JOURNAL OF MATHEMATICAL ANALYSIS AND APPLICATIONS 11, 226-235 (1965)

Iterative Synthesis of Threshold Functions*

E. WONG AND E. EISENBERG†

Department of Electrical Engineering,
University of California, Berkeley, California

Submitted by Lotfi Zadeh

INTRODUCTION

Consider the set of 2^n vectors $x = (x_1, \cdots, x_n)$ with components $x_k = \pm 1$, $k \leqslant n$. We denote this set by $\{-1, 1\}^n$. A function $F(x)$, defined on $x \in \{-1, 1\}^n$ with range $\{-1, 1\}$, will be called a Boolean function. Let $\text{Sgn}(y)$ be defined by

$$\text{Sgn}(y) = \begin{cases} 1, & y > 0 \\ 0, & y = 0. \\ -1, & y < 0 \end{cases} \tag{1}$$

Then, a threshold function $F(x)$ can be defined as a Boolean function for which there exists a vector w and a scalar w_0 such that

$$F(x) = \text{Sgn}(wx^T + w_0), \tag{2}$$

where the superscript denotes transpose. The geometric interpretation of a threshold function is clear. If we consider a Boolean function to be a binary valued function defined on the vertices of an n-cube, a threshold function assumes $+1$ and -1 on the vertices in such a way that the $+1$ vertices can be separated from the -1 vertices by a hyperplane. Thus, threshold functions are often referred to as linearly separable functions.

With recent developments in components which can be used to implement threshold functions, there has been a considerable interest in threshold functions from the points of view of both switching theory [1, 2] and adaptive pattern recognition devices [3, 4]. Since a single threshold function may represent a rather complex Boolean function, the possibility of threshold functions serving as basic building blocks in logical design appears attractive. In pattern recognition, the basic problem can often be reduced to one of

* This work was supported in part by the National Aeronautics and Space Administration Grant NsG 354.

† Deceased.

classifying vertices of an n-cube into two categories. Thus, the basic problem of pattern recognition is often equivalent to one of implementing a Boolean function, or a partially specified Boolean function, i.e., function defined on a subset of $\{-1, 1\}^n$.

While in pattern recognition problems one is primarily concerned with partially specified Boolean functions, the theory of iterative synthesis to be to be developed in this paper differs little whether the Boolean functions are completely or only partially specified. For simplicity of discussion, we shall deal only with completely specified functions. A second point to be noted is that every threshold function of n variables can be uniquely identified with a selfdual threshold function of $n + 1$ variables. Therefore, we shall consider only self-dual functions, i.e., those Boolean functions satisfying $F(-x) = -F(x)$, thus having an equal number of $+1$ and -1 vertices. It is clear that if a self-dual function is linearly separable, then it is separable by a plane passing through the origin. Thus, in these cases, the scalar w_0 in (2) can always be set equal to zero.

Rademacher-Walsh Representation of Boolean Functions

Let $T(t)$ be defined on $0 \leqslant t < 1$ as follows:

$$\begin{aligned} T(t) &= 2t, && 0 \leqslant t < \tfrac{1}{2} \\ &= 2t - 1, && \tfrac{1}{2} \leqslant t < 1. \end{aligned} \tag{3}$$

The Rademacher functions are defined by the following recursion relationships:

$$\begin{aligned} r_0(t) &= 1, && 0 \leqslant t < 1, \\ r_1(t) &= 1, && 0 \leqslant t < \tfrac{1}{2}, \\ &= -1, && \tfrac{1}{2} \leqslant t < 1, \end{aligned} \tag{4}$$

and

$$r_{k+1}(t) = r_k(T(t)), \qquad 0 \leqslant t < 1, \qquad k = 1, 2, \cdots.$$

The $n + 1$ Rademacher functions $r_k(t)$, $k = 0, 1, \cdots, n$, together with all distinct products $r_{i_1} r_{i_2} \cdots r_{i_m}$, $i_k \leqslant n$, $2 \leqslant m \leqslant n$(Walsh function), form a set of orthonormal basis for a 2^n-dimensional real vector space S_n, which includes all square-integrable functions on $[0, 1)$ which are piecewise constant on the 2^n intervals $(k/2^n \leqslant t < (k + 1)/2^n)$, $k = 0, 1, \cdots, 2^n - 1$. It is convenient to define inner product in the usual manner by

$$(f, g) = \int_0^1 f(t)\, g(t)\, dt, \qquad f, g \in S_n. \tag{5}$$

The Rademacher-Walsh functions so defined provide a convenient characterization of functions defined on the vertices of an n-cube [5].

We begin by establishing a one-to-one correspondence between the 2^n vertices of an n-cube with the 2^n intervals $[k/2^n, (k+1)/2^n)$, $k = 0, \cdots, 2^n - 1$. A vertex $x = (x_1, x_2, \cdots, x_n)$ is identified with the interval $k/2^n \leqslant t < (k+1)/2^n$ on which $(r_1(t), r_2(t), \cdots, r_n(t)) = x$. Thus, for example, with $n = 3$ the vertex $(1, -1, 1)$ corresponds to the interval, $[\frac{1}{4}, \frac{3}{8})$. With this correspondence functions defined on $\{-1, 1\}^n$ are mapped one-to-one on S_n. Thus, a Boolean function of n variables $F(x)$ is characterized by a function $\phi(t) \in S_n$, with $|\phi(t)| = 1$, $0 \leqslant t < 1$.

An advantage of the Rademacher-Walsh representation is that threshold functions have a simple characterization. A threshold function $\phi(t)$ is a Boolean function which can be written as

$$\phi(t) = \operatorname{Sgn}\left[\sum_{k=0}^{n} w_k r_k(t)\right]. \tag{6}$$

Equation (7) results merely from a rewriting of (2) using the Rademacher-Walsh representation. Now, let S_n^- be a subspace of S_n generated by the Rademacher functions alone, i.e., an element of S_n^- is a function which can be written as a linear combination of r_k's without using the products. The problem of realizing a threshold function is one of finding a $y(t) \in S_n^-$ such that $\phi(t) = \operatorname{Sgn}[y(t)]$.

On the other hand, Boolean functions which are not threshold admit an equally simple characterization. Let $X_+(F)$ denote the set of vertices on which $F(x) = 1$, and $X_-(F)$ the set of vertices on which $F(x) = -1$. A Boolean function $F(x)$ is not linearly separable if and only if the convex hulls of $X_+(F)$ and $X_-(F)$ intersect. If we consider only self-dual functions then $X_+(F)$ and $X_-(F)$ each contains 2^{n-1} vertices. Denoting elements of X_+ and X_- by $u^{(i)}$ and $v^{(j)}$ respectively, then the convex hulls having a nonempty intersection are equivalent to the existence of a set of λ_i and μ_i such that

$$\sum_{i=1}^{2^{n-1}} \lambda_i u^{(i)} = \sum_{j=1}^{2^{n-1}} \mu_j v^{(j)}, \qquad \lambda_i, \mu_j \geqslant 0,$$

$$\sum_{i=1}^{2^{n-1}} \lambda_i = \sum_{j=1}^{2^{n-1}} \mu_j = 1. \tag{7}$$

Now, a vertex $x = (x_1, \cdots, x_n)$ defines an interval $I(x) \subset [0, 1)$. Let $U_x(t) = 1$, $t \in I(x)$ and $U_x(t) = 0$, $t \notin I(x)$. The function $U_x(t)$ can be written as

$$U_x(t) = \prod_{k=1}^{n} \left[\frac{1 + x_k r_k(t)}{2}\right]. \tag{8}$$

For a given Boolean function $F(x) \leftrightarrow \phi(t)$, define a function $z(t)$ on $[0, 1)$ as follows:

$$\begin{aligned} z(t) &= \lambda_i\,, && t \in I(u^{(i)}),\\ &= -\mu_j\,, && t \in I(v^{(j)}),\\ \lambda_i\,, \mu_j &\geqslant 0, \sum_i \lambda_i = \sum_j \mu_j = 1. \end{aligned} \tag{9}$$

Thus defined, $z(t)$ has the property

$$\operatorname{Sgn}[z(t)] = \phi(t), \qquad \forall t \ni z(t) \neq 0. \tag{10}$$

Furthermore, $z(t)$ can be written using (8) as

$$z(t) = \sum_{i=1}^{2^{n-1}} \frac{\lambda_i}{2^n} \prod_{k=1}^{n} [1 + u_k^{(i)} r_k(t)] - \sum_{i=1}^{2^{n-1}} \frac{\mu_i}{2^n} \prod_{k=1}^{n} [1 + v_k^{(i)} r_k(t)]. \tag{11}$$

where $u_k^{(i)}$, $v_k^{(i)}$ are the kth components of $u^{(i)}$ and $v^{(i)}$ respectively. If we expand the products in (11), we find

$$z(t) = \frac{1}{2^n}\left(\sum_{i=1}^{2^{n-1}} \lambda_i - \sum_{i=1}^{2^{n-1}} \mu_i\right) + \frac{1}{2^n}\sum_{k=1}^{n} r_k(t) \cdot \left[\sum_{i=1}^{2^{n-1}} \lambda_i u_k^{(i)} - \sum_{i=1}^{2^{n-1}} \mu_i v_k^{(i)}\right]$$

$$+ \text{ terms involving products of } r_k(t). \tag{12}$$

The first term on the right hand side of (12) is automatically equal to zero. The second term is identically zero if and only if the condition (7) is satisfied. If we denote by S_n^+ the subspace generated by the Walsh functions only, then a Boolean function $\phi(t)$ is nonthreshold if and only if there exists a $z(t) \in S_n^+$ not identically zero such that (10) is satisfied.

To summarize, the threshold property, or lack of it, of a Boolean function $\phi(t)$ (self-dual) can be characterized as follows:

(a) $\phi(t)$ is a threshold function if and only if there exists a $y(t) \in S_n^-$ such that

$$\operatorname{Sgn}[y(t)] = \phi(t), \qquad 0 \leqslant t < 1. \tag{13}$$

(b) $\phi(t)$ is not a threshold function if and only if there exists a $z(t) \in S_n^+$ such that

$$\int_0^1 |\, z(t) \,|\, dt = 1, \tag{14}$$

and

$$\operatorname{Sgn}[z(t)] = \phi(t), \qquad \forall t \ni z(t) \neq 0. \tag{15}$$

Condition (14) merely serves to prevent $z(t)$ from being identically zero, in which case (15) is trivially true. This simple dual characterizations of threshold property makes the Rademacher-Walsh representation natural for considering threshold problems. They are also basis for the iterative procedures discussed below.

Iterative Procedure

The problem of deciding whether a given Boolean function is linearly separable and of constructing a realization when it is can in most instances (all except some degenerate cases) be solved by a finite iterative procedure. The iterative procedure in a more general form is embodied in the following theorem:

Theorem 1. *Let $\phi(t)$ be a Boolean function, M be a subspace of S_n and denote by P the projection on M. If there exists a $y(t) \in M$ such that*

$$\phi(t) = \operatorname{Sgn}[y(t)], \qquad 0 \leqslant t < 1, \tag{16}$$

then the following sequence terminates:

$$\begin{aligned} f_0(t) &= P\phi(t), \\ f_{m+1}(t) &= f_m(t) + P[\phi(t) - \operatorname{Sgn} f_m(t)]. \end{aligned} \tag{17}$$

That is, there exists an integer N such that $m \geqslant N$ implies $f_{m+1}(t) = f_m(t)$, and $\phi(t) = \operatorname{Sgn} f_m(t)$.

Proof. By assumption there exists a $y \in M$ satisfying (16). Taking inner product of (17) with y, we find

$$(y, f_{m+1}) = (y, f_m) + (y, \operatorname{Sgn} y - \operatorname{Sgn} f_m). \tag{18}$$

Suppose, contrary to the hypothesis of the theorem, the sequence does not terminate. Then, $P[\phi - \operatorname{Sgn} f_m] \neq 0$ for all m. It follows that

$$(y, \operatorname{Sgn} y - \operatorname{Sgn} f_m) \geqslant \frac{1}{2^n} \min_t |\, y(t) \,| = \theta > 0, \tag{19}$$

where the last inequality follows from the fact that to satisfy (16) $y(t)$ can never be zero in [0, 1). Using (18) in (19), we find

$$(y, f_{m+1}) \geqslant (y, f_m) + \theta, \tag{20}$$

or

$$(y, f_m) \geqslant (y, f_0) + m\theta = (y, \operatorname{Sgn} y) + m\theta. \tag{21}$$

Whence, by the Schwarz inequality we find

$$\| f_m \|^2 \geqslant \frac{(y, f_m)^2}{\| y \|^2} > \frac{m^2\theta^2}{\| y \|^2}. \tag{22}$$

On the other hand, (17) implies that

$$\begin{aligned} \| f_{m+1} \|^2 &= \| f_m \|^2 + \| P[\phi - \operatorname{Sgn} f_m] \|^2 - 2(f_m , \operatorname{Sgn} f_m - \phi) \\ &\leqslant \| f_m \|^2 + 4. \end{aligned} \tag{23}$$

It follows from (23) that

$$\| f_m \|^2 \leqslant \| f_0 \|^2 + 4m. \tag{24}$$

Inequalities (22) and (24) clearly cannot both hold for all m. Thus, the sequence (17) must terminate.

Further, if the sequence terminates, it implies the existence of an N such that

$$P \operatorname{Sgn} f_m(t) = P\phi(t), \qquad \forall m \geqslant N. \tag{25}$$

Now, let $y(t)$ satisfy the conditions of the theorem, i.e.,

$$y(t) \in M,$$

and

$$\operatorname{Sgn} y(t) = \phi(t).$$

It follows that $y(t) \neq 0$, $0 \leqslant t < 1$. From (25) we find

$$\int_0^1 y(t)\,\phi(t)\,dt = \int_0^1 |\,y(t)\,|\,dt = \int_0^1 y(t) \operatorname{Sgn} f_m(t)\,dt, \qquad m \geqslant N. \tag{26}$$

The fact that $y(t) \neq 0$, $0 \leqslant t < 1$ and (26) imply

$$\phi(t) = \operatorname{Sgn} f_m(t), \qquad \forall m \geqslant N.$$

COROLLARY 1. *Let $F(x) \leftrightarrow \phi(t)$ be a threshold function. Let $M = S_n^-$. Then (17) terminates in a realization for $\phi(t)$.*

PROOF. It suffices to note that if $\phi(t)$ is a threshold function, then there is a $y(t) \in S_n^-$ satisfying (16).

COROLLARY 2. *Let $F(x) \leftrightarrow \phi(t)$ be a Boolean function for which condition (7) can be satisfied with λ_i, $\mu_j > 0$ for all i, j. (This means that the convex hulls of $X^+(F)$ and $X^-(F)$ intersect at at least one point which is in the* relative interior *of both convex hulls.) Let* $M = S_n^+$. *Then, (17) terminates in an* $f(t) \in S_n^+$ *such that*

$$\phi(t) = \operatorname{Sgn} f(t). \tag{27}$$

PROOF. It suffices to note that if condition (7) can be satisfied with λ's and μ's strictly positive, then conditions (14) and (17) can be satisfied with a $z(t)$ which is nowhere zero on [0, 1).

It is seen that Corollaries (1) and (2) together cover all Boolean functions which are either linearly separable, or whose convex hulls generated by the $+1$ and -1 vertices intersect in an interior point. There remain cases where the Boolean function is not linearly separable, but the convex hulls of X^+ and X^- intersect only on the boundaries. The Boolean function $AB + \bar{A}C$ is an example of this. It does not mean, however, that in these degenerate cases the sequence (17) will necessarily fail to terminate. The function $AB + \bar{A}C$ is an example for which the interation (17) terminates even though the conditions of the theorem are not satisfied.

The iterative procedure (17) can be modified to cover the degenerate cases for which (15), but not (27), is satisfied. However, only convergence, not finite termination, can be proved for the modified procedure.

THEOREM 2. *Let* $\phi(t)$ *be a Boolean function,* M *be a subspace of* S_n. *Denote by* P *the projection on* M, *and* P^+ *the projection on* $M^\perp$, *(the perpendicular subspace to* M*). Suppose there exists a* $y(t) \in M$, *such that* $y(t)$ *is not identically zero, and*

$$\operatorname{Sgn} y(t) = \phi(t), \qquad \textit{for all} \qquad t \ni y(t) \neq 0. \tag{28}$$

Let $f_m(t)$ *be recursively determined as follows:*

$$f_0(t) = \frac{1}{|K_0|} P\phi(t),$$

$$f_{m+1}(t) = \frac{1}{|K_{m+1}|} \{f_m(t) + \tfrac{1}{2} P \,|f_m(t)|\, [\phi(t) - \operatorname{Sgn} f_m(t)]\}, \tag{29}$$

and

$$K_m^2 = 1 - \tfrac{1}{2} \int_0^1 f_m^2(t)\,[1 - \phi(t) \operatorname{Sgn} f_m(t)]\, dt - \tfrac{1}{4} \| P^+ |f_m| \phi \|^2, \tag{30}$$

and

$$K_0^2 = \| P\phi \|^2.$$

Then, $f_m(t)$ *converges to a solution to (28), i.e.,* $\lim_{m\to\infty} f_m(t) = f(t)$, $f(t)$ *not identically equal to zero, and*

$$\operatorname{Sgn} f(t) = \phi(t), \qquad \textit{whenever} \qquad f(t) \neq 0. \tag{31}$$

PROOF. First, we will show that under the stated conditions $K_m^2 \neq 0$, for all m. This can be done by induction as follows: without loss of generality, let $y(t) \in M$ satisfy (28) and $\| y \|^2 = 1$. Then

$$K_0^2 = \| P\phi \|^2 \geqslant (y, P\phi)^2 = (y, \phi)^2 = \left[\int_0^1 | y(t) | \, dt\right]^2 > 0. \tag{32}$$

Assume $K_k^2 \neq 0$, $0 \leqslant k \leqslant m$. Then $\| f_k \|^2 = 1$, $0 \leqslant k \leqslant m$. Now, let $\Delta_m(t)$ be defined by

$$\Delta_m(t) = \tfrac{1}{2} P \, | f_m(t) | \, [\phi(t) - \operatorname{Sgn} f_m(t)]. \tag{33}$$

Then,

$$K_{m+1}^2 = \| f_m(t) + \Delta_m(t) \|^2 \geqslant [(y, f_m) + (y, \Delta_m)]^2. \tag{34}$$

Now,

$$(y, \Delta_k) = \tfrac{1}{2} \int_0^1 y(t) \, | f_k(t) | \, [\phi(t) - \operatorname{Sgn} f_k(t)] \, dt$$

$$= \tfrac{1}{2} \int_0^1 y(t) \, | f_k(t) | \, [\operatorname{Sgn} y(t) - \operatorname{Sgn} f_k(t)] \, dt \geqslant 0, \tag{35}$$

$$(y, f_m) = \frac{1}{| K_m |} [(y, f_{m-1}) + (y, \Delta_{m-1})] \geqslant \frac{1}{| K_m |} (y, f_{m-1})$$

$$\geqslant \left(\prod_{k=0}^{m} \frac{1}{| K_m |}\right) \int_0^1 | y(t) | \, dt. \tag{36}$$

Using (35) and (36) in (34), we find that

$$K_{m+1}^2 \geqslant \left(\int_0^1 | y(t) | \, dt\right)^2 \prod_{k=0}^{m} \frac{1}{K_k^2} > 0, \tag{37}$$

completing the induction.

Now, from the Schwarz inequality and (36), we have that

$$\| f_m \|^2 \geqslant (y, f_m)^2 \geqslant \left(\int_0^1 | y(t) | \, dt\right) \prod_{k=0}^{m} \frac{1}{K_k^2} \, . \tag{38}$$

On the other hand, K_m has been chosen so as to make

$$\| f_m \|^2 = 1. \tag{39}$$

Further,

$$K_m^2 \leqslant 1. \tag{40}$$

Consistency among (38), (39), and (40) requires that

$$\lim_{m \to \infty} K_m^2 = 1, \tag{41}$$

which from (30) implies that

$$\lim_{m\to\infty} \int_0^1 f_m^2(t)\,[1 - \phi(t)\,\mathrm{Sgn}\,f_m(t)]\,dt = 0. \tag{42}$$

Equation (42) in turn implies that

$$\lim_{m\to\infty} \mathrm{Sgn}\,f_m(t) = \mathrm{Sgn}\,f(t) = \phi(t), \qquad \text{whenever} \qquad f(t) \neq 0. \tag{43}$$

The fact that $\| f_m \|^2 = 1$ for all m already insured that $f(t)$ is not identically zero.

If we let $M = S_n^+$ in Theorem 2, then the iterative procedure is directly applicable to cases where (15) is satisfied.

Conclusion

The class of iterative procedures treated in this note may be called "error-correcting" procedures, since at each step the iteration proceeds or terminates according as to whether a solution has already been obtained. Further, the incremental change is a simple projection of the "error". The procedures considered here are closely related to those considered by Agmon [6], Motzkin and Schoenberg [7], Novikoff [8], and Block [9].

It should be noted that the conclusion of Theorem 1 is unchanged if (17) is modified to read

$$f_{m+1}(t) = f_m(t) + a_m P[\phi(t) - \mathrm{Sgn}\,f_m(t)],$$

provided that a_m satisfies the condition

$$a_m \geqslant 0,$$

$$\lim_{m\to\infty} \frac{(\sum_{k=0}^{m} a_k)^2}{\sum_{k=0}^{m} a_k^2} = \infty.$$

A 7090 computer program has been written, which implements (17) in finding realization of threshold functions. Some preliminary results for 6-variable functions indicate that the procedure is fairly efficient, the longest one requiring 12 iterations.

References

1. R. O. Winder. Single stage threshold logic. *Proc. AIEE Symp. Switching Circuit Theory and Logical Design*, 1962, pp. 321-332.

2. S. Muroga. Majority logic and problems of probabilistic behavior. "Self Organizing Systems," pp. 243-281. Spartan Books, 1962.
3. F. Rosenblatt. "Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms." Spartan Books, 1961.
4. B. Widrow and M. Hoff. "Adaptive Switching Circuits," Tech. Rept. No. 1553-1 (Stanford Electronics Lab., 1960).
5. S. W. Golomb. On the classification of Boolean functions. *Trans.* 1959 *Intern. Symp. Circuit and Inform. Theory. IRE Trans. Special Suppl.* **IT-5** (1959), 176-186.
6. S. Agmon. The relaxation method for linear inequalities. *Canad. J. Math.* **6** (1954), 382-392.
7. T. S. Motzkin and I. J. Schoenberg. The relaxation method for linear inequalities. *Canad. J. Math.* **6** (1954), 393-404.
8. A. Novikoff. On convergence proofs for perceptrons. *Proc.* 1962 *Symp. Math. Theory Automata*, pp. 615-622 (Polytechnic Institute of Brooklyn, 1963).
9. H. Block. The perception: A model for brain functioning I. *Rev. Mod, Phys.* **34** (1962), 123-135.