threshold Functions

Rocco A. Servedio*

Division of Engineering and Applied Sciences

Harvard University

Cambridge, MA 02138

October 23, 1999

Abstract

We extend Valiant’s construction of monotone formulae for the majority function to obtain an efficient probabilistic construction of small monotone formulae for arbitrary positive linear threshold functions. We show that any positive linear threshold function on $n$ boolean variables which has weight complexity $q(n)$ can be computed by a monotone boolean formula of size $O(q(n)^3 n^2)$. Our technique also yields a probabilistic construction of small monotone formulae which compute large margin approximators to arbitrary positive linear threshold functions.

Keywords: formula complexity, linear threshold function, computational complexity

*Supported in part by an NSF Graduate Fellowship, by NSF grant CCR-95-04436 and by ONR grant N00014-96-1-0550.
1 Introduction

Positive linear threshold functions are important objects of study in computational learning theory and computational complexity. A boolean function $f$ on boolean variables $x_1, \ldots, x_n$ is a positive linear threshold function if there exist nonnegative integer coefficients $a_1, \ldots, a_n$ and a nonnegative integer threshold $\theta$ such that

$$f(x_1, \ldots, x_n) = 1 \quad \text{if and only if} \quad a_1 x_1 + \cdots + a_n x_n \geq \theta.$$  

Such a collection $a_1, \ldots, a_n, \theta$ is said to represent $f$. One of the most important positive linear threshold functions is the majority function $\text{MAJ}_n$ which takes value 1 if and only if

$$x_1 + \cdots + x_n \geq \frac{n}{2}.$$  

The weight complexity $[2]$ of a positive linear threshold function $f$ is the smallest value of $a_1 + \cdots + a_n$ across all possible representations of $f$. It is easy to verify that the weight complexity of $\text{MAJ}_n$ is $n$. Håstad [1] has shown that the weight complexity of a positive linear threshold function on $n$ variables can be as large as $2^{O(n \log n)}$.

A monotone boolean formula is a boolean formula which uses only the binary connectives $\land, \lor$; the size of such a formula is the total number of occurrences of variables in the formula. A well-studied problem in formula complexity is that of finding the minimum size of a monotone boolean formula which computes the majority function. The strongest upper bound currently known is due to Valiant [3] who has given a probabilistic construction of a monotone boolean formula of size $O(n^{5.3})$ which computes $\text{MAJ}_n$.

In this note we extend Valiant’s results to obtain an upper bound on the monotone formula complexity of an arbitrary positive linear threshold function. Let $f$ be a positive linear threshold function on $n$ boolean variables which has weight complexity $q(n)$. It is easy to see that Valiant’s upper bound immediately implies the existence of a monotone formula of size $O(q(n)^{5.3})$ which
computes $f$. This is achieved by “reduplicating variables,” i.e. expressing the function $f$ as a majority function over $2q(n)$ variables $y_1, \ldots, y_{2q(n)}$, where $a_i$ of the variables are set to $x_i$ for $i = 1, \ldots, n$, $q(n) - \theta$ of the variables are set to 1, and the remaining $\theta$ variables are set to 0.

We show that a modification of Valiant’s construction yields a monotone formula of size $O(q(n)^{3.3}n^2)$ which computes $f$. If $q(n)$ is much larger than $n$ then this is substantially smaller than the $O(q(n)^{5.3})$ size bound which is obtained by the “black-box” method of reduplicating variables. Furthermore, we also show that our modification of Valiant’s construction yields a small monotone formula which approximates $f$ well in the sense that it agrees with $f$ on all points whose margin with respect to $f$ is large.

Throughout this note $\alpha$ denotes the constant $(3 - \sqrt{5})/2 \approx 0.38$. All logarithms are to the base two. For a vector $a \in \mathbb{R}^n$ we write $\|a\|_1$ to denote the $L_1$ norm $|a_1| + \cdots + |a_n|$ and $\|a\|_2$ to denote the $L_2$ norm $\sqrt{a_1^2 + \cdots + a_n^2}$.

## 2 Valiant’s construction

In this section we outline the construction from [3] in a way which will later be conducive to presenting our results.

Let $A_0$ denote a probability distribution over the set $\{x_1, \ldots, x_n, 0, 1\}$ of depth-0 monotone formulae. We recursively define a sequence $A_1, A_2, \ldots$ of probability distributions over monotone formulae in the following way:

- To choose a formula $F$ according to distribution $A_i$ for $i > 0$ : choose formulae $G^1, G^2, G^3, G^4$ independently from distribution $A_{i-1}$ and let $F$ be the formula $(G^1 \lor G^2) \land (G^3 \lor G^4)$.

Thus a formula $F$ chosen from distribution $A_i$ is a complete binary tree of depth $2i$ with $2^{2i}$ leaves where nodes at depth 0, 2, \ldots are $\land$-connectives, nodes at depth 1, 3, \ldots are $\lor$-connectives, and the
leaves of $F$ are independently selected according to the distribution $A_0$.

Given a sequence $A_0, A_1, A_2, \ldots$ of probability distributions over monotone formulae as described above, we define a corresponding sequence of real-valued functions $g_i : \{0, 1\}^n \rightarrow [0, 1]$ by letting $g_i(x) = \Pr_{F \in A_i}[F(x) = 1]$. The recursive definition of the distributions $A_i$ implies that

$$g_i(x) = g_{i-1}(x)^4 - 4g_{i-1}(x)^3 + 4g_{i-1}(x)^2.$$ 

Now fix $x^0, x^1 \in \{0, 1\}^n$. The following theorems are proved implicitly in Valiant’s paper [3]:

**Theorem 1** If the points $x^0, x^1$ and the distribution $A_0$ are such that $g_0(x^0) \leq \alpha - c/q(n)$ and $g_0(x^1) \geq \alpha + c/q(n)$ for some constant $c > 0$, then we have $g_t(x^0) \leq 2^{-5}$ and $g_t(x^1) \geq 1 - 2^{-5}$ for $t = 1.65 \log q(n) + \Theta(1)$.

**Theorem 2** If $g_t(x^0) \leq 2^{-5}$ and $g_t(x^1) \geq 1 - 2^{-5}$, then we have $g_{t+\log n}(x^0) < 2^{-n-1}$ and $g_{t+\log n}(x^1) > 1 - 2^{-n-1}$.

In [3] Valiant uses the distribution $A_0$ which is defined as follows:

- $A_0$ puts weight $2\alpha/(n-1)$ on each of the $n$ variables $x_1, \ldots, x_n$ and puts the remaining $1 - 2\alpha/(n-1)$ weight on the constant 0.

It is straightforward to verify that for even $n$ this distribution $A_0$ is such that

- any $x \in \{0, 1\}^n$ with $x_1 + \cdots + x_n < n/2$ has $g_0(x) \leq \alpha - \alpha/(n-1)$,

- any $x \in \{0, 1\}^n$ with $x_1 + \cdots + x_n \geq n/2$ has $g_0(x) \geq \alpha + \alpha/(n-1)$.

Theorems 1 and 2 now imply that for $k = 2.65 \log n + \Theta(1)$ we have that

- any $x \in \{0, 1\}^n$ with $x_1 + \cdots + x_n < n/2$ has $g_k(x) \leq 2^{-n-1}$,

- any $x \in \{0, 1\}^n$ with $x_1 + \cdots + x_n \geq n/2$ has $g_k(x) \geq 1 - 2^{-n-1}$.

By the union bound there must exist some $F \in A_k$ which computes $\text{MAJ}_n$ correctly on all $2^n$ points in $\{0, 1\}^n$. This formula $F$ is of size $O(n^{5.3})$.
3 Monotone formulae for positive linear threshold functions

In this section we modify Valiant’s construction to prove the following theorem:

**Theorem 3** Let \( f \) be a positive linear threshold function over \( x_1, \ldots, x_n \) of weight complexity \( q(n) \).

There is a monotone boolean formula of size \( O(q(n)^{3.3}n^2) \) which computes \( f \).

**Proof:** The formula is constructed by selecting an appropriate initial distribution \( A_0 \) in the framework described above. Let \( a_1, \ldots, a_n, \theta \) be a representation of \( f \) which has \( a_1 + \cdots + a_n = q(n) \).

Without loss of generality we can assume \( \theta \leq q(n) \) since otherwise the function \( f \) is identically 0.

Consider the distribution \( A'_0 \) over \( \{x_1, \ldots, x_n, 0, 1\} \) which is defined in the following way:

- \( A'_0 \) puts weight \( \alpha a_j / q(n) \) on variable \( x_j \) for \( j = 1, \ldots, n \), puts weight \( \alpha (q(n) - \theta + 1/2) / q(n) \) on the constant 1, and puts the remaining \( 1 - \alpha (2q(n) - \theta + 1/2) / q(n) \) weight on the constant 0.

Since \( a_1, \ldots, a_n, \theta \) are all integers, for \( x \in \{0,1\}^n \) the quantity \( a_1x_1 + \cdots + a_nx_n \) must be integer-valued and hence cannot lie in the open interval \( (\theta - 1, \theta) \). Consequently we have \( f(x) = 0 \) iff \( a_1x_1 + \cdots + a_nx_n \leq \theta - 1 \) and \( f(x) = 1 \) iff \( a_1x_1 + \cdots + a_nx_n \geq \theta \). By the definition of \( A'_0 \) we have that

\[
g_0(x) = \frac{\alpha}{q(n)} \left( q(n) + a_1x_1 + \cdots + a_nx_n - \theta + \frac{1}{2} \right),
\]

which implies that

- any \( x \in \{0,1\}^n \) with \( f(x) = 0 \) has \( g_0(x) \leq \alpha - \alpha/2q(n) \),

- any \( x \in \{0,1\}^n \) with \( f(x) = 1 \) has \( g_0(x) \geq \alpha + \alpha/2q(n) \).

As in Section 2, Theorems 1 and 2 now imply that there must exist some monotone formula \( F \in A_{1.65\log q(n) + \log n + \Theta(1)} \) which computes the function \( f \) correctly on all \( 2^n \) boolean inputs. This formula \( F \) is of size \( O(q(n)^{3.3}n^2) \).
The construction described in Theorem 3 gives the best known upper bound on the monotone formula complexity of positive linear threshold functions for a wide range of weight complexities. However, if the weight complexity $q(n)$ is sufficiently large, then our upper bound of $O(q(n)^{3.3}n^2)$ does not improve on the trivial upper bound obtained by expressing $f$ as a monotone DNF formula. Although we do not know how to show that there must always exist a small monotone function which exactly computes the function $f$, in the next section we show that independent of the weight complexity of $f$ there must exist a small monotone formula which is an accurate approximator for $f$.

4 Monotone formulae which approximate positive linear threshold functions

Following Vapnik [4], we say that the margin of a point $y \in \{0, 1\}^n$ with respect to a hyperplane $a_1x_1 + \cdots + a_nx_n = \theta$ is the minimum Euclidean distance between $y$ and the hyperplane. Let $f$ be a positive linear threshold function. We say that a boolean function $F : \{0, 1\}^n \rightarrow \{0, 1\}$ approximates $f$ with margin $\epsilon$ if there is a sequence of nonnegative integers $a_1, \ldots, a_n, \theta$ such that

- $a_1, \ldots, a_n, \theta$ represents $f$,

- $F(y) = f(y)$ for every $y \in \{0, 1\}^n$ which has margin at least $\epsilon$ with respect to the hyperplane $a_1x_1 + \cdots + a_nx_n = \theta$.

Such a function $F$ thus agrees with $f$ on all points which lie outside the "fat hyperplane" of width $2\epsilon$ which is centered around the hyperplane $a_1x_1 + \cdots + a_nx_n = \theta$.

The following theorem shows that independent of the weight complexity of $f$, there must exist a monotone formula of size $\text{poly}(n, 1/\epsilon)$ which approximates $f$ with margin $\epsilon$. 

5
Theorem 4 Let $f$ be a positive linear threshold function over $x_1, \ldots, x_n$. For any $\epsilon > 0$ there is a monotone formula of size $O(n^{3.65}/\epsilon^{3.3})$ which approximates $f$ with margin $\epsilon$.

Proof: Let $a_1, \ldots, a_n, \theta$ be a sequence of nonnegative integers which represents $f$. We define a slightly different distribution $A'_0$ from the distribution which was used in the proof of Theorem 3:

- $A'_0$ puts weight $\alpha a_j/\|a\|_1$ on variable $x_j$ for $j = 1, \ldots, n$, puts weight $\alpha(\|a\|_1 - \theta)/\|a\|_1$ on the constant 1, and puts the remaining $1 - \alpha(2\|a\|_1 - \theta)/\|a\|_1$ weight on the constant 0.

Let $y \in \{0, 1\}^n$ be a point whose margin with respect to the hyperplane $a_1 x_1 + \cdots + a_n x_n = \theta$ is at least $\epsilon$. Since the distance from $y$ to the hyperplane $a_1 x_1 + \cdots + a_n x_n = \theta$ is equal to $\|a\|_2^{-1} |a_1 y_1 + \cdots + a_n y_n - \theta|$, we have that $|a_1 y_1 + \cdots + a_n y_n - \theta| \geq \epsilon \|a\|_2$.

Suppose first that $a_1 y_1 + \cdots + a_n y_n \geq \theta + \epsilon \|a\|_2$. As in the proof of Theorem 3, the definition of $A'_0$ implies that $g_0(y) \geq \alpha + \alpha \epsilon \|a\|_2/\|a\|_1$. Since $\|a\|_2 \geq \|a\|_1/\sqrt{n}$ for any vector $a \in \mathbb{R}^n$, we have that $g_0(y) \geq \alpha + \alpha \epsilon / \sqrt{n}$. Applying Theorems 1 and 2 we find that $g_t(y) \geq 1 - 2^{-n-1}$ for $t = 1.65 \log(\sqrt{n}/\epsilon) + \log n + \Theta(1)$.

An entirely similar argument shows that if $a_1 y_1 + \cdots + a_n y_n \leq \theta - \epsilon \|a\|_2$, then $g_t(y) \leq 2^{-n-1}$ for the same value of $t$.

Since there can be at most $2^n$ points $y \in \{0, 1\}^n$ whose margin is at least $\epsilon$, there must exist some function $F \in A_t$ which agrees with $f$ on every point whose margin is at least $\epsilon$. This function $F$ is of size $O(n^{3.65}/\epsilon^{3.3})$.

\[ \text{\hfill \blacksquare} \]

5 Acknowledgements

We thank Les Valiant for helpful discussions and suggestions.
References


