JOURNAL OF COMPUTER AND SYSTEM SCIENCES **52**, 349–356 (1996)
ARTICLE NO. 0026

Gavin J. Gibson

Biomathematics and Statistics Scotland, University of Edinburgh, The King's Buildings, Mayfield Road, Edinburgh EH9 3JZ, United Kingdom

Received July 1, 1992

This paper considers the classification properties of two-layer networks of McCulloch–Pitts units from a theoretical point of view. In particular we consider their ability to realise exactly, as opposed to approximate, bounded decision regions in $\mathbb{R}^2$. The main result shows that a two-layer network can realise exactly any finite union of bounded polyhedra in $\mathbb{R}^2$ whose bounding lines lie in general position, except for some well-characterised exceptions. The exceptions are those unions whose boundaries contain a line which is ''inconsistent,'' as described in the text. Some of the results are valid for $\mathbb{R}^n$, $n \geqslant 2$, and the problem of generalising the main result to higher-dimensional situations is discussed.

1. INTRODUCTION

In this paper we consider a mathematical problem which arises in the study of two-layer neural networks of McCulloch–Pitts neurones [1] (see also MADALINE networks [2]), which incorporate one hidden layer of units, a single output unit, and which use a step-function for the node activation function. Such a network accepts an input $\mathbf{x} \in \mathbb{R}^n$ which is passed to the m units in the hidden layer of the network. Each hidden-layer unit is specified by a weight vector $\mathbf{w}_i \in \mathbb{R}^n$ and a bias q_i, $1 \leqslant i \leqslant m$, and computes as its output $y_i = f(\mathbf{w}_i \cdot \mathbf{x} + q_i)$, where $f\colon \mathbb{R} \to \{0, 1\}$ is defined by $f(s) = 1$, if $s \geqslant 0$ and $f(s) = 0$ otherwise. In the networks which we consider the (binary) vector of hidden-layer outputs $\mathbf{y} = (y_1, ..., y_m)$ is passed as input to the single unit in the output layer which is specified by a weight vector $\mathbf{a} \in \mathbb{R}^m$ and bias $b \in \mathbb{R}$. The network therefore implements a function $g\colon \mathbb{R}^n \to \{0, 1\}$, defined by

$$g(\mathbf{x}) = f(\mathbf{a} \cdot \mathbf{y} + b),$$

where

$$y_i = f(\mathbf{w}_i \cdot \mathbf{x} + q_i),\ 1 \leqslant i \leqslant m.$$

Now let $S \subset \mathbb{R}^n$ be a bounded set and let $h\colon \mathbb{R}^n \to \{0, 1\}$ be the function defined by $h(x) = 1$, if $\mathbf{x} \in S$, and $h(\mathbf{x}) = 0$ otherwise. We say that S is *exactly realisable* if there exists a two-layer network of the above form for which $g(\mathbf{x}) = h(\mathbf{x})$ almost everywhere. The question we investigate is that of characterising realisable subsets of $\mathbb{R}^n$.

Many researchers have considered networks with three or more layers of units, whose classification capabilities are more easily understood (see later). In a three-layer network of McCulloch–Pitts units the output vector from the first hidden layer, $\mathbf{y} \in \mathbb{R}^m$ is passed as input to a second hidden layer consisting of m' units, specified by weight vectors $\mathbf{a}_j$ and biases b_j, $1 \leqslant j \leqslant m'$. The output vector from this second hidden layer, $\mathbf{y}' = (y'_1, ..., y'_{m'})$ is then passed as input to the nodes in the output layer. In the case where the output layer consists of a single unit, with weight vector $\mathbf{a}' \in \mathbb{R}^{m'}$ and bias b', the three-layer network implements a function $g\colon \mathbb{R}^n \to \{0, 1\}$ defined by $g(\mathbf{x}) = f(\mathbf{a}' \cdot \mathbf{y}' + b')$, where $y'_j = f(\mathbf{a}_j \cdot \mathbf{y} + b_j)$, $1 \leqslant j \leqslant m'$, and $y_i = f(\mathbf{w}_i \cdot \mathbf{x} + q_i)$, $1 \leqslant i \leqslant m$.

There are, nevertheless, several results in the literature which deal with the capabilities of two-layer networks. The approximation results of Hornik *et al.* [3], Cybenko [4] and Funahashi [5] demonstrate that all bounded, measurable sets are "almost" realisable. Li [6] has appealed to these results to show that any two disjoint, compact subsets of $\mathbb{R}^n$ can be separated by a two-layer network. The abilities of neural networks to classify finite input spaces have been considered by several researchers (see, e.g. [7–9]). Exactly realisable sets (which, in essence can be separated from their complements in $\mathbb{R}^n$) have not received the same attention. The subject was treated in [10] by Gibson and Cowan, where examples of realisable sets and unrealisable sets were constructed, but no general characterisation of realisability was given. A similar approach to the problem, relating the construction of decision regions to problems involving separation of subsets of hypercube vertices, was taken in [11]. Some general classes of realisable sets have since been described in [12–14].

The main contribution of this paper is to provide a simple geometrical characterization of the bounded realisable sets of $\mathbb{R}^2$, whose bounding lines are in general position. The main result of the paper demonstrates that, with the exception of some nongeneric cases, the abilities of two- and three-layer networks of McCulloch–Pitts neurones are similar when inputs are two-dimensional. Examples are given which show that our main result does not hold if the assumptions of boundedness and general position of bounding lines are relaxed. We also discuss how readily the result might be generalised to higher dimensional situations.

0022-0000/96 $18.00

2. BOUNDED REALISABLE SUBSETS OF $\mathbb{R}^2$

We begin by generalising our notion of realisability. Let F_n denote the set of all functions $g\colon \mathbb{R}^n \to \mathbb{R}$ of the form

$$g(\mathbf{x}) = \sum_{i=1}^{m} a_i f(\mathbf{w}_i \cdot \mathbf{x} + q_i), \quad m \in \mathbb{N},\ \mathbf{w}_i \in \mathbb{R}^n, \qquad (2.1)$$
$$a_i, q_i \in \mathbb{R}, \quad 1 \leqslant i \leqslant m.$$

Suppose $S \subset C \subset \mathbb{R}^n$. We say S is *realisable in* C if there exists $g \in F_n$ such that (almost everywhere) $g(\mathbf{x}) > 0$ if $\mathbf{x} \in S$ and $g(\mathbf{x}) < 0$ if $\mathbf{x} \in C \backslash S$. In the notation of the previous section S is realisable if it realisable in $\mathbb{R}^n$.

It is clear from (2.1) that a realisable set must be a finite union of polyhedral sets in $\mathbb{R}^n$, ignoring discrepancies of zero measure. This follows from the fact that the function g in (2.1) is constant over any connected component (cell) of $\mathbb{R}^n \backslash \bigcup_{i=1}^m H_i$ where H_i is the hyperplane defined by $\mathbf{w}_i \cdot \mathbf{x} + q_i = 0$, for those $\mathbf{w}_i \neq \mathbf{0}$. Each cell is a polyhedral set. However, it is well known (e.g., [10, 14, 15]) that not all unions of polyhedral sets are realisable. Before identifying a general class of unrealisable sets we give some definitions.

Definition 1. Let $S \subset \mathbb{R}^n$ be a union of polyhedral sets and let $H_1, ..., H_k$ be the hyperplanes which have a nontrivial (i.e., $(n-1)$-dimensional) intersection with the boundary of S, $B(S)$. Then we say $H_1, ..., H_k$ are the *essential* hyperplanes of $B(S)$.

Definition 2. Let P_j, $1 \leqslant j \leqslant s$, denote the connected components of $\mathbb{R}^n \backslash \bigcup_{i=1}^m H_i$. The hyperplane H_i is said to be *inconsistent* if there exist P_{j_1}, P_{j_2}, P_{j_3} and P_{j_4} such that

(i) $P_{j_1}, P_{j_2} \subset S$ and $P_{j_3}, P_{j_4} \subset \mathbb{R}^n \backslash S$,

(ii) $\bar{P}_{j_1} \cap \bar{P}_{j_3} \cap H_i$ and $\bar{P}_{j_2} \cap \bar{P}_{j_4} \cap H_i$ are both $(n-1)$-dimensional sets, where $\bar{A}$ denotes the closure of the set A, and

(iii) $\mathbf{w}_i \cdot \mathbf{x} + q_i > 0$, $\forall \mathbf{x} \in P_{j_1} \cup P_{j_4}$, $\mathbf{w}_i \cdot \mathbf{x} + q_i < 0$, $\forall \mathbf{x} \in P_{j_2} \cup P_{j_3}$.

The above definition can be restated more intuitively. An inconsistent (essential) hyperplane is one whose intersection with $B(S)$ contains two regions which jointly have the following property. On one of these regions, crossing H_i in a given direction takes one from S into $R^n \backslash S$, whilst on the other region crossing H_i *in the same direction* takes one from $\mathbb{R}^n \backslash S$ into S. Figure 1 is a simple example of a set whose boundary has an inconsistent hyperplane.

Concerning inconsistent hyperplanes we have the following well-known "folk theorem" of neurocomputing.

Lemma 3 (*cf.* [10, 14, 15]). *Let S be a finite union of polyhedral sets whose boundary has an essential hyperplane, H, which is inconsistent. Then S is unrealisable in $\mathbb{R}^n$.*

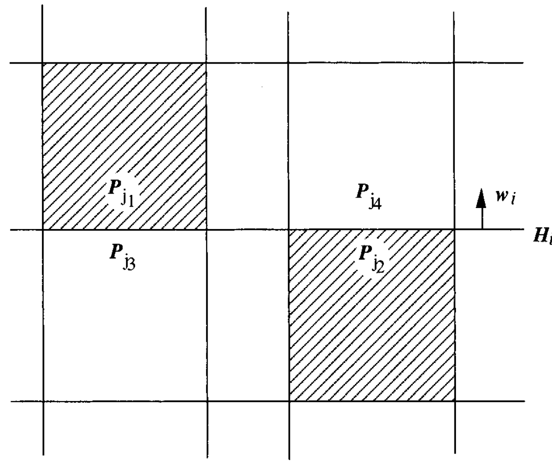

FIG. 1. A region whose boundary incorporates an inconsistent essential hyperplane.

Proof. Suppose otherwise so that there exists a function g, as defined in (2.1) such that $g(\mathbf{x}) > 0$, for all $\mathbf{x} \in S$, and $g(\mathbf{x}) < 0$ otherwise. We may assume without loss of generality that H is defined by $\mathbf{w}_i \cdot \mathbf{x} + q_i = 0$ for some i in (2.1). It is easily verified that the conditions (i)–(iii) force the coefficient a_i to be simultaneously greater than, and less than zero—a contradiction. ∎

Lemma 3 identifies one particular pathology which renders a union of polyhedral sets unrealisable. A logical question to ask is whether the converse of Lemma 3 is true—must all unrealisable sets have an essential hyperplane which is inconsistent? In our main result, Theorem 4, we prove that, apart from some nongeneric exceptions, the converse of Lemma 3 is true when $n = 2$.

Theorem 4. *Let $S \subset \mathbb{R}^2$ be a finite union of bounded polyhedral sets for which no three essential hyperplanes (lines) intersect. Then S is realisable if and only if no essential hyperplane is inconsistent.*

Before we prove Theorem 4, we give some auxiliary results which are required by the proof. The first of these, which generalises a result of [13], allows us to focus our attention on the boundary of a set to determine whether it is realisable or not.

Proposition 5. (*cf.* [13]). *Let $C \subset \mathbb{R}^n$ be compact, let $S \subset C$ have boundary $B(S)$ in C and let N be any open subset of C containing $B(S)$. Suppose that $S \cap N$ is realisable in N. Then S is realisable in C.*

Proof. Since $S \cap N$ is realisable in N, there exist $g \in F_n$ and $\delta > 0$ such that $g > \delta$ on $S \cap N$ and $g < -\delta$ on $N \backslash S$. Now g is bounded so that there exists $M \in R$ such that $|g(\mathbf{x})| < M$ for all $\mathbf{x} \in C$. Consider the function $\mu\colon C \to \mathbb{R}$, defined by

$$\mu(\mathbf{x}) = d(\mathbf{x}, C \backslash S) - d(\mathbf{x}, S),$$

where $d(\mathbf{x}, A) = \inf_{a \in A} \|\mathbf{x} - \mathbf{a}\|$. Now μ is continuous on C, $\mu \geqslant 0$ on S and $\mu \leqslant 0$ on $C \backslash S$. Furthermore, since $B(S)$ is compact, there exists $\gamma > 0$ such that $|\mu(\mathbf{x})| > \gamma$ for all $\mathbf{x} \in C \backslash N$. Now let $\mu' = (M + \delta)\, \mu/\gamma$. Since μ' is continuous we can choose $h \in F_n$ such that $|h(\mathbf{x}) - \mu'(\mathbf{x})| < \delta$, for all $\mathbf{x} \in C$, by appealing to the approximation results of [3–5]. It is easily verified that $g + h \in F_n$ satisfies $h(\mathbf{x}) + g(\mathbf{x}) > 0$ for almost all $\mathbf{x} \in S$, and $g(\mathbf{x}) + h(\mathbf{x}) < 0$ for almost all $\mathbf{x} \in C \backslash S$. It follows that S is realisable in C. ▮

LEMMA 6. *Let $\mathbf{x}_1, ..., \mathbf{x}_k$ be distinct points in $\mathbb{R}^n$ and let $y_1, ..., y_k \in \mathbb{R}$. Then there exist $\varepsilon > 0$, and a function $g \in F_n$ (as defined in* (2.1) *such that $g(\mathbf{x}_i) = y_i$, for all $\mathbf{x} \in N(\mathbf{x}_i, \varepsilon)$, $1 \leqslant i \leqslant k$.*

Proof. Let, $n = 1$ and suppose that $x_1 < x_2 \cdots < x_k$, without loss of generality. Choose $\varepsilon < \min\{(x_{i+1} - x_i)/2\}$. Now let g be defined by

$$g(x) = \sum_{i=1}^{k} a_i f(x - q_i),$$

where $a_1 = y_1$, $q_1 < x_1 - \varepsilon$, and for $i \geqslant 2$, $a_i = y_i - y_{i-1}$, $q_i = (x_{i-1} + x_i)/2$. Then $g(x) = y_i$ for all $x \in N(x_i, \varepsilon)$, $1 \leqslant i \leqslant k$, as required.

For $n \geqslant 2$, we note that there exists $\mathbf{w} \in \mathbb{R}^n$ such that $\mathbf{w} \cdot \mathbf{x}_i \neq \mathbf{w} \cdot \mathbf{x}_j$, $i \neq j$. Taking the inner product of each $\mathbf{x}_i$ with any such $\mathbf{w}$ reduces the problem to the case $n = 1$ considered above. An identical argument has been used previously to reduce from n dimensions to a single dimension in [8, Lemma 5.1].

Proof of Theorem 4. The general argument given below is illustrated in the context of a particular example, where S is the union of shaded rectangles in Fig. 2. From Lemma 3 it suffices to show that if no essential hyperplane of $B(S)$ is inconsistent then S is realisable. Therefore let S satisfy the conditions of the theorem and further suppose that no essential hyperplane (line) is inconsistent. Let C denote a compact polyhedral set containing S (see Fig. 2). If S is realisable in C, then a simple argument shows that it is realisable in $\mathbb{R}^2$.

Let $H_i = \{\mathbf{x} \mid \mathbf{w}_i \cdot \mathbf{x} + q_i = 0\}$, $1 \leqslant i \leqslant m$, denote the essential hyperplanes of $B(S)$. In the example of Fig. 2, $m = 6$, since we only need consider essential hyperplanes which intersect with the interior of C. Suppose further that each (unit) normal, $\mathbf{w}_i$, is chosen to point towards the interior of S at any point in the interior (in H_i) of $B(S) \cap H_i$ (see Fig. 2). Since none of the H_i is inconsistent, $\mathbf{w}_i$ can be chosen thus without loss of generality.

We demonstrate the existence of a neighbourhood N containing $B(S)$ such that $S \cap N$ is realisable in N, then we appeal to Proposition 5. Let H_i' be the line $\mathbf{w}_i \cdot \mathbf{x} + q_i + \delta = 0$, $1 \leqslant i \leqslant m$, where $\delta > 0$, and consider the function $g \in F_2$ defined by

$$g(\mathbf{x}) = \sum_{i=1}^{m} f(\mathbf{w}_i \cdot \mathbf{w} + q_i) + \sum_{i=1}^{m} f(-\mathbf{w}_i \cdot \mathbf{x} - q_i - \delta) - m + 1.$$

Since no three of the lines H_i intersect we can choose δ to be sufficiently small so that, for $m \geqslant 2$, the open cells of $C \backslash \bigcup_{i=1}^{m} (H_i \cup H_i')$ are partitioned into three sets V_{-1}, V_0, and V_1 over which g takes the values -1, 0, and 1, respectively. This can be deduced in general by an inductive argument and is illustrated in Fig. 3. Concerning open polygons, P, in V_{-1} and V_1, we make the following claims, which follow from the hypotheses on S:

(i) If $P \in V_{-1}$ is contained in S (e.g., P_1 in Fig. 3), then $d(P, B(S)) > 0$.

(ii) If $P \in V_1$ is not contained in S (e.g., P_2 in Fig. 3), then $d(P, B(S)) > \delta > 0$.

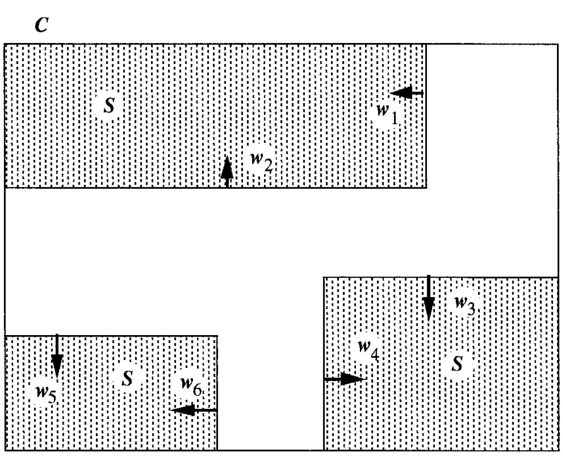

FIG. 2. Realisable set used to illustrate the proof of Theorem 4.

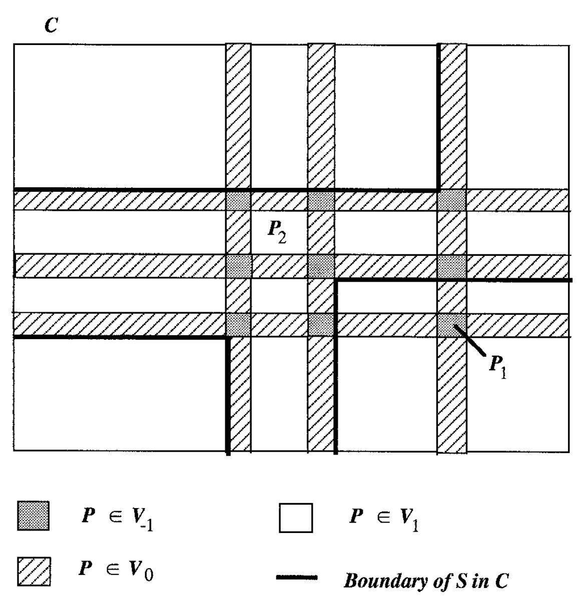

FIG. 3. Partitioning $C \backslash \{\bigcup_{i=1}^{m} (H_i \cup H_i')\}$ according to the value of g in proof of Theorem 4.

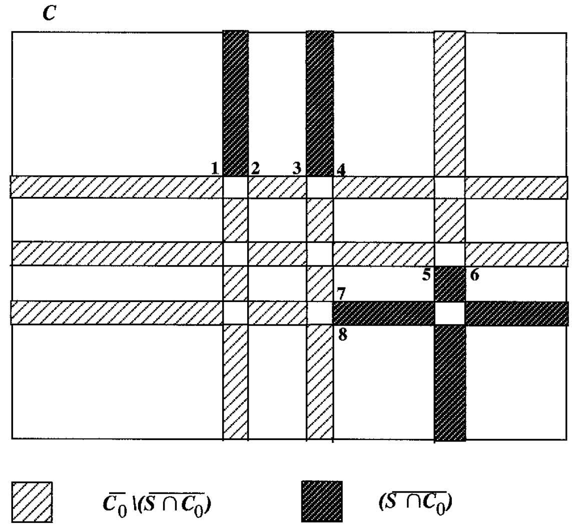

FIG. 4. The sets $\overline{(S \cap C_0)}$ and $\overline{C_0}\backslash\overline{(S \cap C_0)}$ (see proof of Theorem 4). The boundary of $\overline{(S \cap C_0)}$ in $\overline{C_0}$ consists of the points 1–8.

Hence, there exists an open set N containing $B(S)$ such that N does not intersect with any of those sets P to which claims (i) and (ii) apply.

We now consider the set

$$C_0 = \bigcup \{P \mid P \in V_0\}.$$

We claim that $\overline{(S \cap C_0)}$ is realisable in $\overline{C_0}$, the closure of C_0. Now the boundary of $\overline{(S \cap C_0)}$ in $\overline{C_0}$ consists of a finite set of points $\{z_j \mid 1 \leqslant j \leqslant r\}$, where $r = 8$ in the case of our example (see Fig. 4). Now consider a neighbourhood of this boundary of the form

$$A = \bigcup_{j=1}^{r} N(\mathbf{z}_j, \varepsilon),$$

where $\varepsilon \ll \delta$ selected above. For each z_j we select a line H_j defined by $\mathbf{v}_j \cdot \mathbf{x} + s_j = 0$, which contains $\mathbf{z}_j$ and whose

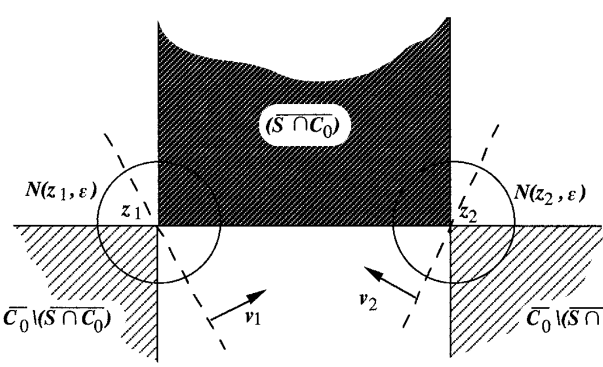

FIG. 5. Separating $\overline{(S \cap C_0)}$ from $\overline{C_0}\backslash\overline{(S \cap C_0)}$ on a neighbourhood of the boundary (see proof of Theorem 4).

normal, $\mathbf{v}_j$, points into the interior of $S \cap C_0$ at $\mathbf{z}_j$. This is illustrated for the particular case of points 1 and 2 in Fig. 5. Assume further that H_j does not intersect any of the neighbourhoods $N(\mathbf{z}_i, \varepsilon)$, $i \neq j$. The last condition can always be achieved after reducing ε if necessary. Consider the function $u \in F_2$ defined by

$$u(\mathbf{x}) = \sum_{j=1}^{r} 2f(\mathbf{v}_j \cdot \mathbf{x} + s_j).$$

It is clear that on each neighbourhood $N(\mathbf{z}_j, \varepsilon)$, $u(\mathbf{x}) = \alpha_j$, if $\mathbf{x} \notin S$, and $u(\mathbf{x}) = \alpha_j + 2$, if $\mathbf{x} \in S$, for integers $\alpha_1, ..., \alpha_r$. By Lemma 6, we can choose $t \in F_2$, such that $t(\mathbf{x}) = \alpha_j$ for all $\mathbf{x} \in N(\mathbf{z}_j, \varepsilon)$ (reducing ε again, if necessary). Let $s(\mathbf{x}) = u(\mathbf{x}) - t(\mathbf{x}) - 1$. Then $s(\mathbf{x}) = 1$ for all $\mathbf{x} \in S \cap A$, and $s(\mathbf{x}) = -1$ for all $\mathbf{x} \in A \backslash S$. It follows from Proposition 1 that $\overline{(S \cap C_0)}$ is realisable in $\overline{C_0}$. There exists, therefore, some $h \in F_2$, such that $h(\mathbf{x}) > 0$ for almost all $\mathbf{x} \in S \cap C_0$, and $h(\mathbf{x}) < 0$ for almost all $\mathbf{x} \in C_0 \backslash S$. We can assume without loss of generality that $|h(\mathbf{x})| < 1$ for all $\mathbf{x} \in C$. Then $h(\mathbf{x}) + g(\mathbf{x}) > 0$ for almost all $\mathbf{x} \in S \cap N$ and $h(\mathbf{x}) + g(\mathbf{x}) < 0$ for almost all $\mathbf{x} \in N \backslash S$, so that $S \cap N$ is realisable in N. It follows from Proposition 5 that S is realisable and the proof is complete. ▮

Remark. It is well known (e.g., [16]) that the set of possible decision regions for the three-layer architecture described in Section 1 consists of all finite unions of polyhedral sets in $\mathbb{R}^n$. Theorem 4 demonstrates that, for $n = 2$, the two-layer architecture is almost as versatile. If we restrict attention to bounded sets whose bounding hyperplanes lie in general position then the decision regions for which a three-layer architecture is necessary are precisely those whose boundaries contain an inconsistent line, as exemplified by Fig. 1.

However, while Theorem 4 enables a set S to be recognised as realisable from the geometry of its boundary, it says little about the number of first-layer nodes (equivalently, the number of terms in (2.1) for which $\mathbf{w}_i \neq \mathbf{0}$, $a_i \neq 0$) required in a two-layer network realising S. If $B(S)$ has k essential hyperplanes, then (see, e.g., [10]) k represents a lower bound for this number. Previous research (e.g, [12]) has identified some general classes of realisable sets for which this lower bound on network size can be attained. In general, however, the number of first-layer nodes required may be strictly greater than k. For example, Fig. 6a (see [10]) depicts a set S which is realisable by Theorem 4, but which cannot be realised by a two-layer network whose first-layer nodes represent only eight essential hyperplanes of $B(S)$. One further node is required whose weight vector $\mathbf{w}$ and bias q define, for example, the line L in Fig 6a. Increasing the ratio d_1/d_2 (Fig. 6b) results in a set for which three additional nodes (corresponding, for example, to L_1, L_2, and L_3 in the figure) are required in any network realising it. Moreover, it is proved in [17], that as $d_1/d_2 \to \infty$ in Fig. 6b, so does the minimum number of first-layer nodes in two-layer network realising S. Thus, the number of first-layer nodes required in a two-layer network to realise a set S may be arbitrarily greater than the number of essential hyperplanes of $B(S)$. In contrast, regardless of the value of d_1/d_2 in Fig. 6b, S can be realised exactly as the decision region of a three-layer network with eight first-layer and two second-layer nodes, and a single output node. Hence, although two- and three-layer networks offer an exact solution to this classification problem, the three-layer solution in general will be more efficient in terms of the total number of units required.

The following examples demonstrate that Theorem 4 does not hold if S is unbounded, or if we dispense with the condition concerning the intersection of essential lines.

EXAMPLE 7. The unbounded region

$$S = \{(x, y) \mid x, y \geqslant 1\} \cup \{(x, y) \mid x, y \leqslant -1\},$$

illustrated in Fig. 7 is not realisable in $\mathbb{R}^2$.

Proof. Suppose that S were realisable by a function

$$g(\mathbf{x}) = \sum_{i=1}^{m} a_i f(\mathbf{w}_i \cdot \mathbf{x} + q_i), \quad m \in \mathbb{N},$$
$$w_i \in \mathbb{R}^n, \quad a_i, q_i \in \mathbb{R}, \quad 1 \leqslant i \leqslant m,$$

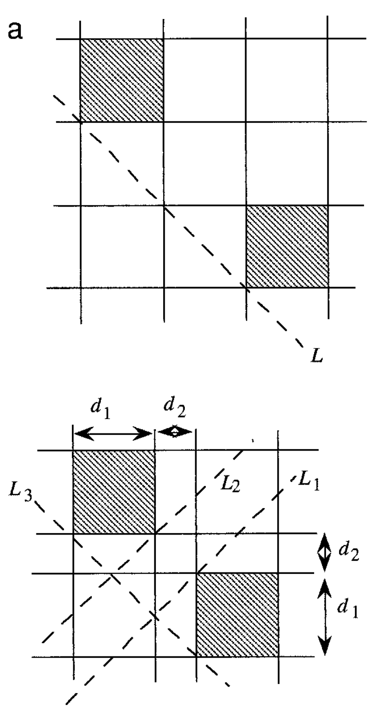

FIG. 6. Examples of realisable decision regions where number of first-layer units strictly exceeds number of essential hyperplanes.

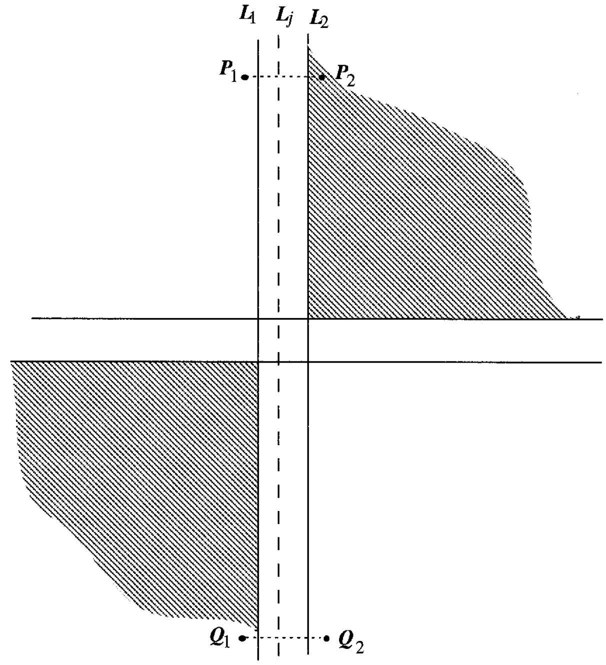

FIG. 7. Example of unbounded region which is unrealisable.

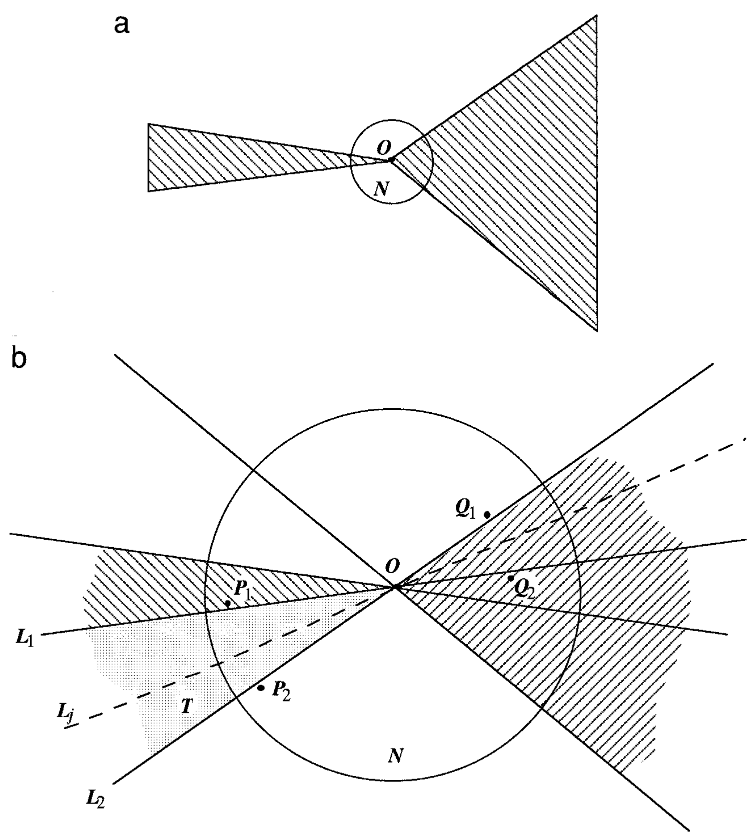

FIG. 8. Example of region which does not satisfy intersection condition and which is unrealisable.

where, without loss of generality, L_1 and L_2 are defined by $\mathbf{w}_1 \cdot \mathbf{x} + q_1 = 0$ and $\mathbf{w}_2 \cdot \mathbf{x} + q_2 = 0$. Consider the coefficients a_1, a_2 and those coefficients, a_j, for which $\mathbf{w}_j \cdot \mathbf{x} + q_j = 0$ defines a line, L_j, parallel to L_1 and L_2 and located in the slab which they bound (see Fig. 7). We can assume that $\mathbf{w}_1 = \mathbf{w}_2 = \mathbf{w}_j$. Consider the value of g at the points P_1, P_2, Q_1, and Q_2, where these points are chosen to lie sufficiently far from the origin that the line segments $[P_1, P_2]$ and $[Q_1, Q_2]$ do not intersect with any lines $\mathbf{w}_i \cdot \mathbf{x} + q_i = 0$ appearing in the definition of g, apart from L_1, L_2, and the L_j identified above. Comparing the values of g at P_1 and P_2, we obtain

$$a_1 + a_2 + \sum_j a_j = g(P_1) - g(P_2) > 0.$$

For points Q_1 and Q_2 we have

$$a_1 + a_2 + \sum_j a_j = g(Q_1) - g(Q_2) < 0.$$

This is a contradiction and it follows that the set S cannot be realisable.

EXAMPLE 8. The bounded region S in Fig. 8a is not realisable in $\mathbb{R}^2$.

Proof. The proof is analogous to that for Example 7. Again we assume S can be realised by a function

$$g(x) = \sum_{i=1}^{m} a_i f(w_i \cdot x + q_i), \quad m \in \mathbb{N},$$

$$w_i \in \mathbb{R}^n, \quad a_i, q_i \in \mathbb{R}, \quad 1 \leqslant i \leqslant m.$$

However, in this case we focus attention on a neighbourhood of $\mathbf{O}$, which is sufficiently small that it is intersected only by those lines $\mathbf{w}_i \cdot \mathbf{x} + q_i = 0$ which contain $\mathbf{O}$. We consider the coefficients a_1 and a_2 associated with the lines L_1 and L_2 in the definition of g and those coefficients a_j associated with lines L_j which pass through $\mathbf{O}$ and lie in the space T bounded by L_1 and L_2 as shown in Fig. 8b. Once again we can identify points P_1, P_2, Q_1, and Q_2 such that

$$a_1 + a_2 + \sum_j a_j = g(P_1) - g(P_2) > 0$$

and

$$a_1 + a_2 + \sum_j a_j = g(Q_1) - g(Q_2) < 0.$$

This contradiction shows that S is unrealisable. ∎

Example 8 is a case of the "twisted-bow tie" condition which Zweitering [14, Chap. 5] has shown to characterise a class of unrealisable regions.

3. DISCUSSION

The results presented in this paper contribute to our understanding of the functions which are exactly realisable by a two-layer network of McCulloch–Pitts neurones. In particular, Theorem 4 provides a simple geometrical characterization of all the bounded, realisable subsets of $\mathbb{R}^2$, apart from those nongeneric examples which violate the condition regarding the intersection of essential lines. It shows that the ability of two-layer nets to create bounded decision regions is virtually the same as that of the three-layer architecture when inputs are two-dimensional. However, it should be noted that the difference in complexity between two- and three-layer structures realising the same classification may be arbitrarily large, with the two-layer structure requiring many more nodes. The results also indicate how the boundary lines of an unrealisable, bounded union of polyhedra might be deformed to produce a realisable approximation to it. Any such deformation which removes inconsistent lines and

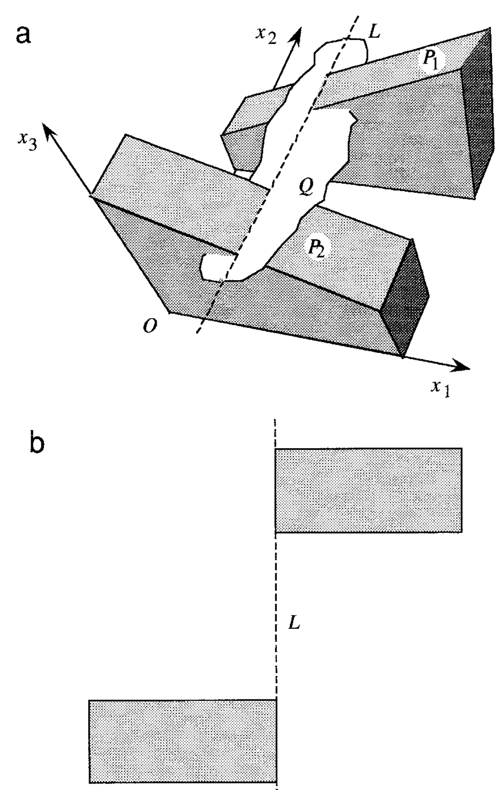

FIG. 9. (a) Example of three-dimensional region which is unrealisable. (b) Intersection of region in Fig. 9(a) with plane Q to form two-dimensional unrealisable region.

"triplet-wise" intersections of lines suffices. Hence the results can be of value in identifying a realisable approximation to a specified classification function.

There are a number of avenues for further research in the area. There is clearly a need to investigate possible generalizations of the results to higher dimensions. However, initial investigations suggest that the problem of obtaining a valid generalization of Theorem 4 to higher dimensions is far from straightforward. For example, suppose we restate the condition concerning the intersection of essential hyperplanes in n dimensions as "no $n+1$ essential hyperplanes have a nonempty intersection," which reduces to the condition of Theorem 4 when $n=2$. Consider now the set $S \subset \mathbb{R}^3$, defined by

$$S = \{(x_1, x_2, x_3) \mid 0 \leqslant x_1 \leqslant 1, 0 \leqslant x_2 \leqslant \tfrac{1}{4}, 0 \leqslant x_3 \leqslant 1, x_1 + 2x_3 \leqslant 2\} \cup \{(x_1, x_2, x_3) \mid 0 \leqslant x_1 \leqslant 1, \tfrac{3}{4} \leqslant x_2 \leqslant 1, 0 \leqslant x_3 \leqslant 1, 2x_3 - x_1 \leqslant 1\}$$

formed from the union of two polyhedral sets in R^3 as illustrated in Fig. 9a. Clearly this set satisfies the new condition, it is bounded, and none of its essential hyperplanes is inconsistent. However, it is not realisable. To show this, we assume first that S is realisable so that there exists a function $g \in F_3$ defined by

$$g(\mathbf{x}) = \sum_{i=1}^{m} a_i f(\mathbf{w}_i \cdot \mathbf{x} + q_i)$$

such that $g(\mathbf{x}) > 0$ for almost all $\mathbf{x} \in S$ and $g(\mathbf{x}) < 0$ for almost all $\mathbf{x} \in \mathbb{R}^3 \backslash S$. Let L be the line of intersection of the essential planes P_1: $2x_3 - x_1 = 1$ and P_2: $x_1 + 2x_3 = 2$ and select a plane, Q, which contains L (see Fig. 9a), so that the intersection $S \cap Q$, considered as a subset of $\mathbb{R}^2$, has the form of Fig. 9b. Now there are an infinitely many Q which satisfy these conditions and we can ensure, therefore, that Q is distinct from each plane H_i: $\mathbf{w}_i \cdot \mathbf{x} + q_i = 0$, $1 \leqslant i \leqslant m$. Furthermore, since each connected open cell of $\mathbb{R}^3 \backslash \bigcup_{i=1}^m H_i$ is contained either in S or in $\mathbb{R}^3 \backslash S$, and g is constant over each cell, it follows that

$$(S \cap \{\mathbf{x} \in \mathbb{R}^3 \mid g(\mathbf{x}) < 0\}) \cup (\{\mathbb{R}^3 \backslash S\} \cap \{\mathbf{x} \in \mathbb{R}^3 \mid g(\mathbf{x}) > 0\}) \subseteq \bigcup_{i=1}^{m} H_i. \tag{3.1}$$

Since Q is distinct from each H_i, $1 \leqslant i \leqslant m$, $Q \cap \bigcup_{i=1}^m H_i$ consists of a union of lines. Hence, (3.1) implies that, with respect to two-dimensional Lebesgue measure on Q, $g(\mathbf{x}) > 0$ for almost all $\mathbf{x} \in S \cap Q$ and $g(\mathbf{x}) < 0$ for almost all $\mathbf{x} \in Q \backslash S$. Thus the set $S \cap Q$ (see Fig. 9b) must represent a realisable set in $\mathbb{R}^2$. This is a contradiction since L is an inconsistent line in $S \cap Q$ and it follows that S cannot be realisable in $\mathbb{R}^3$. Thus it may be that simple geometric characterisations of n-dimensional realisable sets cannot be so readily obtained as in the two-dimensional case.

Further questions to explore include that of extending the theory to provide a geometrical characterisation of realisable subsets of $\mathbb{R}^2$ which are unbounded, or which do not satisfy the condition that no three essential bounding lines intersect. In view of the reliance of the proof of Theorem 4 on both the boundedness of the set and the intersection condition, it seems likely that a different approach will be required in order to extend the theory to these cases.

In this paper we have considered sets which are "almost everywhere" realisable; i.e., we consider classification to be exact if the misclassified set is of zero measure. This is a reasonable definition in the context of classifying data generated from a probability density which is absolutely continuous with respect to Lebesgue measure. The question of whether the theory can be applied in the case where a stricter definition of realisability (i.e., one which demands that the misclassified set be empty) has been suggested by an anonymous referee. We claim that Theorem 4 remains true for this stronger definition of realisability if we restrict attention to sets S which are either open or closed in $\mathbb{R}^n$. This can be proved by adapting the construction in the proof of Theorem 4 to take account of the stronger definition and ensure that all points in $B(S)$ are correctly classified. The details are somewhat tedious and are not included in this paper. However, Theorem 4 is not true for the stricter definition without the condition that S is open or closed. It is a simple matter to construct a union of polyhedral sets which is neither open nor closed and which is realisable in sense of this paper but unrealisable when the stricter definition is used.

The question of constructing two-layer architectures which realise given sets has not been considered in this paper. While some progress in this area has been made (e.g., [14, 17, 18]) no general method exists to the knowledge of the author. Such techniques would be of value in the design of neural networks to realise known functions in cases where inefficient adaptive learning algorithms are currently applied.

ACKNOWLEDGMENTS

Financial support for this research was provided by the Scottish Office Agriculture, Environment and Fisheries Department and the Science and Engineering Research Council. The author is grateful to two anonymous referees for their valuable comments on an earlier version of this paper.

REFERENCES

1. W. S. McCulloch and W. Pitts, A logical calculus of the ideas imminent in nervous activity, *Bull. Math. Biophys.* **5** (1943), 115–133.
2. B. Widrow, R. G. Winter, and R. A. Baxter, Layered neural nets for pattern recognition, *IEEE Trans. Acoust. Speech Signal Proc.* **36** (1988), 1109–1118.
3. K. Hornik, M. Stinchcombe, and H. White, Multilayer feed-forward networks are universal approximators, *Neural Networks* **2** (1989), 359–366.
4. G. Cybenko, Approximations by superpositions of a sigmoidal function, *Math. Control Signals Systems* **2** (1989), 303–314.
5. K.-I. Funahashi, On the approximate realisation of continuous mappings by neural networks, *Neural Networks* **2** (1989), 183–192.
6. L. K. Li, On computing decision regions with neural nets, *J. Comput. System Sci.* **43** (1991), 509–512.
7. S.-C. Huang and Y.-F. Huang, Bounds on the number of hidden neurones in multilayer perceptrons, *IEEE Trans. Neural Networks* **2** (1991), 47–55.
8. E. D. Sontag, Feedforward nets for interpolation and classification, *J. Comput. System Sci.* **45** (1992), 20–48.
9. M. Budinich and E. Milotti, Properties of feedforward neural networks, *J. Phys. A: Math. Gen.* **25** (1992), 1903–1914.
10. G. J. Gibson and C. F. N. Cowan, On the decision regions of multilayer perceptrons, *Proc. IEEE* **78** (1990), 1590–1594.
11. J. Makhoul, R. Schwartz, and A. El-Jaroudi, Classification capabilities of two-layer neural nets, *in* "Proc. IEEE Int. Conf. ASSP, Glasgow, May 1989," pp. 653–638.
12. R. Shonkwiler, Separating the vertices of N-cubes by hyperplanes and its application to artificial neural networks, *IEEE Trans. Neural Networks* **4** (1993), 343–347.
13. G. J. Gibson, A combinatorial approach to understanding perceptron decision regions, *IEEE Trans. Neural Networks* **4** (1993), 989–992.
14. P. J. Zweitering, "The Complexity of Multi-layered Perceptrons," Ph.D. thesis, Technische Universiteit Eindhoven, 1994.
15. E. K. Blum and L. K. Li, Approximation theory and feedforward networks, *Neural Networks* **4** (1991), 511–515.
16. R. Lippmann, Computing with neural nets, *IEEE Acoust. Speech Signal Proc.* (1987).
17. G. J. Gibson, Constructing functions using multilayer perceptrons—Towards a theory of complexity, *J. Systems Eng.* **2** (1992), 263–271.
18. H. Takahashi, E. Tomita, and T. Kawabata, Separability of internal representations in multilayer perceptrons with application to learning, *Neural Networks* **6** (1993), 689–703.