Annals of Mathematics and Artificial Intelligence 24 (1998) 115–128

Multilayer neural networks and polyhedral dichotomies

Claire Kenyon [a] and Hélène Paugam-Moisy [b]

[a] *LRI, CNRS URA 410, Université Paris-Sud, F-91405 Orsay Cedex, France*
E-mail: Claire.Kenyon@lri.fr
[b] *LIP, CNRS URA 1398, ENS Lyon, 46 allée d'Italie, F-69364 Lyon Cedex 07, France*
E-mail: Helene.Paugam-Moisy@ens-lyon.fr

We study the number of hidden layers required by a multilayer neural network with threshold units to compute a dichotomy from $\mathbb{R}^d$ to $\{0, 1\}$, defined by a finite set of hyperplanes. We show that this question is far more intricate than computing Boolean functions, although this well-known problem is underlying our research. We present advanced results on the characterization of dichotomies, from $\mathbb{R}^2$ to $\{0, 1\}$, which require two hidden layers to be exactly realized.

1. Introduction

The number of hidden layers is a crucial parameter for the architecture of multilayer neural networks. Early research, in the 60's, addressed the problem of exactly realizing Boolean functions with binary networks or binary multilayer networks. On the one hand, more recent work focused on approximately realizing real functions with multilayer neural networks with one hidden layer [7,8,14] or with two hidden layers [2]. On the other hand, some authors [1,15] were interested in finding bounds on the architecture of multilayer networks for exact realization of a finite set of points. Another approach is to search the minimal architecture of multilayer networks for exactly realizing real functions, from $\mathbb{R}^d$ to $\{0, 1\}$. Our work, of the latter kind, is a continuation of the effort of [5,6,9,10] towards characterizing the dichotomies which can be exactly realized with a single hidden layer neural network composed of threshold units. In this article, the research is focused on 2-cycles, in $\mathbb{R}^2$ and, extending [4], we show how this subject is related to linear programming and combinatorial optimization. We prove two results on local realizability of a polyhedral dichotomy in $\mathbb{R}^2$, by applying Farkas' lemma.

First we define the notion of *polyhedral dichotomy* and precise which neural networks we consider. We emphasized the link with Boolean functions and we present several points of view of the problem, our approach being geometric.

A finite set of hyperplanes $\{H_i\}_{1\leqslant i\leqslant h}$ defines a partition of the d-dimensional space into convex polyhedral open regions, the union of the H_i's being neglected as a subset of measure zero. A *polyhedral dichotomy* is a function $f : \mathbb{R}^d \to \{0, 1\}$, obtained by associating a class, equal to 0 or to 1, to each of those regions. Thus both

© J.C. Baltzer AG, Science Publishers

$f^{-1}(0)$ and $f^{-1}(1)$ are unions of a finite number of convex polyhedral open regions. The h hyperplanes whose removal would result in merging at least one pair of regions in different classes are called the *essential hyperplanes* of f. A point P is an *essential point* if it is the intersection of some set of essential hyperplanes.

All *multilayer networks* are supposed to be feedforward neural networks of threshold units, fully interconnected from one layer to the next, without skipping interconnections. A network is said to *realize* a function $f: \mathbb{R}^d \to \{0,1\}$ if, for an input vector x, the network output is equal to $f(x)$, almost everywhere in $\mathbb{R}^d$. The functions realized by our multilayer networks are the polyhedral dichotomies.

2. Polyhedral dichotomies and Boolean functions

By definition of threshold units, each unit of the first hidden layer computes a binary function y_j of the real inputs $(x_1, \ldots, x_d)$. For all j, the jth unit of the first hidden layer can be seen as separating the space by the hyperplane H_j: $\sum_{i=1}^{d} w_{ij} x_i = \theta_j$. Hence the first hidden layer necessarily contains at least one hidden unit for each essential hyperplane of f. Thus each region can be labelled by a binary number $y = y_h, \ldots, y_1$ (see figure 1). Afterwards, subsequent layers compute a Boolean function of $\{0,1\}^h$ and must associate the right class to each region. Since any Boolean function can be written in DNF-form, two hidden layers are sufficient for a multilayer network to realize any polyhedral dichotomy. The network of figure 1 computes $\overline{y_1} y_2 + y_1 \overline{y_2}$.

Usually there are fewer than 2^h regions and not all possible labels actually exist. More precisely, the number of convex regions defined by h hyperplanes in a d-dimensional space is bounded by 2^h only if $h \leqslant d$ and by $\sum_{i=0}^{d} \mathrm{C}_h^i$ if $h > d$.

Definition 1. The *Boolean family* $\mathcal{B}_f$ of a polyhedral dichotomy f is defined to be the set of all Boolean functions on h variables which are equal to f on all existing labels.

A one-hidden-layer network realizes the dichotomy iff a linear system of inequalities is solvable, the unknown variables being the weights w_i and the threshold θ of

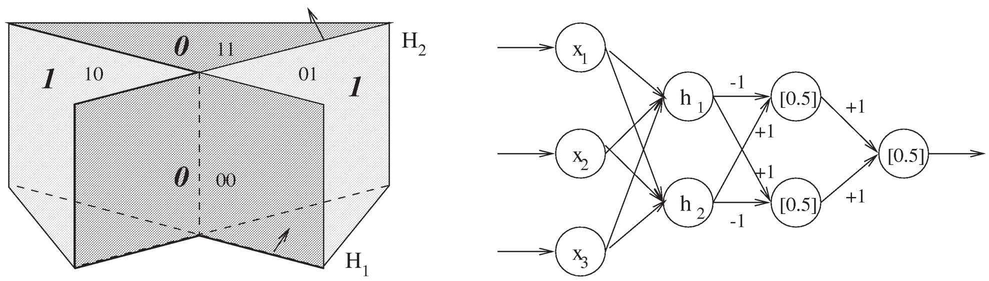

Figure 1. The "four-quadrant" dichotomy, which generalizes the XOR function, in dimension $d = 3$, and a two-hidden-layer network realizing this dichotomy. The connections are labelled by their weights and the thresholds are bracketed inside the units.

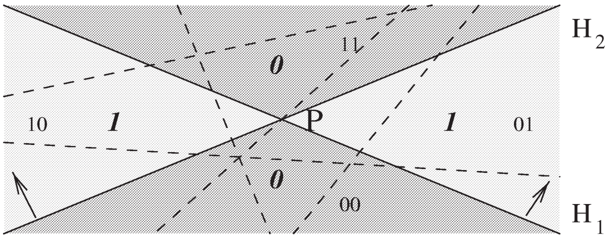

Figure 2. No one-hidden-layer network can realize the "four-quadrant" dichotomy (here $d = 2$). Adding redundant hyperplanes (whatever their number k) does not help solving the system.

the output unit. For instance, the system corresponding to the example of figure 1 is as follows and cannot be solved:

$$\begin{cases} 0 < \theta, & \text{region 00, in class 0,} \\ w_1 > \theta, & \text{region 10, in class 1,} \\ w_2 > \theta, & \text{region 01, in class 1,} \\ w_1 + w_2 < \theta, & \text{region 11, in class 0.} \end{cases} \tag{1}$$

Adding k hidden units on the first hidden layer corresponds to adding k *redundant hyperplanes* and could allow to find a solution to the system of inequalities since the dimension of the internal representation (number of input variables of the Boolean function) is increased from h to $h + k$. However, for realizing the "four-quadrant" dichotomy, it can be proved that two hidden layers are necessary (see [5]). Figure 2 shows that, whatever the number and the position of redundant hyperplanes we add, there still exist four regions near the essential point P which create an inconsistency.

Coming back to the notion of Boolean families, it is straightforward that all polyhedral dichotomies which have at least one linearly separable function in their Boolean family can be realized by a one-hidden-layer network. However, the converse is false. A counter-example was produced in [6]: adding redundant hyperplanes (i.e., extra units on the first hidden layer) can eliminate the need for a second hidden layer (see figure 3).

Example 2. The linear system ($\mathcal{S}$) associated to the dichotomy f defined by figure 3 is composed of 16 inequalities as follows:

$$(\mathcal{S}) \quad \left\{ \begin{array}{rl} 0 > \theta & (1) \\ w_1 < \theta & \\ w_1 + w_2 < \theta & (3) \\ w_1 + w_2 + w_3 < \theta & \\ w_4 < \theta & \\ w_1 + w_4 < \theta & \\ w_1 + w_2 + w_4 < \theta & \\ w_1 + w_2 + w_3 + w_4 < \theta & \end{array} \right. \quad \left\{ \begin{array}{rl} w_4 + w_5 < \theta & (4) \\ w_1 + w_4 + w_5 < \theta & \\ w_1 + w_2 + w_4 + w_5 > \theta & (2) \\ w_1 + w_2 + w_3 + w_4 + w_5 < \theta & \\ w_4 + w_5 + w_6 < \theta & \\ w_1 + w_4 + w_5 + w_6 < \theta & \\ w_1 + w_2 + w_4 + w_5 + w_6 < \theta & \\ w_1 + w_2 + w_3 + w_4 + w_5 + w_6 < \theta & \end{array} \right.$$

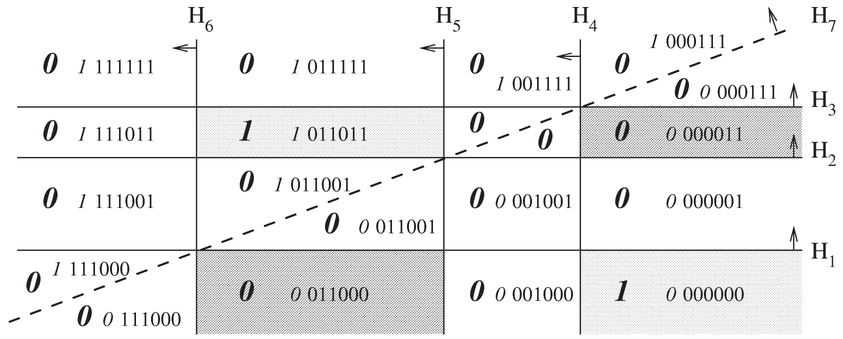

Figure 3. A one-hidden-layer network, with 6 hidden units on the first layer (corresponding to the essential hyperplanes), cannot realize the dichotomy, but a network with a 7th extra unit (associated to the redundant hyperplane H_7, in dotted line) can.

System ($\mathcal{S}$) cannot be solved because it contains an inconsistency created by the four grey regions (figure 3). Adding the four numbered inequalities by pairs implies that $w_1 + w_2 + w_4 + w_5$ should be both greater and less than 2θ.

Adding the redundant hyperplane H_7 corresponds to adding a 7th Boolean variable, as written in italic on the left of each region label (figure 3). Each of the four regions along H_7 are split into two regions, one with a 0 digit, on the negative side of H_7, and one with a 1 digit, on the positive side. A new linear system ($\mathcal{S}'$) can be defined for these 7 variables. It contains 20 inequalities, each of them including w_7 or not, according to the side of H_7 the region stands. The inconsistency appears no longer, since w_7 is added to only one of the four numbered inequalities. Moreover, the system can be solved using Maple. The simplex method for minimizing gives the following solution:

$$w_1 = -4, \quad w_2 = 2, \quad w_3 = -4, \quad w_4 = -4, \quad w_5 = 2, \quad w_6 = -4, \quad w_7 = 4, \quad \theta = -1.$$

Hence the problem of finding a minimal architecture for realizing dichotomies is not equivalent to the neural computation of Boolean functions. Finding a characterization of all the polyhedral dichotomies which can be realized exactly by a one-hidden-layer network is a challenging problem.

3. Geometrical approach

3.1. Three geometric configurations of XOR

Our approach consists in finding geometric configurations which imply that a function is not realizable with a single hidden layer. There are three known such geometric configurations which involve two pairs of regions: the XOR-situation, the XOR-bow-tie and the XOR-at-infinity, as summarized in figure 4. We propose below to give precise and unified definitions for all these geometric situations.

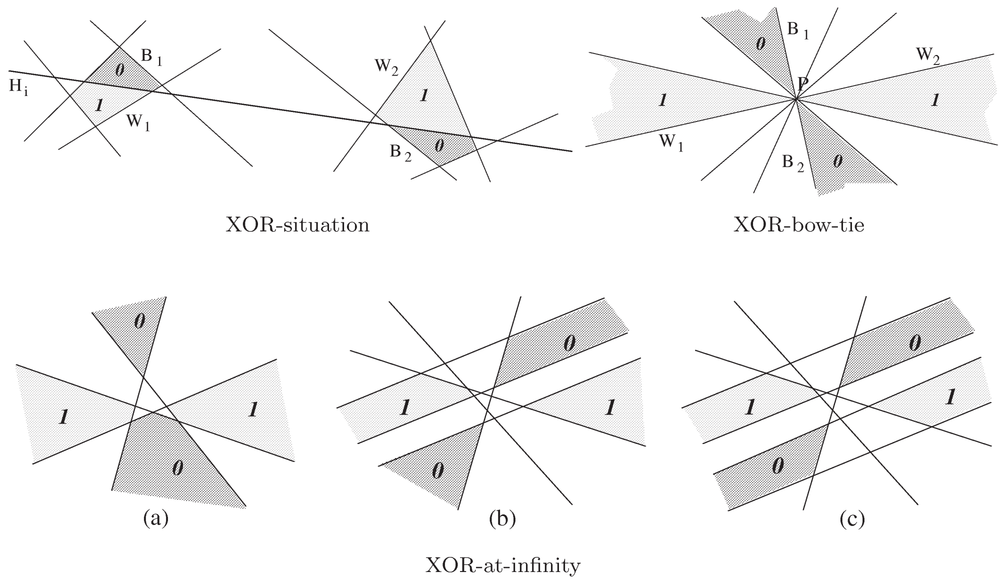

Figure 4. Geometrical representation of XOR-situation, XOR-bow-tie and XOR-at-infinity in the plane (dark regions are in class 0, light regions are in class 1).

Definition 3. A polyhedral dichotomy is *in XOR-situation* if one of its essential hyperplanes H_i is *inconsistent*, i.e., there exist four distinct regions such that:

- B_1 and B_2 are in class 0, whereas W_1 and W_2 are in class 1,
- B_1 and W_1 are adjacent along H_i, somewhere, and B_2 and W_2 are adjacent along H_i, elsewhere, but B_1 and W_2 are on the same side of H_i.

The two next definitions require the notion of *opposite regions, with regard to a mask*, defined as follows:

Definition 4. Let f be a dichotomy defined by h essential hyperplanes. Let M_k be a *mask*, i.e., a partition of the h digits into k "visible" digits and $h - k$ "hidden" digits. Two regions are said to be M_k*-opposite* iff their labels are opposite on the visible digits of M_k and are identical on the digits hidden by M_k.

Example 5. If $h = 5$, $k = 3$ and M_3 is the multiplicative mask 01101 (first and fourth digits hidden, second, third and fifth digits visible), two regions labelled by 10001 and 11100 are M_3-opposite, since their visible parts ⋄00⋄1 and ⋄11⋄0 are opposite and their hidden parts 1⋄⋄0⋄ and 1⋄⋄0⋄ are identical.

Definition 6. If there exist a non-empty mask M_k and four distinct regions B_1, B_2, W_1 and W_2, with a common point P in their closure and such that:

- B_1 and B_2 are in class 0, whereas W_1 and W_2 are in class 1,

- B_1 and B_2 are M_k-opposite, and W_1 and W_2 are M_k-opposite,

then the polyhedral dichotomy is *in XOR-bow-tie* around the essential point P.

In this case, M_k can be considered as masking all the hyperplanes which do not contain point P. The two regions in class 0 create a bow-tie. The two regions in class 1 create another bow-tie which crosses the previous one at point P.

Definition 7. A polyhedral dichotomy is *in XOR-at-infinity* if there exist a non-empty mask M_k and four distinct regions B_1, B_2, W_1 and W_2, which are unbounded and such that:

- B_1 and B_2 are in class 0, whereas W_1 and W_2 are in class 1,
- B_1 and B_2 are M_k-opposite, and W_1 and W_2 are M_k-opposite.

There are several slightly different possible outlines of this geometrical situation, as represented on cases (a), (b) and (c) of figure 4. In the case of unbounded regions, the mask is transparent, i.e., $M_h = 111\ldots11$, except if some of their border are parallel essential hyperplanes. Even in this case, when considering the compacted space $\mathbb{R}^d \cup \{\infty\}$, a XOR-at-infinity is no more than a XOR-bow-tie with the point ∞ as common point P.

Theorem 8. If a polyhedral dichotomy f, from $\mathbb{R}^d$ to $\{0, 1\}$, is in XOR-situation, or in XOR-bow-tie, or in XOR-at-infinity, then f cannot be realized by a one-hidden-layer network.

The proof can be found in [6,11] for the XOR-situation, in [18] for the XOR-bow-tie, and in [6] for the XOR-at-infinity. The sketch of these proofs is always the same: the four regions (two in each class) and their respective labellings induce an inconsistency in the system of inequalities associated to a one-hidden-layer solution (see equation (1)) which cannot be solved by adding redundant hyperplanes, whatever their number and position.

3.2. Critical cycle

Figure 5 shows an example of another geometric configuration which prevents a dichotomy from being realizable by a one-hidden-layer network, the *critical cycle* which has been exhibited first in [4]. We give below a detailed definition.

We first need some preliminary definitions of *critical pairs* of regions.

Definition 9. Two regions whose closures both contain a point P are called *opposite with respect to* P iff they are M_k-opposite, where M_k is the mask which hides exactly the digits of the hyperplanes not containing P.

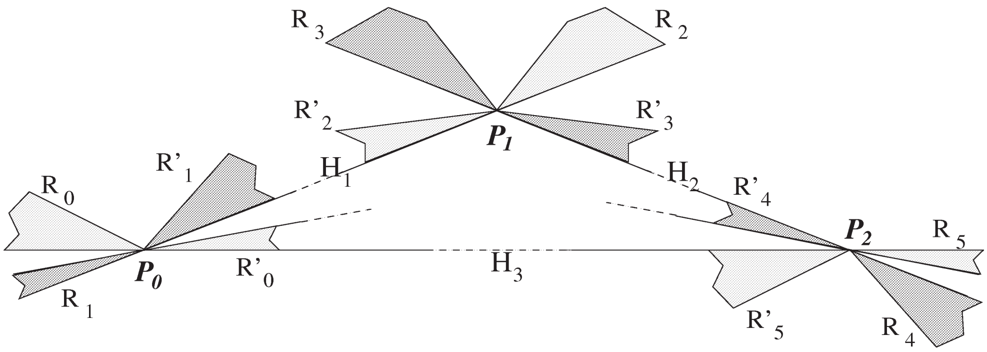

Figure 5. Geometrical configuration of a critical cycle, in the plane. N.B. One can augment the figure in such a way that there is no XOR-situation, no XOR-bow-tie, and no XOR-at-infinity.

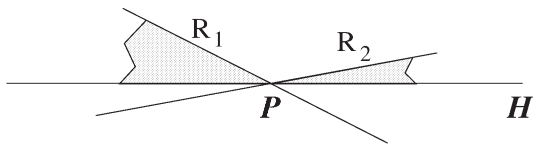

Figure 6. A critical pair of regions, with respect to a hyperplane H and a point P.

Definition 10. Let $\{R_1, R_2\}$ be a pair of regions, in the same class, and whose closures both contain an essential point P. If there is an essential hyperplane H going through P, such that R_2 is adjacent along H to the region opposite to R_1, with respect to P (see figure 6), then $\{R_1, R_2\}$ is called a *critical pair with respect to* P *and* H. Note that R_1 and R_2 are both on the same side of H.

We define a graph G whose nodes correspond to the critical pairs of essential regions of f and whose edges are colored green or red. There is a *red edge* between $\{B_1, B_2\}$ and $\{W_1, W_2\}$ if the pairs, in different classes, are both critical with respect to the same point but to different hyperplanes (see figure 7). Note that this strongly resembles a XOR-bow-tie, except that B_1 and B_2 are not quite M_k-opposite, and W_1 and W_2 are not quite M_k-opposite. If two pairs are both critical with respect to the same hyperplane H but with respect to different points, there are linked by a *green edge* either if the two pairs $\{B_1, B_2\}$ and $\{W_1, W_2\}$ are in different classes and on the same side of H, or if the two pairs $\{B_1, B_2\}$ and $\{B_3, B_4\}$ are in the same class but on different sides of H (see figure 8).

Definition 11. A *critical cycle* is a geometric configuration associated to a graph G which contains a cycle, with alternating colors, as depicted in figures 5 and 9.

Theorem 12. If a polyhedral dichotomy f, from $\mathbb{R}^d$ to $\{0, 1\}$, has a critical cycle, then f cannot be realized by a one-hidden-layer network.

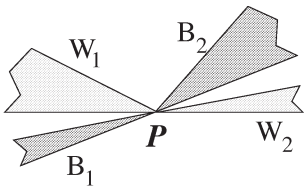

Figure 7. Two critical pairs linked by a red edge in graph G.

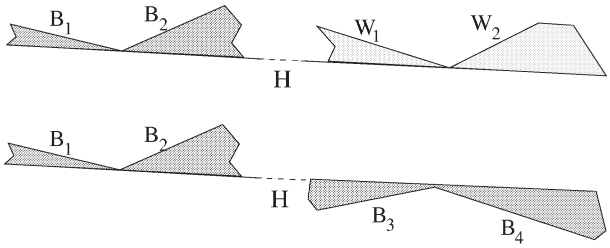

Figure 8. Two different cases of critical pairs linked by a green edge in graph G.

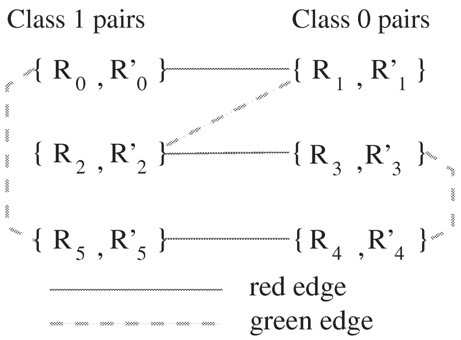

Figure 9. Graph associated to the critical cycle proposed as example in figure 5.

Note that all the figures are in $\mathbb{R}^2$, but the definition is general, in $\mathbb{R}^d$. The proof of the theorem relies on the bicolor graph G and can be found in [4]. The underlying idea is always the same: no set of redundant hyperplanes can help solving the inconsistency.

4. Different points of view

4.1. Network approach

In contrast with our geometrical definitions, Takahashi et al. [17] have presented a notion of *cyclicity*, in the same context of research, but with a different point of view. They start from the notion of *summability* of Boolean functions [16], and "n-cyclicity" can be viewed as a reinterpretation of n-summability. Given the hyperplanes associated to the hidden units of a *fixed* network (essential hyperplanes, plus redundant hyperplanes), finding the weights which realize the dichotomy amounts to solving a system of linear inequalities. The authors of [17] claim that this system has a solution iff there is no cyclicity, but their notion of cyclicity is only defined with respect to a fixed network.

On the one hand, if a dichotomy f is in any of the three cases of XOR, then "cyclicity" occurs and remains, no matter what hidden units are added on the first hidden layer. On the other hand, for the example of figure 3, "cyclicity" occurs with six hidden units but not with seven hidden units. Hence, some cyclicities, under the definition of [17], can be realized and others cannot. Our approach is different since we want a characterization of the polyhedral dichotomies f which can be realized by a one-hidden-layer network, independently of the network realizing f. The problem can be addressed in a different way, even less geometric than [17], as presented below.

4.2. Topological approach

Another research direction, implying a function is realizable by a single hidden layer network, is based on a topological approach. The proof uses the universal approximator property of one-hidden-layer networks [7,8,14] applied to intermediate functions obtained constructively by adding extra hyperplanes to the essential hyperplanes of f. This direction was explored by Gibson [10], for a two dimensional input space. Gibson's result can be reformulated as follows:

Theorem 13. If a polyhedral dichotomy f is defined on a compact subset of $\mathbb{R}^2$, if f is not in XOR-situation, and if no three essential hyperplanes (lines) intersect, then f is realizable with a single hidden layer network.

Unfortunately Gibson's proof is not constructive, and extending it to remove some of the assumptions seems challenging. Both XOR-bow-tie and XOR-at-infinity are excluded by his assumptions of compactness and of no multiple intersections. In the next sections, we explore the cases, in $\mathbb{R}^2$, which are excluded by Gibson's assumptions. Recent research by Hassell Sweatman et al. [13] is turned towards extending the definitions and proofs to go to higher dimensions, where new cases of inconsistency emerge in subspaces of intermediate dimension.

5. Advanced results, for 2-cycles, in $\mathbb{R}^2$

From now on, the research is focused on the 2-cycles, i.e., the sets of four regions, two of each class, which induce an inconsistency in the system of inequalities. Our aim is to extend Gibson's theorem and to state converse results to theorem 8, at least in $\mathbb{R}^2$. We need to introduce a notion of *local realizability*. The two theorems of this section prove that, in $\mathbb{R}^2$, the XOR-bow-tie and the XOR-at-infinity are the only configurations barring local realizability. Their proofs, based on the use of Farkas' lemma, are detailed below.

Definition 14. A polyhedral dichotomy f is *locally realizable* around a point $P \in \mathbb{R}^2 \cup \{\infty\}$ if there exists a neighborhood V of P such that the restriction of f to V is realizable by a one-hidden-layer network.

5.1. Local realization and XOR-bow-tie

Theorem 15. A polyhedral dichotomy f on $\mathbb{R}^2$ is locally realizable around a point P in $\mathbb{R}^2$ iff f is not in XOR-bow-tie at P.

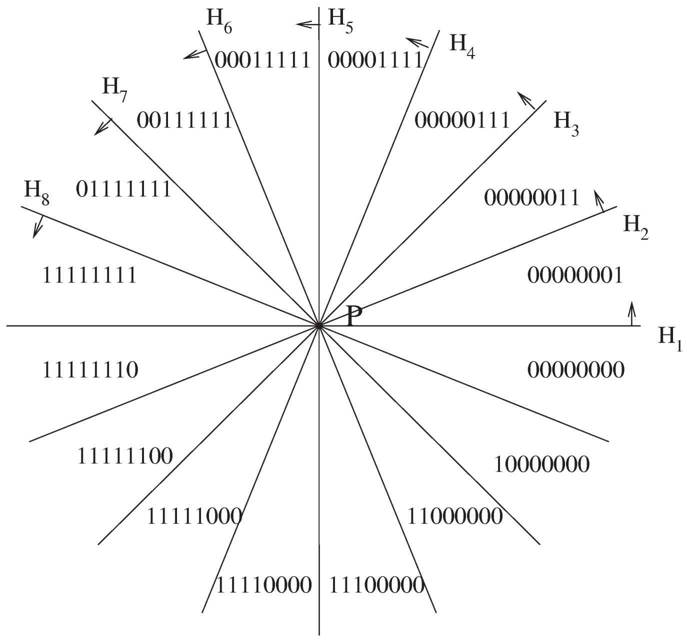

Figure 10. A bundle of essential hyperplanes intersecting at P, and the labels of local regions, after having renumbered and reoriented the hyperplanes.

This result is specially interesting if P is a point of multiple intersection, i.e., where more than two essential hyperplanes intersect.

Proof. Consider as neighborhood of P any open ball C_P which includes P and which does not intersect any essential hyperplane other than those going through P. The proof is in three steps. First, we reorder the hyperplanes in the neighborhood C_P of P, so as to get an appropriate labelling of all the regions whose closure contains P (see figure 10). This is possible because we are in two dimensions, and so the hyperplanes going through P are just lines and can be ordered by slope. Second, we apply Farkas' lemma to the "nice" looking system of inequalities resulting from the appropriate labelling. Third, we show how an XOR-bow-tie can be deduced.

The problem of finding a one-hidden-layer network with no redundant hyperplanes and which locally realizes f can be rewritten as a system ($\mathcal{S}$) of inequalities obtained from the $2k$ regions which have the point P in their closure.

$$(\mathcal{S})\quad \begin{cases} 1 \leqslant i \leqslant k & \left[\begin{array}{ll} \sum_{m=1}^{i} w_m < \theta & \text{if region } i \text{ of class } 0, \\ \sum_{m=1}^{i} w_m > \theta & \text{if region } i \text{ of class } 1, \end{array}\right. \\ k+1 \leqslant i \leqslant 2k & \left[\begin{array}{ll} \sum_{m=i-k+1}^{k} w_m < \theta & \text{if region } i \text{ of class } 0, \\ \sum_{m=i-k+1}^{k} w_m > \theta & \text{if region } i \text{ of class } 1. \end{array}\right. \end{cases} \tag{2}$$

The system ($\mathcal{S}$) can be rewritten in matrix form $Ax \leqslant b$, where

$$x^{\mathrm{T}} = [w_1, w_2, \ldots, w_k, \theta]$$

is the unknown vector of weights and threshold, for the output unit, and

$$b^{\mathrm{T}} = [b_1, b_2, \ldots, b_k, b_{k+1}, \ldots, b_{2k}]$$

is defined by $b_i = -\eta$, for all i, where η is an arbitrarily small positive number.

The matrix A has the following expression:

$$A = \begin{bmatrix} \varepsilon_1 & 0 & \ldots & \ldots & 0 & -\varepsilon_1 \\ \varepsilon_2 & \varepsilon_2 & 0 & \ldots & 0 & -\varepsilon_2 \\ \vdots & & \ddots & \ddots & \vdots & \vdots \\ \vdots & & & \varepsilon_{k-1} & 0 & \vdots \\ \varepsilon_k & \ldots & \ldots & \varepsilon_k & \varepsilon_k & -\varepsilon_k \\ 0 & \varepsilon_{k+1} & \ldots & \ldots & \varepsilon_{k+1} & -\varepsilon_{k+1} \\ 0 & 0 & \varepsilon_{k+2} & & \vdots & -\varepsilon_{k+2} \\ \vdots & & \ddots & \ddots & \vdots & \vdots \\ \vdots & & & \ddots & \varepsilon_{2k-1} & -\varepsilon_{2k-1} \\ 0 & \ldots & \ldots & \ldots & 0 & -\varepsilon_{2k} \end{bmatrix},$$

where $\varepsilon_j = +1$ if the region j is in class 0 or $\varepsilon_j = -1$ if this region is in class 1.

The next step is to apply Farkas' lemma, or an equivalent version [12], which gives a necessary and sufficient condition for finding a solution of $Ax \leqslant b$.

Proposition 16 (Farkas' lemma, equivalent version). There does not exist a vector $x \in \mathbb{R}^n$ such that $Ax \leqslant b$ iff there exists a vector $y \in \mathbb{R}^m$ such that $y^{\mathrm{T}}A = 0$, $y \geqslant 0$ and $y^{\mathrm{T}}b < 0$.

Assume that $Ax \leqslant b$ is not solvable and apply Farkas' lemma with $n = k+1$ and $m = 2k$. The equations corresponding to $y^{\mathrm{T}}A = 0$ are

$$(\mathcal{E}) \quad \begin{cases} 1 \leqslant i \leqslant k & \sum_{m=i}^{i+k-1} \varepsilon_m y_m = 0, \\ & \sum_{m=1}^{2k} -\varepsilon_m y_m = 0. \end{cases}$$

Substracting the ith equation from the $(i+1)$th equation, we get

$$(1 \leqslant i \leqslant k-1) \quad \varepsilon_i y_i = \varepsilon_{i+k} y_{i+k}.$$

Adding the first and kth equations and substracting them to the last one, we get

$$\varepsilon_k y_k = \varepsilon_{2k} y_{2k}.$$

Since $y \geqslant 0$, we see that whenever $y_i \neq 0$, the two regions i and $i+k$ are in the same class.

Now, since $y^{\mathrm{T}} b < 0$, vector y is not uniformly 0. Moreover, the last equation of $(\mathcal{E})$ can be rewritten as

$$\sum_{i \mid \text{region } i \text{ in class } 0} y_i = \sum_{i \mid \text{region } i \text{ in class } 1} y_i.$$

Hence there exist j_1 and j_2 such that $y_{j_1} \neq 0$, $y_{j_2} \neq 0$, and regions j_1 and j_2 are in different classes.

Hence, we have proved that, if the system $(\mathcal{S})$ $Ax \leqslant b$ is not solvable, then there exist two opposite regions in class 0 and two opposite regions in class 1, "opposite" with respect to their labels around the point P, which is exactly the definition of a XOR-bow-tie configuration at P (cf. figure 4). □

5.2. Local realization and XOR-at-infinity

If no two essential hyperplanes are parallel, the case of unbounded regions is exactly the same as a multiple intersection, the point P being replaced by ∞. The case of parallel hyperplanes (see an example in figure 11) is more intricate.

Example 17. For the example of figure 11, matrix A takes the expression

$$A = \begin{bmatrix}
\varepsilon_1 & 0 & 0 & 0 & 0 & 0 & 0 & -\varepsilon_1 \\
\varepsilon_2 & \varepsilon_2 & 0 & 0 & 0 & 0 & 0 & -\varepsilon_2 \\
\varepsilon_3 & \varepsilon_3 & \varepsilon_3 & 0 & 0 & 0 & 0 & -\varepsilon_3 \\
\varepsilon_4 & \varepsilon_4 & \varepsilon_4 & \varepsilon_4 & 0 & 0 & 0 & -\varepsilon_4 \\
\varepsilon_5 & \varepsilon_5 & \varepsilon_5 & \varepsilon_5 & \varepsilon_5 & 0 & 0 & -\varepsilon_5 \\
\varepsilon_6 & \varepsilon_6 & \varepsilon_6 & \varepsilon_6 & \varepsilon_6 & \varepsilon_6 & 0 & -\varepsilon_6 \\
\varepsilon_7 & \varepsilon_7 & \varepsilon_7 & \varepsilon_7 & \varepsilon_7 & \varepsilon_7 & \varepsilon_7 & -\varepsilon_7 \\
0 & \varepsilon_8 & \varepsilon_8 & \varepsilon_8 & \varepsilon_8 & \varepsilon_8 & \varepsilon_8 & -\varepsilon_8 \\
0 & \varepsilon_9 & \varepsilon_9 & \varepsilon_9 & 0 & \varepsilon_9 & \varepsilon_9 & -\varepsilon_9 \\
0 & \varepsilon_{10} & \varepsilon_{10} & 0 & 0 & \varepsilon_{10} & \varepsilon_{10} & -\varepsilon_{10} \\
0 & \varepsilon_{11} & 0 & 0 & 0 & \varepsilon_{11} & \varepsilon_{11} & -\varepsilon_{11} \\
0 & 0 & 0 & 0 & 0 & \varepsilon_{12} & \varepsilon_{12} & -\varepsilon_{12} \\
0 & 0 & 0 & 0 & 0 & 0 & \varepsilon_{13} & -\varepsilon_{13} \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & -\varepsilon_{14}
\end{bmatrix},$$

where $\varepsilon_j = +1$ if the region j is in class 0 or $\varepsilon_j = -1$ if this region is in class 1.

Theorem 18. Let f be a polyhedral dichotomy on $\mathbb{R}^2$. Let C_∞ be the complementary region of the convex hull of the essential points of f. The restriction of f to C_∞ is realizable by a one-hidden-layer network iff f is not in XOR-at-infinity.

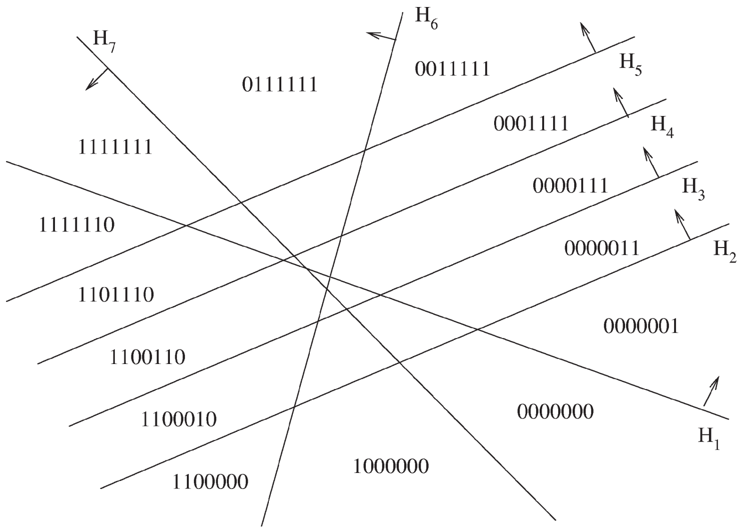

Figure 11. A bundle of parallel essential hyperplanes, and the labels of unbounded regions, after having correctly renumbered and re-oriented the hyperplanes.

The proof is based on a similar reasoning, but it requires a heavy case-by-case analysis (see [3] for details). From theorems 15 and 18 we deduce that a polyhedral dichotomy is locally realizable in $\mathbb{R}^2$ by a one-hidden-layer network iff f has no XOR-bow-tie and no XOR-at-infinity. Unfortunately this result cannot be extended to the global realization of f in $\mathbb{R}^2$.

6. Conclusion

In this article, after setting the basic definitions, we have summarized several approaches aiming to determine which polyhedral dichotomies can be realized by a one-hidden-layer network and which cannot. Four geometric configurations are defined in $\mathbb{R}^d$: the XOR-situation, the XOR-bow-tie, the XOR-at-infinity and the critical cycle. The first three configurations belong to the family of 2-cycles, since they involve four regions, two in each class. The last one is at least a 6-cycle. All these configurations prevent a dichotomy to be realized with a single hidden layer. We conjecture that, in $\mathbb{R}^2$, they are the only ones, but converse results are much more challenging. Partial results about local realizability are developed. Their proof is based on using Farkas' lemma. Our further research will be re-oriented towards another point of view, giving a greater importance to the internal representation of the dichotomy.

References

[1] E.B. Baum, On the capabilities of multilayer perceptrons, J. Complexity 4 (1988) 193–215.

[2] E.K. Blum and L.K. Li, Approximation theory and feedforward networks, Neural Networks 4(4) (1991) 511–516.

[3] G. Brightwell, C. Kenyon and H. Paugam-Moisy, Multilayer neural networks: one or two hidden layers? Research Report 96-37, LIP, ENS Lyon (1996) (available by ftp at lip.ens-lyon.fr).

[4] G. Brightwell, C. Kenyon and H. Paugam-Moisy, Multilayer neural networks: one or two hidden layers?, in: *Advances in Neural Information Processing Systems*, Vol. 9, *Proc. of NIPS '96*, eds. M.C. Mozer, M.I. Jordan and T. Petsche (MIT Press, Cambridge, MA, 1997) pp. 148–154.

[5] M. Cosnard, P. Koiran and H. Paugam-Moisy, Complexity issues in neural network computations, in: *Proc. of LATIN '92*, ed. I. Simon, Lecture Notes in Computer Science, Vol. 583 (Springer, New York, 1992) pp. 530–544.

[6] M. Cosnard, P. Koiran and H. Paugam-Moisy, A step towards the frontier between one-hidden-layer and two-hidden layer neural networks, in: *Proc. of IJCNN '93-Nagoya*, ed. I. Simon, Vol. 3 (Springer, Berlin, 1993) pp. 2292–2295.

[7] G. Cybenko, Approximation by superpositions of a sigmoidal function, Math. Control Signals Systems 2 (1988) 303–314.

[8] K. Funahashi, On the approximate realization of continuous mappings by neural networks, Neural Networks 2(3) (1989) 183–192.

[9] G.J. Gibson, A combinatorial approach to understanding perceptron decision regions, IEEE Trans. Neural Networks 4 (1993) 989–992.

[10] G.J. Gibson, Exact classification with two-layer neural nets, J. Comput. System Sci. 52(2) (1996) 349–356.

[11] G.J. Gibson and C.F.N. Cowan, On the decision regions of multilayer perceptrons, Proc. IEEE 78(10) (1990) 1590–1594.

[12] M. Grötschel, L. Lovsz and A. Schrijver, *Geometric Algorithms and Combinatorial Optimization* (Springer, Berlin/Heidelberg, 1988).

[13] C.Z.W. Hassell Sweatman, G.J. Gibson and B. Mulgrew, Exact classification with two-layer neural nets in n dimensions, J. Discrete Appl. Math. (1998, to appear).

[14] K. Hornik, M. Stinchcombe and H. White, Multilayer feedforward networks are universal approximators, Neural Networks 2(5) (1989) 359–366.

[15] 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.

[16] S. Muroga, *Threshold Logic and Its Applications* (Wiley, New York, 1971).

[17] H. Takahashi, E. Tomita and T. Kawabata, Separability of internal representations in multilayer perceptrons with application to learning, Neural Networks 6 (1993) 689–703.

[18] P.J. Zweitering, The complexity of multi-layered perceptrons, Ph.D. thesis, Technische Universiteit Eindhoven (1994).