Monotone Boolean Formulas can Approximate Monotone Linear Threshold Functions

Rocco A. Servedio*
Department of Computer Science
Columbia University
New York, NY 10027
rocco@cs.columbia.edu

July 23, 2003

Abstract

We show that any monotone linear threshold function on n Boolean variables can be approximated to within any constant accuracy by a monotone Boolean formula of poly(n) size.

1 Introduction

Over the past two decades, researchers in computational complexity have studied monotone computation models in a variety of contexts. While many notable results have been achieved, some seemingly basic questions about low-level monotone complexity remain unanswered. In this paper we examine the relative power of two simple models of monotone computation for Boolean functions: *monotone linear threshold functions,* which compute a weighted sum $\sum_{i=1}^{n} w_i x_i$ of inputs and compare it with a threshold θ, and *monotone Boolean formulas* over the basis {AND,OR}.

The question which motivates our study is the following: does every monotone linear threshold function on n Boolean variables have a monotone Boolean formula of size poly(n)? This is an interesting and natural question for several reasons:

- In a celebrated result Ajtai *et al.* [1] gave a polynomial size monotone formula which computes the majority function (their construction also gives a monotone circuit of size $O(n \log n)$ and depth $O(\log n)$). Subsequently Valiant [10] gave an elegant probabilistic construction of monotone formulas of size $O(n^{5.3})$ for the majority function on n bits. Since majority is simply a monotone linear threshold function in which each weight w_i

*Supported in part by an NSF Mathematical Sciences Postdoctoral Fellowship, by NSF grant CCR-95-04436 and by ONR grant N00014-96-1-0550. Work was performed while the author was at Harvard University.

is 1, it is natural to now ask whether all monotone linear threshold functions (regardless of the size of the weights) have polynomial size monotone formulas.

- This question is equivalent to the question of whether monotone TC^0 (the class of functions computed by monotone threshold circuits of constant depth and polynomial size) is contained in monotone NC^1 (which can be defined as the class of functions computed by monotone formulas of polynomial size; see Section 2.4 of [5]). A positive answer would give an interesting contrast to results of Yao [11] who exhibits polynomial size monotone formulas which cannot be computed by constant depth polynomial size monotone threshold circuits.

- Goldmann and Karpinski [3] have posed the following question: does every monotone linear threshold function have a monotone constant depth polynomial size circuit of majority gates? (While several simulations of an arbitrary linear threshold gate by constant-depth polynomial-size circuits of majority gates are known [4, 3], these simulations do not preserve monotonicity.) By the results of Ajtai *et al.* [1] and Valiant [10], a negative answer to our question would imply a negative answer to Goldmann and Karpinski's question.

In this paper we prove that any monotone linear threshold function can be approximated to within any constant accuracy by a monotone Boolean formula of polynomial size. Our proof uses the existence of polynomial size monotone formulas for the majority function [1, 10] together with inequalities for sums of independent random variables (which assert that such sums are unlikely to have very small deviations from their expected values) and a recursive decomposition technique.

2 Preliminaries

We write log for $\log_2$ and ln for $\log_e$. For a vector $v \in \mathbb{R}^n$ we write $\|v\|_2$ to denote the 2-norm $\sqrt{\sum_{i=1}^n v_i^2}$. The function $\mathrm{sgn}(z)$ takes value 1 if $z \geqslant 0$ and -1 if $z < 0$.

A Boolean function $f : \{-1,1\}^n \to \{-1,1\}$ is a *linear threshold function* (henceforth simply a threshold function) if there exist coefficients $w = (w_1, \ldots, w_n) \in \mathbb{R}^n$ and a threshold $\theta \in \mathbb{R}$ such that $f(x) = \mathrm{sgn}(w \cdot x - \theta)$. Such a pair w, θ is said to represent f. A *monotone threshold function* is a threshold function which computes a monotone Boolean function; equivalently, a monotone threshold function is one which has some representation in which each $w_i \geqslant 0$.

Since we are only concerned with the discrete Boolean cube, for any threshold function f there are infinitely many different representations of f. The *weight* of a threshold function f is the smallest value of $\sum_{i=1}^n |w_i|$ across all representations of f such that w_i, θ are all integers. Note that the weight of any threshold function is well defined since every threshold function f has a representation with integer coefficients and threshold. It has long been known [7] that every threshold function on n variables has weight at most $2^{O(n \log n)}$, and Håstad [9] has exhibited a threshold function on n variables which has weight $2^{\Theta(n \log n)}$.

The *majority* function $MAJ_n : \{-1,1\}^n \to \{-1,1\}$ is the monotone threshold function

$$MAJ_n(x_1,\ldots,x_n) = \mathrm{sgn}(\sum_{i=1}^n x_i).$$

A *monotone formula* on $x_1,\ldots,x_n$ is a Boolean formula F which uses only the binary connectives $\wedge$ (AND) and $\vee$ (OR). Equivalently, F is is a rooted binary tree in which each internal node has degree exactly two and is labelled with either $\wedge$ or $\vee$, and each leaf is labelled with one of the variables $x_1,\ldots,x_n$. We view -1 as representing FALSE and 1 as representing TRUE. The *size* of a formula is the total number of occurences of variables in the formula, i.e. the number of leaves in the binary tree. Note that this is exactly one more than the number of gates (internal nodes) of the binary tree.

If f is a Boolean function on inputs $x_1,\ldots,x_n$ and b is a bit we write $f|_{x_1\leftarrow b}$ to denote the function $f(b,x_2,\ldots,x_n)$ on inputs $x_2,\ldots,x_n$. If f,g are Boolean functions we say that g is an *ϵ-approximator* for f if $\Pr[f(x) \neq g(x)] \leqslant \epsilon$ where the probability is uniform over $x \in \{-1,1\}^n$.

We will use the following result due to Valiant:

Theorem 1 (Valiant) *There exist monotone formulas for MAJ_n of size $O(n^{5.3})$.*

By reduplicating inputs to MAJ_n and inserting some number of constant inputs 0 or 1, we obtain the following corollary:

Corollary 2 *Let f be a monotone linear threshold function of weight W. Then f has a monotone formula of size $O(W^{5.3})$.*

3 Small monotone formulas can approximate monotone threshold functions

Our main result is the following:

Theorem 3 *Let f be any monotone threshold function on n variables and let $0 < \epsilon < \frac{1}{2}$. There is a poly$(n, 2^{(\log 1/\epsilon)^2 \epsilon^{-4}})$ size monotone formula F which is an ϵ-approximator for f.*

This theorem implies that for any constant ϵ (in fact for any $\epsilon = \Omega(\frac{(\log\log n)^{1/2}}{(\log n)^{1/4}})$), there is an ϵ-approximating monotone formula of size poly(n) for any monotone threshold function.

Proof of Theorem 3: Let $f(x) = \mathrm{sgn}(\sum_{i=1}^n w_i x_i - \theta)$. Without loss of generality we may suppose that $1 = w_1 \geqslant w_2 \geqslant \cdots \geqslant w_n > 0$. The proof has several cases.

Case I: $\epsilon \geqslant \frac{6}{\|w\|_2}$, i.e. $\epsilon = \frac{6\lambda}{\|w\|_2}$ for some $\lambda \geqslant 1$. For $i = 1,\ldots,n$ let w'_i be obtained by rounding w_i to the nearest integer multiple of $\frac{1}{n}$, and let $f'(x) = \mathrm{sgn}(\sum_{i=1}^n w'_i x_i - \theta)$. Clearly f' has weight at most $O(n^2)$, and hence by Corollary 2 there is a monotone formula F for f' of size $O(n^{10.6})$. Since $|w'_i - w_i| \leqslant \frac{1}{2n}$ for all i, we have $|\sum_{i=1}^n w'_i x_i - \sum_{i=1}^n w_i x_i| \leqslant 1/2$ and thus

$$\Pr[f(x) \neq f'(x)] \leqslant \Pr[|\sum_{i=1}^n w_i x_i - \theta| \leqslant \frac{1}{2}] \leqslant \Pr[|\sum_{i=1}^n w_i x_i - \theta| \leqslant \lambda]. \quad (1)$$

We now use the following bound which asserts that the distribution of a sum of independent random variables cannot be too tightly clustered around its expected value:

Theorem 4 *Let $0 < w_1, \ldots, w_n \leqslant 1$, and let $X_1, \ldots, X_n$ be independent random variables such that X_k is w_k with probability $\frac{1}{2}$ and $-w_k$ with probability $\frac{1}{2}$. Let $x = \sum_{k=1}^{n} X_k$. Then for every $\lambda \geqslant 1$ and every $\theta \in \mathbb{R}$, we have $\Pr[|x - \theta| \leqslant \lambda] < \frac{6\lambda}{\|w\|_2}$.*

Theorem 4 can be derived from Theorem 2.14 in [8]. Since the proof in [8] is long and complicated we give a self-contained proof of Theorem 4 in Appendix A.

By Theorem 4 and inequality (1) we have that $\Pr[f(x) \neq f'(x)] \leqslant \frac{6\lambda}{\|w\|_2} = \epsilon$. Hence the monotone formula F of size $O(n^{10.6})$ for f' is the desired ϵ-approximator for f.

Case II: $\epsilon < \frac{6}{\|w\|_2}$. We define a sequence $T_0, T_1, \ldots, T_n$ of augmented monotone formulas as follows: (The formulas are augmented in that the leaves may contain either variables as usual or monotone threshold functions.)

1. T_0 is a single leaf which contains the monotone threshold function f on variables $x_1, \ldots, x_n$.

2. For $i \geqslant 0$, T_{i+1} is obtained from T_i as follows: for each leaf of T_i which contains a monotone threshold function f' on variables $x_{i+1}, \ldots, x_n$, replace the leaf by the augmented monotone formula $f'|_{x_{i+1} \leftarrow -1} \vee \left(x_{i+1} \wedge f'|_{x_{i+1} \leftarrow 1}\right)$.

An easy induction using the above definition (2) of T_{i+1} and the base case (1) for T_0 shows that for each leaf of T_i which contains a monotone threshold function f', the inputs to f' are indeed $x_{i+1}, \ldots, x_n$ as required by (2). Now observe that for any monotone threshold function f' on variables $x_{i+1}, \ldots, x_n$ the augmented monotone formula $f'|_{x_{i+1} \leftarrow -1} \vee \left(x_{i+1} \wedge f'|_{x_{i+1} \leftarrow 1}\right)$ is logically equivalent to f'. (If $x_{i+1} = -1$ then the augmented monotone formula reduces to $f'|_{x_{i+1} \leftarrow -1}$ which equals f' on any input such that $x_{i+1} = -1$. If $x_{i+1} = 1$ then the augmented monotone formula reduces to $f'|_{x_{i+1} \leftarrow -1} \vee f'|_{x_{i+1} \leftarrow 1}$ which equals $f'|_{x_{i+1} \leftarrow 1}$ since f' is a monotone function.) Thus replacing f' by the augmented monotone formula $f'|_{x_{i+1} \leftarrow -1} \vee \left(x_{i+1} \wedge f'|_{x_{i+1} \leftarrow 1}\right)$ is similar to building a decision tree by splitting on x_{i+1}: if x_{i+1} equals 1 then the value of the whole formula is given by the right-hand term $f'|_{x_{i+1} \leftarrow 1}$, and if x_{i+1} equals -1 then the value of the whole formula is given by the left-hand term $f'|_{x_{i+1} \leftarrow -1}$.

Using these observations, the following facts are also easily verified by induction:

1. Each augmented formula T_i computes the function f.

2. Each augmented formula T_i has 2^i leaves which are monotone threshold functions over inputs $x_{i+1}, \ldots, x_n$ and $2^i - 1$ leaves which are variables.

3. Each monotone threshold function leaf in T_i corresponds to a unique i-bit string $b_1 \ldots b_i \in \{-1, 1\}^i$, and the threshold function at that leaf is

$$f(b_1, \ldots, b_i, x_{i+1}, \ldots, x_i) = \operatorname{sgn}(w_{i+1}x_{i+1} + \cdots + w_n x_n - (\theta - w_1 b_1 - \cdots - w_i b_i)).$$

   Moreover, for any input in which the first i bits are $b_1, \ldots, b_i$, the value of the entire formula T_i is given by the value computed at this leaf.

Let $\ell = \frac{5184}{\epsilon^4}\left(\ln \frac{4}{\epsilon}\right)^2$. We henceforth suppose that ϵ is such that $\ell < n$ since otherwise the bound of Theorem 3 is trivially true. We consider two subcases:

Case IIa: $\frac{w_i^2}{\sum_{j=i}^n w_j^2} < \frac{\epsilon^2}{36}$ for some $1 \leqslant i \leqslant \ell$. In this case we rescale the coefficients $w_i, \ldots, w_n$, i.e. we define $w'_j = \frac{w_j}{w_i}$ for $j = i$ to n; note that $1 = w'_i \geqslant w'_{i+1} \geqslant \cdots \geqslant w'_n$. Let w' denote the $(n-i+1)$-dimensional vector $(w'_i, \ldots, w'_n)$. We have

$$\|w'\|_2^2 = \sum_{j=i}^{n} \left(\frac{w_j}{w_i}\right)^2 > \frac{36}{\epsilon^2}$$

and hence $\epsilon > \frac{6}{\|w'\|_2}$. Consequently for each leaf in T_{i-1} which contains a monotone threshold function f', as in Case I there is some monotone formula F' of size $O(n^{10.6})$ which is an ϵ-approximator for f'. (Note that we are using the fact that all 2^{i-1} monotone threshold functions in the leaves of T_{i-1} have the same coefficients $w'_i, \ldots, w'_n$.) By fact (3) above, replacing each leaf f' in T_{i-1} with the appropriate ϵ-approximating monotone formula F' gives a monotone formula which ϵ-approximates f. This formula has size $O(2^i n^{10.6}) = O(2^\ell n^{10.6})$.

Case IIb: $\frac{w_i^2}{\sum_{j=i}^n w_j^2} \geqslant \frac{\epsilon^2}{36}$ for all $1 \leqslant i \leqslant \ell$. We thus have $w_i^2 \geqslant \frac{\epsilon^2}{36}\sum_{j=i}^n w_j^2$ and hence $\sum_{j=i+1}^n w_j^2 \leqslant (1 - \frac{\epsilon^2}{36})\sum_{j=i}^n w_j^2$ for $i = 1, \ldots, \ell$. Hence

$$\sum_{j=\ell+1}^{n} w_j^2 \leqslant \left(1 - \frac{\epsilon^2}{36}\right)^{\ell} \sum_{j=1}^{n} w_j^2 \leqslant \left(1 - \frac{\epsilon^2}{36}\right)^{\ell} \cdot \frac{36}{\epsilon^2} \tag{2}$$

where the last inequality is because we are in Case II.

Let W denote $\sum_{j=\ell+1}^n w_j^2$. Since $w_\ell^2 \geqslant \frac{\epsilon^2}{36}\sum_{j=\ell}^n w_j^2$, we have $w_\ell^2 > \frac{\epsilon^2}{36}\sum_{j=\ell+1}^n w_j^2 = \frac{\epsilon^2}{36} \cdot W$ and hence $w_1 \geqslant w_2 \geqslant \cdots \geqslant w_\ell > \frac{\epsilon}{6}\sqrt{W}$. Note also that by the definition of W we have $\sqrt{W} \geqslant w_{\ell+1}, \ldots, w_n$.

Consider the function $g : \{+1,-1\}^n \to \{+1,-1\}$ defined by $g(x_1, \ldots, x_n) = \mathrm{sgn}(\sum_{i=1}^\ell w_i x_i - (\theta - \eta))$ where $\eta > 0$ will be defined later. Clearly g is a monotone function, and since g depends on only ℓ variables there is a monotone formula for g of size 2^ℓ. Now note that for $x \in \{+1,-1\}^n$ we have $g(x) \neq f(x)$ only if $\left|\sum_{i=1}^\ell w_i x_i - \theta\right| \leqslant \eta$ or $\left|\sum_{i=\ell+1}^n w_i x_i\right| > \eta$. (Suppose that both $\left|\sum_{i=1}^\ell w_i x_i - \theta\right| > \eta$ and $\left|\sum_{i=\ell+1}^n w_i x_i\right| \leqslant \eta$. If $\sum_{i=1}^\ell w_i x_i - \theta > \eta$, then clearly $g(x) = 1$, and $f(x)$ must also be 1 since $\sum_{i=1}^n w_i x_i$ is at most η less than $\sum_{i=1}^\ell w_i x_i$. If $\sum_{i=1}^\ell w_i x_i - \theta < -\eta$, then we have $g(x) = -1$ and $f(x)$ must also be -1 since $\sum_{i=1}^n w_i x_i$ is at most η more than $\sum_{i=1}^\ell w_i x_i$.) We will bound each of $\Pr[|\sum_{i=1}^\ell w_i x_i - \theta| \leqslant \eta]$ and $\Pr[|\sum_{i=\ell+1}^n w_i x_i| > \eta]$ by $\frac{\epsilon}{2}$ and thus establish $\Pr[g(x) \neq f(x)] \leqslant \epsilon$.

To bound $\Pr[|\sum_{i=\ell+1}^n w_i x_i| > \eta]$ we use Bernstein's inequality (see e.g. Section 7 of [6] or Theorem 2.8 of [8]):

Bernstein's Inequality: Let $V_1, \ldots, V_r$ be independent random variables with zero means and bounded ranges $|V_i| \leqslant M$. Let $V = \sum_{i=1}^r V_i$. Then for every $\eta > 0$ we have

$$\Pr[|V| > \eta] \leqslant 2\exp\left[-\eta^2/(2(\mathrm{Var}[V] + M\eta))\right]. \tag{3}$$

We will apply (3) to the random variables $w_{\ell+1}x_{\ell+1}, \ldots, w_n x_n$. As noted above we have $|w_j| \leqslant \sqrt{W}$ for all $j = \ell+1, \ldots, n$ and moreover $\mathrm{Var}[V] = \sum_{j=\ell+1}^n w_j^2 = W$. Hence we obtain from Bernstein's inequality

$$\Pr[|\sum_{j=\ell+1}^{n} w_j x_j| > \eta] \;\leqslant\; 2\exp[-\eta^2/(2W + 2\sqrt{W}\eta)]. \tag{4}$$

Let $\eta = 3\sqrt{W}\ln\frac{4}{\epsilon}$. Since $\epsilon < \frac{1}{2}$ we have $\ln\frac{4}{\epsilon} > 1$ and hence

$$\eta^2 = 9W\left(\ln\frac{4}{\epsilon}\right)^2 > 2W\ln\frac{4}{\epsilon} + 6W\left(\ln\frac{4}{\epsilon}\right)^2 = \left(2W + 2\sqrt{W}\eta\right)\ln\frac{4}{\epsilon}$$

so consequently the right side of (4) is less than $\frac{\epsilon}{2}$.

It remains to show that $\Pr[|\sum_{i=1}^{\ell} w_i x_i - \theta| \leqslant \eta] \leqslant \frac{\epsilon}{2}$. To establish this we use a bound which is somewhat different from Theorem 4:[1]

Theorem 5 *Let $0 < b < w_1, \ldots, w_n$ and let $X_1, \ldots, X_n$ be independent random variables such that X_k is w_k with probability $\frac{1}{2}$ and $-w_k$ with probability $\frac{1}{2}$. Let $x = \sum_{k=1}^n X_k$. Then for every $\lambda \geqslant 1$ and every $\theta \in \mathbb{R}$, we have* $\Pr[|x - \theta| \leqslant \lambda b] \leqslant \frac{2\lambda}{\sqrt{n}}$.

(This theorem can also be derived from results in [8]. We give a simple self-contained proof due to Benjamini *et al.* [2] in Appendix B.) We take $b = \frac{\epsilon}{6}\sqrt{W}$ and $\lambda = \frac{18\ln\frac{4}{\epsilon}}{\epsilon}$ so $\lambda b = 3\sqrt{W}\ln\frac{4}{\epsilon} = \eta$. As mentioned earlier we have that $w_1 \geqslant \cdots \geqslant w_\ell > b$, so Theorem 5 gives

$$\Pr[|\sum_{i=1}^{\ell} w_i x_i - \theta| \leqslant \eta] \leqslant \frac{36\ln\frac{4}{\epsilon}}{\epsilon\sqrt{\ell}}.$$

Our choice of ℓ implies that this bound is at most $\frac{\epsilon}{2}$.

Thus we see that the worst case size bound for an ϵ-approximating monotone formula for f comes from Case IIa in our analysis, which gives a bound of $O(2^\ell n^{10.6}) = n^{10.6} \cdot \mathrm{poly}(2^{(\ln\frac{1}{\epsilon})^2\epsilon^{-4}})$, and Theorem 3 is proved. ■

4 Conclusion

In Case I of our proof, we saw that for certain monotone threshold functions we can round the weights and obtain a low-weight monotone threshold function which is an ϵ-approximator. A natural question is whether this technique works in general, i.e. can every monotone threshold function be approximated by a low-weight monotone threshold function? If one could show, for example, that any monotone threshold function can be ϵ-approximated by a monotone threshold function of weight polynomial in n and exponential in $\frac{1}{\epsilon}$, then this would give an alternate proof of Theorem 3. While we have not been able to prove such a result, we can show that in general there does not exist an ϵ-approximating monotone threshold function of weight $\mathrm{poly}(n, \frac{1}{\epsilon})$:

[1] We cannot apply Theorem 4 here because $\eta < 1$; recall that from the definition of ℓ and equation (2) we have $W = 2^{-\Omega(\epsilon^{-2})}$.

Claim 6 *There is no polynomial $p(\cdot,\cdot)$ such that every monotone threshold function on n variables can be ϵ-approximated by a monotone threshold function of weight $p(n, \frac{1}{\epsilon})$.*

Proof: As mentioned in Section 2, Håstad [9] has given a threshold function h on n variables which requires weight $2^{\Omega(n \log n)}$. Since every threshold function is unate (i.e. can be made monotone by flipping some coordinate axes) we may take h to be monotone; note that this does not change its weight. If there were a polynomial $p(n, \frac{1}{\epsilon})$ as described above then by taking $\epsilon = \frac{1}{2^n+1}$ we would obtain a monotone threshold function of weight $p(n, 2^n+1) = 2^{O(n)}$ which $\frac{1}{2^n+1}$-approximates h. Since $|\{-1,1\}^n| = 2^n$ this approximator must in fact compute h exactly, contradicting Håstad's lower bound on the weight of h. ■

5 Acknowledgements

Theorem 4 was derived with Ryan O'Donnell; we thank him. We thank Les Valiant for helpful discussions and suggestions.

References

[1] M. Ajtai, J. Komlos, and E. Szemeredi. An $O(n \log n)$ sorting network. *Combinatorica*, 3(1):1–19, 1983.

[2] I. Benjamini, G. Kalai, and O. Schramm. Noise sensitivity of boolean functions and applications to percolation. *Inst. Hautes Études Sci. Publ. Math.*, 90:5–43, 1999.

[3] M. Goldmann and M. Karpinski. Simulating threshold circuits by majority circuits. *SIAM Journal on Computing*, 27(1):230–246, 1998.

[4] M. Goldmann, J. Håstad, and A. Razborov. Majority gates vs. general weighted threshold gates. *Computational Complexity*, 2:277–300, 1992.

[5] M. Grigni and M. Sipser. *Monotone complexity*, pages 57–75. London Mathematical Society lecture notes, 1992.

[6] W. Haerdle, H. Liang, and J. Gao. Partially linear models. e-book available at `http://www.quantlet.de/scripts/plm/`, 1999.

[7] S. Muroga. *Threshold logic and its applications.* Wiley-Interscience, New York, 1971.

[8] V. V. Petrov. *Limit theorems of probability theory.* Oxford Science Publications, Oxford, England, 1995.

[9] J. Håstad. On the size of weights for threshold gates. *SIAM Journal on Discrete Mathematics*, 7(3):484–492, 1994.

[10] L. Valiant. Short monotone formulae for the majority function. *Journal of Algorithms*, 5:363–366, 1984.

[11] A. Yao. Circuits and local computation. In *Proceedings of the Twenty-First Symposium on Theory of Computing*, pages 186–196, 1989.

A Proof of Theorems 4 and 5

Proof of Theorem 4: We first handle the case $\lambda = 1$. Define:

$$p(x) = \frac{2(1-\cos x)}{x^2} \geqslant 0 \qquad \text{and} \qquad h(t) = \begin{cases} 1-|t|, & |t| \leqslant 1 \\ 0, & \text{else} \end{cases}.$$

Elementary integration by parts shows that $p(x)$ is the inverse Fourier transform of $h(t)$; i.e.,

$$p(x) = \int_{-\infty}^{\infty} e^{-itx} h(t)\, dt.$$

By considering the first two terms of the Taylor series for $\cos x$, we see that $p(x) \geqslant \frac{11}{12}$ on $[-1, 1]$. Hence:

$$\begin{aligned}
\Pr[|x-\theta| \leqslant 1] &= \mathop{\mathbf{E}}_{x}[\mathbf{1}_{x\in[\theta-1,\theta+1]}] \\
&\leqslant \tfrac{12}{11}\mathbf{E}[p(x-\theta)] \\
&= \tfrac{12}{11}\mathbf{E}\Big[\int_{-\infty}^{\infty} e^{-it(x-\theta)} h(t)\, dt\Big] \\
&= \frac{12}{11}\int_{-\infty}^{\infty} \mathbf{E}[e^{-itx}e^{it\theta}h(t)]\, dt \\
&= \frac{12}{11}\Big|\int_{-\infty}^{\infty} e^{it\theta}h(t)\mathbf{E}[e^{-itx}]\, dt\Big| \qquad (5)\\
&\leqslant \frac{12}{11}\int_{-1}^{1} \big|\mathbf{E}[e^{-itx}]\big|\, dt, \qquad (6)
\end{aligned}$$

with (5) following because the quantity is already real and nonnegative, and (6) following because $|e^{it\theta}| \leqslant 1$, $h(t) = 0$ outside $[-1, 1]$, and $|h(t)| \leqslant 1$ otherwise. Now observe that

$$\begin{aligned}
\mathop{\mathbf{E}}_{x}[e^{-itx}] &= \mathop{\mathbf{E}}_{x_1\leftarrow X_1,\ldots,x_n\leftarrow X_n}[\exp(-it\sum_{k=1}^{n} x_k)] \\
&= \mathop{\mathbf{E}}_{x_1\leftarrow X_1,\ldots,x_n\leftarrow X_n}\Big[\prod_{k=1}^{n}\exp(-itx_k)\Big] \\
&= \prod_{k=1}^{n}\mathop{\mathbf{E}}_{x_k\leftarrow X_k}[\exp(-itx_k)] \qquad (7)\\
&= \prod_{k=1}^{n}(\tfrac{1}{2}\exp(-itw_k) + \tfrac{1}{2}\exp(itw_k)) \\
&= \prod_{k=1}^{n}\cos(w_k t)
\end{aligned}$$

where equation (7) is by independence. By comparing Taylor expansions, we find that $\cos u \leqslant \exp(-u^2/2)$ on the interval $[-1,1]$. Since $w_k \leqslant 1$ for all i, we may conclude that:

$$\begin{aligned}
\Pr[|x-\theta| \leqslant 1] &\leqslant \frac{12}{11}\int_{-1}^{1}\prod_{k=1}^{n}\exp(-w_k^2t^2/2)\,dt \\
&= \frac{12}{11}\int_{-1}^{1}\exp(-t^2/2(\|w\|_2^{-1})^2)\,dt \\
&\leqslant \frac{12}{11}\int_{-\infty}^{\infty}\exp(-t^2/2(\|w\|_2^{-1})^2)\,dt \\
&= \sqrt{2\pi}(\tfrac{12}{11})\|w\|_2^{-1}
\end{aligned}$$

since $(\sqrt{2\pi}\sigma)^{-1}\exp(-t^2/2\sigma^2)$ is a probability density function for every positive σ. Since we have made no assumptions about θ anywhere in the proof, this establishes that for every $\theta' \in \mathbb{R}$ we have $\Pr[|x-\theta'| \leqslant 1] < \frac{3}{\|w\|_2}$.

For the general case fix any $\lambda \geqslant 1$ and any $\theta \in \mathbb{R}$; we must bound $\Pr[x \in [\theta-\lambda, \theta+\lambda]]$. For any integer $j \geqslant 1$, by taking $\theta' = \theta - \lambda + (2j-1)$ in the case we have already proved, we have that $\Pr[x \in [\theta-\lambda+(2j-2), \theta-\lambda+2j]] < \frac{3}{\|w\|_2}$. Taking a union bound over $j = 1, 2, \ldots, \lceil\lambda\rceil$, we have that $\Pr[x \in \theta-\lambda, \theta-\lambda+2\lceil\lambda\rceil] < \frac{3\lceil\lambda\rceil}{\|w\|_2}$. Since $-\lambda + 2\lceil\lambda\rceil \geqslant \lambda$ and $\lceil\lambda\rceil < 2\lambda$, we have $\Pr[x \in \theta-\lambda, \theta+\lambda] < \frac{6\lambda}{\|w\|_2}$ and the theorem is proved. ∎

Proof of Theorem 5: As in the previous proof, we first prove the case $\lambda = 1$ and then use this case to establish the general theorem.

The distribution of x as described in the theorem is easily seen to be identical to the distribution obtained via the following process:

1. Choose a random permutation π on $\{1, \ldots, n\}$.

2. Choose an integer $\ell \in \{0, 1, \ldots, n\}$ according to the binomial distribution $B(n, \frac{1}{2})$.

3. Set x to $\sum_{i=1}^{\ell} w_{\pi(i)} - \sum_{i=\ell+1}^{n} w_{\pi(i)}$.

Let $W(\pi, \ell)$ denote $\sum_{i=1}^{\ell} w_{\pi(i)} - \sum_{i=\ell+1}^{n} w_{\pi(i)}$. Note that $W(\pi,\ell)$ increases with ℓ for each π.

Fix any $\theta \in \mathbb{R}$. Consider any fixed permutation π. Since each w_i is greater than b, we have $W(\pi, i+1) > W(\pi, i) + 2b$ for all i, and hence there is exactly one value $\ell \in \{0, 1\ldots, n\}$ such that $|W(\pi,\ell) - \theta| \leqslant b$. Hence for any fixed π we have that $\Pr[|W(\pi,\ell)-\theta| \leqslant b]$ is at most $\max_{k=0,1,\ldots,n} \Pr[\ell = k]$. Since ℓ is drawn from $B(n, \frac{1}{2})$ this is at most $\binom{n}{\lceil n/2\rceil}2^{-n} \leqslant \sqrt{\frac{2}{\pi n}} < \frac{4}{5\sqrt{n}}$. Averaging over all choices of π, we find that $\Pr[|x-\theta| \leqslant b] \leqslant \frac{4}{5\sqrt{n}}$.

For the general case of arbitrary $\lambda \geqslant 1$, the same argument and analysis as in the last paragraph of the proof of Theorem 4 using a union bound proves the theorem. ∎