Universität Ulm
Fakultät für Informatik

A Slightly Improved Upper Bound on the Size of Weights Sufficient to Represent Any Linearly Separable Boolean Function

MICHAEL SCHMITT
Universität Ulm

Nr. 92-10

Ulmer Informatik-Berichte

November 1992

A Slightly Improved Upper Bound on the Size of Weights Sufficient to Represent Any Linearly Separable Boolean Function

Michael Schmitt*

Abstract

The maximum absolute value of integral weights sufficient to represent any linearly separable Boolean function is investigated. It is shown that upper bounds exhibited by Muroga (1971) for rational weights satisfying the normalized system of inequalities also hold for integral weights. Therewith, the previous best known upper bound for integers is improved by approximately a factor of 1/2.

1 Introduction

A linearly separable Boolean function $f : \{0,1\}^n \to \{0,1\}$ is represented by a real vector $(w_1, \ldots, w_n, t)$ of weights such that for all $x \in \{0,1\}^n$

$$w_1 x_1 + \cdots + w_n x_n \geq t \qquad \text{iff} \qquad f(x_1 \ldots x_n) = 1. \tag{1}$$

f is also called threshold function and t the threshold. It is a well-known fact that the possibly infinite information contained in the real components of the weight vector can be made finite without restricting the class of representable functions by requiring all weights to be integers (see e.g. [7, 11, 16] for proofs). In the past, there has been considerable interest to bound the maximum absolute integral value sufficient for a weight from above for various reasons. Predominant was the search for a polynomial upper bound on the length of a weight in binary representation leading to $O(n \log n)$ as the best known asymptotic bound up to now [6, 17, 16]. The proof of [6] can also be found in [15] in a more elaborated version concluding with $((n+3)\log(n+1)+1)$. The tightest result has been given by Muroga in [13]. He investigated weight vectors satisfying the so-called normalized system of inequalities

$$\begin{array}{lcll} w_1 x_1 + \cdots + w_n x_n & \geq & t & \text{if} \quad f(x) = 1 \\ w_1 x_1 + \cdots + w_n x_n & \leq & t-1 & \text{if} \quad f(x) = 0 \end{array} \tag{2}$$

obtaining the following results (see [13, Section 9.3.2]):

*Department of Neural Information Processing, University of Ulm, Oberer Eselsberg, W-7900 Ulm(Donau), Germany, e-mail: mschmitt@neuro.informatik.uni-ulm.de

- Every linearly separable Boolean function $f : \{0,1\}^n \rightarrow \{0,1\}$ can be represented by integral weights satisfying

$$|w_i| \leq 2^{-n}(n+1)^{(n+1)/2}, \qquad |t| \leq 2^{-n}n(n+1)^{(n+1)/2}. \qquad (3)$$

- If rational numbers are permitted then there exist weights satisfying (2) and

$$|w_i| \leq 2^{-n}(n+1)^{(n+1)/2}, \qquad |t| \leq 2^{-n-1}(n+1)^{(n+3)/2} + \frac{1}{2}. \qquad (4)$$

- Furthermore, if we denote the truth value 0 by -1 then for a linearly separable function $g : \{-1,1\}^n \rightarrow \{-1,1\}$ rational weights can be found such that (2) and

$$|w_i| \leq 2^{-n}(n+1)^{(n+1)/2}, \qquad |t| \leq 2^{-n}(n+1)^{(n+1)/2} \qquad (5)$$

  hold simultaneously.

In this paper we show that inequalities (4) and (5) can be met even if we require all weights (including the threshold) to be integers. With that we obtain an improvement of (3) by approximately a factor of $\frac{1}{2}$.

In the next section we demonstrate some old and new properties of integral separating weight vectors. These will be used in the proof of the main theorem in Section 3. In the last section we make some remarks on the methods and speculations on further improvements.

2 Properties of integral separating weight vectors

The existence of an integral vector $(w_1, \ldots, w_n, t)$ satisfying the system of linear inequalities (2) is sufficient and necessary for a Boolean function $f : \{0,1\}^n \rightarrow \{0,1\}$ to be linearly separable. The same holds if we replace 0 by -1 and consider Boolean functions $g : \{-1,1\}^n \rightarrow \{-1,1\}$. Given an integral separating vector for g we obtain a — not necessarily integral — vector for the corresponding f by the following method.

Lemma 1 *Let $g : \{-1,1\}^n \rightarrow \{-1,1\}$ be linearly separable, represented by $(w_1, \ldots, w_n, t)$, and let $f : \{0,1\}^n \rightarrow \{0,1\}$ be defined by*

$$2f(x_1, \ldots, x_n) - 1 = g(2x_1 - 1, \ldots, 2x_n - 1) \qquad \textit{for all } x \in \{0,1\}^n.$$

Then f is represented by $(w_1, \ldots, w_n, \frac{1}{2}(t + \sum w_i))$.

Proof. The proof is due to Muroga [13, Theorem 1.3.1]. In the inequality system (2) corresponding to g

$$\begin{array}{lcll} w_1y_1 + \cdots + w_ny_n & \geq & t & \text{if } g(y) = 1 \\ w_1y_1 + \cdots + w_ny_n & \leq & t-1 & \text{if } g(y) = -1 \end{array} \qquad (y \in \{-1,1\}^n)$$

we add $\sum w_i$ on both sides of each inequation and divide by 2. Then we obtain the equivalent system

$$\begin{array}{lcll} w_1 \cdot \frac{1}{2}(y_1+1) + \cdots + w_n \cdot \frac{1}{2}(y_n+1) & \geq & \frac{1}{2}(t+\sum w_i) & \text{if } g(y)=1 \\ w_1 \cdot \frac{1}{2}(y_1+1) + \cdots + w_n \cdot \frac{1}{2}(y_n+1) & \leq & \frac{1}{2}(t+\sum w_i - 1) & \text{if } g(y)=-1. \end{array}$$

From that it can be derived that

$$\begin{array}{lcll} w_1x_1 + \cdots + w_nx_n & \geq & \frac{1}{2}(t+\sum w_i) & \text{if } f(x)=1 \\ w_1x_1 + \cdots + w_nx_n & < & \frac{1}{2}(t+\sum w_i) & \text{if } f(x)=0 \end{array} \qquad (x \in \{0,1\}^n),$$

in other words, (1) holds for $(w_1, \ldots, w_n, \frac{1}{2}(t+\sum w_i))$. ■

As can be seen, we could easily obtain an integral vector for f from an integral vector for g by multiplying by 2 the representation for f constructed in the proof. However, as we will show in the next lemma, either adding 1 to or subtracting from the threshold of g is always possible, thereby making $t + \sum w_i$ even.

Lemma 2 *Let $g : \{-1,1\}^n \to \{-1,1\}$ be linearly separable by the vector $(w_1, \ldots, w_n, t) \in \mathbb{Z}^{n+1}$.*

1. *There exists $t' \in \mathbb{Z}$ such that each of the following holds:*
   - *(a) $|t' - t| \leq 1$*
   - *(b) $t' + \sum w_i$ is even.*
   - *(c) $(w_1, \ldots, w_n, t')$ represents g.*
2. *There exists $t' \in \mathbb{Z}$ such that for all $y \in \{-1,1\}^n$*

$$\begin{array}{lcll} w_1y_1 + \cdots + w_ny_n & \geq & t'+1 & \text{if} \quad g(y)=1 \\ w_1y_1 + \cdots + w_ny_n & \leq & t'-1 & \text{if} \quad g(y)=-1. \end{array} \qquad (6)$$

Proof. Both assertions are proved if we show that $w \cdot y$ has the same parity for all y. This can be seen from

$$w_1y_1 + \cdots + w_i(-y_i) + \cdots w_ny_n \;=\; w \cdot y - 2w_i.$$

Therefore, either an increase or a decrease of t of at most 1 still yields a representation for g. ■

The reader should note that the t' of the two assertions need not to be identical.

By $\|(w_1, \ldots, w_n, t)\|_\infty$ we denote the maximum of $|w_1|, \ldots, |w_n|, |t|$. With the previous lemma we immediately have the following.

Lemma 3 *Let $g : \{-1,1\}^n \to \{-1,1\}$ be linearly separable by the vector $(w_1, \ldots, w_n, t) \in \mathbb{Z}^{n+1}$ and let $\gamma = \|(w_1, \ldots, w_n, t)\|_\infty$. If we replace -1 by 0 then the corresponding $f : \{0,1\}^n \to \{0,1\}$ is representable by a vector $(w'_1, \ldots, w'_n, t') \in \mathbb{Z}^{n+1}$ such that*

$$|w'_i| \;\leq\; \gamma, \qquad |t'| \;\leq\; \frac{n+1}{2} \cdot \gamma + \frac{1}{2}.$$

Proof. By Lemma 1, f can be represented by $(w_1, \ldots, w_n, \frac{1}{2}(t + \sum w_i))$. If $t + \sum w_i$ is not divisible by 2 we can increase or decrease t by 1 by virtue of Lemma 2. Therefore the absolute value of the new threshold t' is bounded by

$$|t'| \leq \frac{1}{2}(n \cdot \gamma + \gamma + 1) = \frac{n+1}{2} \cdot \gamma + \frac{1}{2}.$$

■

3 The upper bound

Before we state the theorem, we report a property of finite systems of linear inequalities in real linear spaces demonstrated by K. Fan [2] and employed in our proof, the so-called *principle of bounding solutions.*

Theorem 1 (Theorem 2 in [2]) *Let X be a real linear space of arbitrary dimension, finite or infinite, let $F_1, \ldots, F_p$ be linear functionals on X, and $\alpha_1, \ldots, \alpha_p$ real numbers, and let r be the maximum number of linearly independent linear functionals among $F_1, \ldots, F_p$. If the system*

$$F_i(x) \leq \alpha_i \qquad (1 \leq i \leq p) \tag{7}$$

has a solution then there exist r linearly independent functionals $F_{\nu_1}, \ldots, F_{\nu_r}$ among $F_1, \ldots, F_p$ such that every solution of the system

$$F_{\nu_k}(x) = \alpha_{\nu_k} \qquad (1 \leq k \leq r) \tag{8}$$

is also a solution of (7).

For the proof we refer the reader to reference [2].

Theorem 2 *Every linearly separable function $g : \{-1,1\}^n \to \{-1,1\}$ is representable by a vector $(w_1, \ldots, w_n, t)$ of integers satisfying*

$$\|(w_1, \ldots, w_n, t)\|_\infty \leq 2^{-n}(n+1)^{(n+1)/2}.$$

Proof. Let $g : \{-1,1\}^n \to \{-1,1\}$ be linearly separable. According to assertion 2 of Lemma 2 the system

$$\begin{array}{rcll} -w_1y_1 - \cdots - w_ny_n + t & \leq & -1 & \text{if} \quad g(y) = 1 \\ w_1y_1 + \cdots + w_ny_n - t & \leq & -1 & \text{if} \quad g(y) = -1 \end{array} \qquad (y \in \{-1,1\}^n) \tag{9}$$

has a solution in $(w_1, \ldots, w_n, t)$. For the sake of simplicity, we denote the threshold by w_{n+1} and let $y_{n+1} = -1$ for the rest of the proof. Then (9) appears as

$$\begin{array}{rcll} -w_1y_1 - \cdots - w_{n+1}y_{n+1} & \leq & -1 & \text{if} \quad g(y) = 1 \\ w_1y_1 + \cdots + w_{n+1}y_{n+1} & \leq & -1 & \text{if} \quad g(y) = -1 \end{array} \qquad (y \in \{-1,1\}^n) \tag{10}$$

having a solution in $(w_1, \ldots, w_{n+1})$. Let r denote the rank of the coefficient matrix on the left hand side of (10). By virtue of Theorem 1 there exist r elements $y^{(1)}, \ldots, y^{(r)} \in \{-1,1\}^n$ such that every solution of the system of equalities

$$\begin{array}{rcll} -w_1 y_1^{(k)} - \cdots - w_{n+1} y_{n+1}^{(k)} & = & -1 & \text{if} \quad g(y^{(k)}) = 1 \\ w_1 y_1^{(k)} + \cdots + w_{n+1} y_{n+1}^{(k)} & = & -1 & \text{if} \quad g(y^{(k)}) = -1 \end{array} \qquad (1 \le k \le r) \quad (11)$$

is also a solution of (10). Further, the rank of the coefficient matrix on the left hand side of (11) is equal to r, i.e. it contains an $(r \times r)$-submatrix consisting of columns, say $\lambda_1, \ldots, \lambda_r$, such that the system of equalities

$$\begin{array}{rcll} -w_{\lambda_1} y_{\lambda_1}^{(k)} - \cdots - w_{\lambda_r} y_{\lambda_r}^{(k)} & = & -1 & \text{if} \quad g(y^{(k)}) = 1 \\ w_{\lambda_1} y_{\lambda_1}^{(k)} + \cdots + w_{\lambda_r} y_{\lambda_r}^{(k)} & = & -1 & \text{if} \quad g(y^{(k)}) = -1 \end{array} \qquad (1 \le k \le r) \quad (12)$$

has a unique solution $(\widehat{w}_{\lambda_1}, \ldots, \widehat{w}_{\lambda_r})$. We can extend this solution of (12) to a solution of (11) (and thereby of (10)) by letting the components $i \notin \{\lambda_1, \ldots, \lambda_r\}$ be equal to 0. Because of the matrix on the left hand side of (12) being regular, we obtain the solution by Cramer's Rule

$$\widehat{w}_i = \frac{\Delta_i}{\Delta}, \qquad i \in \{\lambda_1, \ldots, \lambda_r\}$$

where Δ is the determinant of the coefficient matrix on the left hand side of (12) and Δ_i is the determinant obtained by replacing the i-th column by the right hand side of (12). We then have

$$\frac{1}{\Delta}(\Delta_1, \ldots, \Delta_{n+1}) \qquad \text{with } \Delta_i = 0 \text{ if } i \notin \{\lambda_1, \ldots, \lambda_r\}$$

as solution of (11) and thereby of (10) and, therefore,

$$(\Delta_1, \ldots, \Delta_{n+1})$$

as integral representation for g. Each $\Delta_i \neq 0$ is the determinant of an $(r \times r)$-matrix consisting solely of -1 and 1. Now we show how to get the factor 2^{r-1} from each of them. We multiply each row having -1 in the last component by -1 and obtain matrices with the last column consisting solely of 1. After adding this column to each of the remaining $r-1$ columns, columns $1, \ldots, r-1$ contain only 0 and 2 and we can factor 2^{r-1} outside of each determinant. If we reverse the previous multiplications we have

$$\Delta_i \;=\; 2^{r-1} \cdot w_i \qquad \text{for } i \in \{\lambda_1, \ldots, \lambda_r\}$$

where w_i is the determinant of an $(r \times r)$-matrix with entries $-1, 0, 1$. With the vector

$$(w_1, \ldots, w_{n+1}) \qquad \text{with } w_i = 0 \text{ if } i \notin \{\lambda_1, \ldots, \lambda_r\}$$

we obtain again an integral representation for g. Hadamard's Determinant Theorem implies $|\Delta_i| \le r^{r/2}$, therefore

$$|w_i| \;\le\; 2^{-r+1} r^{r/2} \qquad (1 \le i \le n+1).$$

Taking into account that r, the rank of the coefficient matrix on the left hand side of (10), is not greater than $n+1$ we have the statement of the theorem. ■

Having shown that (5) holds for integral weight vectors we can now easily derive that (4) is also satisfied by integers.

Corollary 1 *Every linearly separable function $f : \{0,1\}^n \rightarrow \{0,1\}$ is representable by a vector $(w_1, \ldots, w_n, t)$ of integers satisfying*

$$|w_i| \leq 2^{-n}(n+1)^{(n+1)/2}, \qquad |t| \leq 2^{-n-1}(n+1)^{(n+3)/2} + \frac{1}{2}.$$

Proof. The statement follows from Theorem 2 and Lemma 3. ■

Corollary 2 *Every linearly separable function $f : \{0,1\}^n \rightarrow \{0,1\}$ is representable by weights requiring not more than*

$$\left\lceil \log\left(2^{-n-1}(n+1)^{(n+3)/2} + \frac{3}{2}\right)\right\rceil + 1$$

bits.

Proof. An integral number z can be represented by $\lceil \log(|z|+1)\rceil + 1$ bits. ■

4 Remarks

In the three coarse steps

1. reducing a system of inequalities to a system of equalities by Fan's Theorem
2. solving this system by Cramer's Rule
3. bounding the solution by Hadamard's Theorem

the proof of Theorem 2 follows the original proof given by Muroga *et al.* in [14, Theorem 16] for linearly separable Boolean functions $f : \{0,1\}^n \rightarrow \{0,1\}$. However, in their proof the authors did not take into account the rank of the coefficient matrix which has to be regular for the application of Cramer's Rule. This was amended later in the proof by Muroga [13, Theorem 9.3.2.1] where the premisses require all weights to be greater than 0. This second proof, however, makes no longer use of Fan's Theorem but employs the existence of extreme points in convex sets.[1]

Polyhedral Theory is also suitable to make the essential inference in the proof. There, minimal faces and vertices of polyhedra are obtained by equating inequalities (see e.g. [19]). In Linear Programming these are closely related to

[1]This property can easily be derived by Fan's Theorem if the convex set is given by a system of linear inequalities.

the so-called *basic feasible solutions* (see e.g. [5, 8]). We chose to employ Fan's Theorem because of its generality and because it is purely algebraic and does not presuppose any further background.

The improvement we achieved with our proof is essentially based on observations on weight vectors for functions over $\{-1,1\}^n$ that were expressed in Lemma 2. Furthermore, we fell back upon a property of determinants exhibited by Williamson [20]: Every determinant of an $(n \times n)$-matrix consisting solely of -1 and 1 is divisible by 2^{n-1}. A proof of this fact can also be found in [1, p. 332]. Concerning Cramer's Rule we refer the reader to [9]. Hadamard's Determinant Theorem can be found in almost every book on matrix theory, [3] gives a quite elementary proof.

The question of the least upper bound is still open. It is well known by a counting argument [16] that $\Omega(n)$ is a lower bound for the binary length of a weight. Also, functions have been constructed by Goto [4] and Muroga [12] that require at least a value of $\alpha \cdot 2^n, 0 < \alpha < 1$, for the largest weight of every integral representation (see also [10, p. 406] for a simpler example).

How could one try to improve our result? Of course, the number of equations resulting from Fan's Theorem cannot be decreased below $n+1$. Furthermore, Hadamard's inequality cannot be improved because it is optimal for infinitely many n. Therefore, there does not seem to be a way to get a better upper bound by improving our proof steps. However, it turns out that matrices that satisfy equality in Hadamard's inequality yield weights of quite low absolute value. Therefore, we conjecture that better results can only be achieved by a rather different proof method.

Finally, if we consider functions that are incompletely specified then our proof is also applicable and gives the same result. Only if we severely restrict the size of the domain we get smaller bounds. We pursue this issue in a subsequent paper [18].

References

[1] D. K. Fadeev and I. S. Sominskii. *Problems in Higher Algebra.* W. H. Freeman, San Francisco, 1965.

[2] Ky Fan. On systems of linear inequalities. In Harold W. Kuhn and Albert W. Tucker, editors, *Linear Inequalities and Related Systems*, pages 99–156. Princeton University Press, Princeton, NJ, 1956. Annals of Mathematics Studies, Number 38.

[3] Joel N. Franklin. *Matrix Theory.* Prentice-Hall, Englewood Cliffs, NJ, 1968.

[4] Eiichi Goto. Threshold, majority, and bilateral switching devices. In H. Aiken and W. F. Main, editors, *Switching Theory in Space Technology*, pages 47–67. Stanford University Press, Stanford, CA, 1963.

[5] Martin Grötschel, László Lovász, and Alexander Schrijver. *Geometric Algorithms and Combinatorial Optimization.* Springer-Verlag, Berlin, 1988.

[6] Jiawei Hong. On connectionist models. Technical Report 87-012, University of Chicago, Department of Computer Science, Chicago, IL, June 1987.

[7] Sze-Tsen Hu. *Threshold Logic.* University of California Press, Berkeley and Los Angeles, 1965.

[8] Howard Karloff. *Linear Programming.* Birkhäuser, Boston, 1991.

[9] Serge Lang. *Linear Algebra.* Springer-Verlag, New York, third edition, 1989.

[10] P. M. Lewis and C. L. Coates. *Threshold Logic.* John Wiley & Sons, New York, 1967.

[11] Marvin Minsky and Seymour Papert. *Perceptrons: An Introduction to Computational Geometry.* The MIT Press, Cambridge, MA, expanded edition, 1988.

[12] Saburo Muroga. Lower bounds of the number of threshold functions and a maximum weight. *IEEE Transactions on Electronic Computers*, 14:136–148, 1965.

[13] Saburo Muroga. *Threshold Logic and Its Applications.* John Wiley & Sons, New York, 1971.

[14] Saburo Muroga, Iwao Toda, and Satoru Takasu. Theory of majority decision elements. *Journal of the Franklin Institute*, 271:376–418, 1961.

[15] Balas K. Natarajan. *Machine Learning: A Theoretical Approach.* Morgan Kaufmann, San Mateo, CA, 1991.

[16] Ian Parberry. A primer on the complexity theory of neural networks. In R. B. Banerji, editor, *Formal Techniques in Artificial Intelligence: A Sourcebook*, pages 217–268. Elsevier Science Publishers B. V. (North-Holland), Amsterdam, 1990.

[17] Prabhakar Raghavan. Learning in threshold networks. In D. Haussler and L. Pitt, editors, *Proceedings of the 1988 Workshop on Computational Learning Theory*, pages 19–27. Morgan Kaufmann, San Mateo, CA, 1988.

[18] Michael Schmitt. On the weight complexity of linearly separable sets, in preparation.

[19] Alexander Schrijver. *Theory of Linear and Integer Programming.* Wiley-Interscience Series in Discrete Mathematics and Optimization. John Wiley & Sons, Chichester, 1986.

[20] John Williamson. Determinants whose elements are 0 and 1. *The American Mathematical Monthly*, 53:427–434, 1946.

Liste der bisher erschienenen Ulmer Informatik-Berichte:
List of technical reports currently available from the University of Ulm:

91-01 KER-I KO, P. ORPONEN, U. SCHÖNING, O. WATANABE:
Instance Complexity.

91-02 K. GLADITZ, H. FASSBENDER, H. VOGLER:
Compiler-Based Implementation of Syntax-Directed Functional Programming.

91-03 ALFONS GESER:
Relative Termination.

91-04 JOHANNES KÖBLER, UWE SCHÖNING, JACOBO TORAN:
Graph Isomorphism is low for PP.

91-05 JOHANNES KÖBLER, THOMAS THIERAUF:
Complexity Restricted Advice Functions.

91-06 UWE SCHÖNING:
Recent Highlights in Structural Complexity Theory.

91-07 FREDERIC GREEN, JOHANNES KÖBLER, JACOBO TORAN:
The Power of the Middle Bit.

91-08 V. ARVIND, Y. HAN, L. HEMACHANDRA, J. KÖBLER, A. LOZANO, M. MUNDHENK, M. OGIWARA, U. SCHÖNING, R. SILVESTRI, T. THIERAUF:
Reductions to Sets of Low Information Content.

92-01 VIKRAMAN ARVIND, JOHANNES KÖBLER, MARTIN MUNDHENK:
Bounded Truth-Table and Conjunctive Reductions to Sparse and Tally Sets.

92-02 THOMAS NOLL, HEIKO VOGLER:
Top-down Parsing with Simultaneous Evaluation of Noncircular Attribute Grammars

92-03 PROGRAM AND ABSTRACTS:
17. Workshop über Komplexitätstheorie, effiziente Algorithmen und Datenstrukturen am 26. Mai 1992 in Ulm

92-04 V. ARVIND, J. KÖBLER, M. MUNDHENK
Lowness and the Complexity of Sparse and Tally Descriptions

92-05 JOHANNES KÖBLER
Locating P/poly Optimally in the Extended Low Hierarchy

92-06 ARMIN KÜHNEMANN, HEIKO VOGLER
Synthesized and inherited functions - a new computational model for syntax-directed semantics

92-07 HEINZ FASSBENDER, HEIKO VOGLER
A Universal Unification Algorithm Based on Unification-Driven Leftmost Outermost Narrowing.

92-08 UWE SCHÖNING
On Random Reductions from Sparse Sets to Tally Sets

92-09 HERMANN VON HASSELN, LAURA MARTIGNON
Consistency in Stochastic Networks

92-10 MICHAEL SCHMITT
A Slightly Improved Upper Bound on the Size of Weights Sufficient to Represent Any Linearly Separable Boolean Function

Ulmer Informatik-Berichte

ISSN 0939-5091

Herausgeber: Fakultät für Informatik

Universität Ulm, Oberer Eselsberg, W-7900 Ulm