Max-Planck-Institut
für Mathematik
in den Naturwissenschaften
Leipzig

Hyperplane arrangements separating arbitrary vertex classes in n-cubes

by

Walter Wenzel, Nihat Ay and Frank Pasemann

Preprint no.: 35 1999

Hyperplane Arrangements Separating Arbitrary Vertex Classes in n-Cubes*

Walter Wenzel†

Institute of Mathematics, Technical University of Chemnitz,
D-09107 Chemnitz, Germany,

Nihat Ay, and Frank Pasemann

Max-Planck-Institute for Mathematics in the Sciences,
D-04103 Leipzig, Germany.

Abstract

Strictly layered feedforward networks with binary neurons are viewed as maps from the vertex set of an n-cube to the vertex set of an l-cube. With only one output neuron in principle they can realize any Boolean function on n inputs. We address the problem of determining the necessary and sufficient numbers of hidden units for this task by using separability properties of affine oriented hyperplane arrangements.

Keywords: feedforward networks, binary units, classification problems, hypercube, separability, affine oriented hyperplane arrangements, linear codes

*submitted for publication

†Research supported in part by MPI for Mathematics in the Sciences.

1 Introduction

The problem addressed in this paper stems from an unsolved question in the theory of neural networks [7]. There it was proven that so called feedforward networks may serve as universal approximators, that is, under quite general regularity assumptions a network with sufficiently many hidden neurons can approximate any member of a class of functions to any desired degree of accuracy [4], [5], [6], [8], [12]. Although these theorems guarantee the existence of neural network solutions for such problems, it is still an open question, how to find a good upper bound for the number of hidden units to use. This, of course, depends on the problem and on the desired degree of accuracy [2], [3], [10], [13].
To get better access to analytical considerations for this problem, we will reduce it in several steps. First, one may consider only categorization tasks: A set of points in the input space $\mathbb{R}^n$ has to be mapped e.g. to the values 1 and -1. In a second step, this can be further reduced to the problem of approximating a Boolean function on n inputs, i.e. mapping the vertices of a hypercube in $\mathbb{R}^n$ to values 1 and -1.
This kind of classification problems can be treated by 2-layer networks [7], where the so called *hidden layer* has e.g. l units getting signals from the n inputs; and the single unit of the *output layer* gets l signals from the hidden layer. Usually the units of feedforward networks are given as composition of affine functions on their input space with a differentiable transfer function of sigmoidal characteristic, for example $\sigma_r(x) := \tanh(r\,x)$, $r \in \mathbb{R}$. Thus the output of hidden unit i is given in the form

$$o_i(x) := \sigma_r\Big(\theta_i + \sum_{j=1}^{n} w_{ij}\,x_j\Big)\ , \quad i = 1, \ldots, l, \quad x \in \mathbb{R}^n\ ,$$

where θ_i is a constant, the *bias term* of the unit, and $w_i = (w_{i1}, \ldots, w_{in}) \in \mathbb{R}^n$ denotes its *weight vector.* Every such unit partitions its input space $\mathbb{R}^n$ into two half spaces separated by its so called *center* H_i, which is here defined by

$$H_i := \{x \in \mathbb{R}^l \mid w \cdot x = -\theta_i\} \quad .$$

In the last step we let the slope of the sigmoid go to infinity, i.e. $r \to \infty$, so that the sigmoid approximates a step function, without moving the center H_i, and associates to the half spaces separated by the center the values 1 and -1. Thus we are referring to feedforward networks with binary neurons.
Using this approach, the hidden layer of a neural network maps the binary input patterns of an n-cube to binary patterns of an l-cube. These l-dimensional patterns then have to be separated by the center of the output unit in such a way that the values 1 and -1 give the correct classification of the input patterns.
In section 2 we formulate the problem in geometrical terms and present some elementary results. In the following section we specify assumptions under which

compositions of hyperplane arrangements separate unions of patterns belonging to different classes. This leads to the result that each subset of the vertex set W_n of the n-cube may be separated by at most $\frac{3}{n+2} \cdot 2^n$ affine hyperplanes. In section 4, we obtain the result that there exist binary problems for which one needs at least $(2^{\frac{n}{2}} - \frac{n^2}{2})$ affine hyperplanes to separate the patterns belonging to two different classes. Based on these results, some further issues related to the neural network context of this article are shortly discussed in the final section.

2 Problem Formulation and Elementary Results

For $n \geq 1$ we shall study – in some sense to be specified – separations of the n-cube by affine hyperplane arrangements. First we state the following

Convention: An *affine oriented hyperplane* H in $\mathbb{R}^n$ consists of an affine hyperplane H together with a partition

$$\mathbb{R}^n = H^- \uplus H \uplus H^+ \tag{2.1}$$

where H^- and H^+ are specified open and convex half-spaces. Of course, H^- and H^+ are – up to the order – uniquely determined.

Thus, if we speak about an affine oriented hyperplane H, we shall always assume that a partition as in (2.1) is given. To *choose* some affine oriented hyperplane H will mean that H^- and H^+ may be selected arbitrarily.

In the sequel, $W_n = \{1, -1\}^n$ will denote the vertex set of the n-cube for fixed $n \geq 1$.

Definition 2.1 *Assume $l \geq 1$, and A, B are subsets of $\mathbb{R}^l$. A and B will be called* linearly separable*, if there exists some affine oriented hyperplane H in $\mathbb{R}^l$ with $A \subseteq H^+$ and $B \subseteq H^-$.*

Conventions: Assume $\mathcal{H} = (H_1, \ldots, H_l)$ is some l-tuple of affine oriented hyperplanes in $\mathbb{R}^n$ which is generic, that means

$$W_n \cap \bigcup_{i=1}^{l} H_i = \emptyset.$$

For $1 \leq i \leq l$ we define the map $\varphi_i(\mathcal{H}) : W_n \to \{1, -1\}$ by

$$\varphi_i(\mathcal{H})(x) := \varphi_i(\mathcal{H}, x) := \begin{cases} 1 & \text{if} \quad x \in H_i^+ \\ -1 & \text{if} \quad x \in H_i^- \end{cases} . \tag{2.2}$$

Now $\varphi(\mathcal{H}) : W_n \to W_l$ will denote the map given by

$$\varphi(\mathcal{H})(x) := \varphi(\mathcal{H}, x) := (\varphi_1(\mathcal{H}, x), \ldots, \varphi_l(\mathcal{H}, x)). \tag{2.3}$$

For $C \subseteq W_n$ we write of course

$$\varphi(\mathcal{H})(C) := \varphi(\mathcal{H}, C) := \{\varphi(\mathcal{H}, x) \; : \; x \in C\}. \tag{2.4}$$

Definition 2.2 *Assume $C \subseteq W_n$ and $\mathcal{H} = (H_1, \ldots, H_l)$ is some generic hyperplane arrangement of affine oriented hyperplanes in $\mathbb{R}^n$. We say that $\mathcal{H}$* separates the vertex set C*, if the sets $\varphi(\mathcal{H}, C)$ and $\varphi(\mathcal{H}, W_n - C)$ are linearly separable (as subsets of $\mathbb{R}^l$).*

From now on, we call a generic hyperplane arrangement $(H_1, \ldots, H_l)$ of affine oriented hyperplanes in $\mathbb{R}^n$ also an *l-arrangement* for brevity.

Definition 2.3 *For $C \subseteq W_n$ with $\emptyset \neq C \neq W_n$ we put*

$$\begin{aligned} h(C) :=& \\ &\min\{l \in \mathbb{N} \; : \; \textit{there exists some } l\textit{-arrangement in } \mathbb{R}^n \textit{ which separates } C\}. \end{aligned} \tag{2.5}$$

By convention, we write

$$h(\emptyset) = h(W_n) = 0. \tag{2.6}$$

Remarks:
(i) By Definition 2.2, it is not trivial that every $C \subseteq W_n$ may be separated by some l-arrangement for an appropriate number $l \in \mathbb{N}$. However, we shall see later (c.f. Theorem 3.16) that every C may be separated by at most $\frac{3}{n+2} \cdot 2^n$ affine hyperplanes; that means, we have

$$h(C) \leq \frac{3}{n+2} \cdot 2^n.$$

(ii) By the above definitions, a subset $C \subseteq W_n$ with $\emptyset \neq C \neq W_n$ satisfies $h(C) = 1$ if and only if C and $W_n \backslash C$ are linearly separable.
(iii) By Definition 2.2, every $C \subseteq W_n$ satisfies

$$h(C) = h(W_n \backslash C).$$

If $\mathcal{H}$ separates C, then one has $\varphi(\mathcal{H}, C) \cap \varphi(\mathcal{H}, W_n \backslash C) = \emptyset$.
(iv) By symmetry of the l-cube, for some l-arrangement $(H_1, \ldots, H_l)$ to separate a set $C \subseteq W_n$ it does not matter in which way the half spaces corresponding to $H_1, \ldots, H_l$ are oriented.

Example 2.4 (The XOR-Problem) *Assume $n = 2$, and put*

$$A := \{(1,1), (-1,-1)\}, \qquad B := \{(1,-1), (-1,1)\}.$$

If $C \subseteq W_2$ satisfies $C \neq A$ and $C \neq B$, then C and $W_2 \backslash C$ are linearly separable. However A and B are not linearly separable, because

$$(0,0) \in \operatorname{conv} A \; \cap \; \operatorname{conv} B,$$

where conv *denotes the convex closure operator.*
For $i \in \{1, 2\}$ *put*

$$H_i := \{(x_1, x_2) \in \mathbb{R}^2 \ : \ x_1 + x_2 = 3 - 2i\}$$

as well as

$$\begin{aligned} H_i^+ &:= \{(x_1, x_2) \in \mathbb{R}^2 \ : \ x_1 + x_2 > 3 - 2i\} \\ H_i^- &:= \mathbb{R}^2 \backslash (H_i \cup H_i^+). \end{aligned}$$

Then $\mathcal{H} = (H_1, H_2)$ *is some* 2*-arrangement separating* A *and* B*: We get*

$$\begin{aligned} \varphi(\mathcal{H})(\{(1,1)\}) &= \{(1,1)\}, \\ \varphi(\mathcal{H})(\{(1,-1),(-1,1)\}) &= \{(-1,1)\}, \\ \varphi(\mathcal{H})(\{(-1,-1)\}) &= \{(-1,-1)\} \end{aligned}$$

and thus

$$\begin{aligned} \varphi(\mathcal{H}, A) &= \{(1,1),(-1,-1)\} =: A', \\ \varphi(\mathcal{H}, B) &= \{(-1,1)\} =: B'. \end{aligned}$$

Of course, A' *and* B' *are linearly separable in* $\mathbb{R}^2$*. We obtain* $h(A) = h(B) = 2$*.*

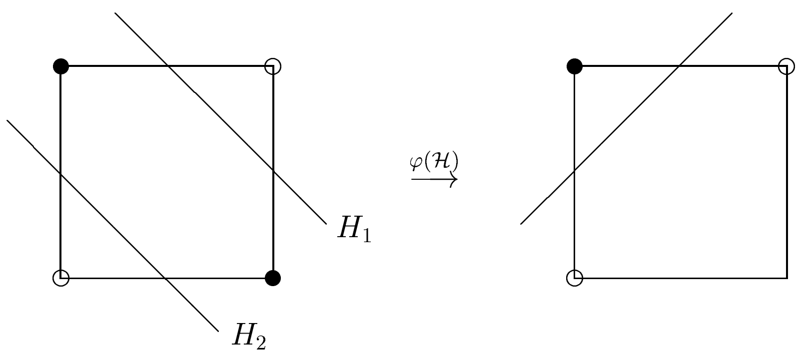

Figure 1: Two hyperplanes separating the input patterns of the XOR-problem (Example 2.4), and the separating hyperplane for their image under φ.

In what follows, $\langle \cdot, \cdot \rangle$ will denote the standard scalar product in $\mathbb{R}^l$; that means, for $v = (v_1, \ldots, v_l) \in \mathbb{R}^l$ and $w = (w_1, \ldots, w_l) \in \mathbb{R}^l$ we write

$$\langle v, w \rangle := \sum_{i=1}^{l} v_i \cdot w_i.$$

Next we prove the following simple

Lemma 2.5 *Assume the l-arrangement $\mathcal{H} = (H_1, \ldots, H_l)$ separates the set $C \subseteq W_n$. Then the following holds:*
*(***i***) If $\sigma \in S_l$ is some permutation, then $(H_{\sigma(1)}, \ldots, H_{\sigma(l)})$ separates the set C, too.*
*(***ii***) If $W_n \subseteq H_l^+$ or $W_n \subseteq H_l^-$, then $\mathcal{H}' = (H_1, \ldots, H_{l-1})$ separates C.*
*(***iii***) If $W_n \cap H_{l-1}^+ = W_n \cap H_l^+$ or $W_n \cap H_{l-1}^+ = W_n \cap H_l^-$, then $\mathcal{H}' = (H_1, \ldots, H_{l-1})$ separates C. In particular, $\mathcal{H}' = (H_1, \ldots, H_{l-1})$ separates C in case $H_{l-1} = H_l$.*
*(***iv***) If H_{l+1} is any affine oriented hyperplane in $\mathbb{R}^n$ with $H_{l+1} \cap W_n = \emptyset$, then $\mathcal{H}'' = (H_1, \ldots, H_l, H_{l+1})$ separates C, too.*

Proof.
(i) If $A, B \subseteq \mathbb{R}^l$ are linearly separable by some affine oriented hyperplane H in $\mathbb{R}^l$ and $\alpha : \mathbb{R}^l \to \mathbb{R}^l$ is some bijective affine map, then $\alpha(A)$ and $\alpha(B)$ are of course linearly separable by the affine hyperplane $\alpha(H)$; here we may put $\alpha(H)^+ := \alpha(H^+)$ and $\alpha(H)^- := \alpha(H^-)$. This holds in particular if α is some linear isomorphism which merely permutes coordinates.
(ii) Without loss of generality, we may assume $\emptyset \neq C \neq W_n$ and $W_n \subseteq H_l^+$. Thus for all $x \in W_n$ one has $\varphi_l(\mathcal{H}, x) = 1$. Assume $\varphi(\mathcal{H}, C)$ and $\varphi(\mathcal{H}, W_n \backslash C)$ are linearly separable by the affine oriented hyperplane H in $\mathbb{R}^l$. Then for suitable $w = (w_1, \ldots, w_l) \in \mathbb{R}^l \backslash \{0\}$ and $t \in \mathbb{R}$ we get

$$H = \{v \in \mathbb{R}^l \ : \ \langle v, w \rangle = t\},$$

$$\varphi(\mathcal{H}, C) \subseteq H^+ = \{v \in \mathbb{R}^l \ : \ \langle v, w \rangle > t\},$$
$$\varphi(\mathcal{H}, W_n \backslash C) \subseteq H^- = \{v \in \mathbb{R}^l \ : \ \langle v, w \rangle < t\}.$$

Thus, for $w' := (w_1, \ldots, w_{l-1})$ we get

$$\varphi(\mathcal{H}', C) \subseteq \{v' \in \mathbb{R}^{l-1} \ : \ \langle v', w' \rangle > t - w_l\},$$

$$\varphi(\mathcal{H}', W_n \backslash C) \subseteq \{v' \in \mathbb{R}^{l-1} \ : \ \langle v', w' \rangle < t - w_l\}.$$

In particular, we have $w' \neq 0$, because $C \neq \emptyset \neq W_n \backslash C$. Thus, $\varphi(\mathcal{H}', C)$ and $\varphi(\mathcal{H}', W_n \backslash C)$ are linearly separable by the affine hyperplane

$$H' := \{v' \in \mathbb{R}^{l-1} \ : \ \langle v', w' \rangle = t - w_l\}.$$

(iii) Without loss of generality, we may suppose $W_n \cap H_{l-1}^+ = W_n \cap H_l^+$ and $\emptyset \neq C \neq W_n$. Assume again that $w = (w_1, \ldots, w_l) \in \mathbb{R}^l \backslash \{0\}$ and $t \in \mathbb{R}$ satisfy

$$\varphi(\mathcal{H}, C) \subseteq \{v \in \mathbb{R}^l \ : \ \langle v, w \rangle > t\},$$

$$\varphi(\mathcal{H}, W_n \backslash C) \subseteq \{v \in \mathbb{R}^l \ : \ \langle v, w \rangle < t\}.$$

Now put $w' := (w_1, \ldots, w_{l-2}, w_{l-1} + w_l)$. Since every $v = (v_1, \ldots, v_l) \in \varphi(\mathcal{H}, W_n)$ satisfies $v_{l-1} = v_l$, we get

$$\varphi(\mathcal{H}', C) \subseteq \{v' \in \mathbb{R}^{l-1} \ : \ \langle v', w' \rangle > t\},$$

$$\varphi(\mathcal{H}', W_n \backslash C) \subseteq \{v' \in \mathbb{R}^{l-1} \ : \ \langle v', w' \rangle < t\}.$$

Now $C \neq \emptyset \neq W_n \backslash C$ implies $w' \neq 0$; therefore, $\varphi(\mathcal{H}', C)$ and $\varphi(\mathcal{H}', W_n \backslash C)$ are linearly separable by the affine hyperplane

$$H' := \{v' \in \mathbb{R}^{l-1} \ : \ \langle v', w' \rangle = t\}.$$

(iv) Choose once more $w = (w_1, \ldots, w_l) \in \mathbb{R}^l \backslash \{0\}$ and $t \in \mathbb{R}$ with

$$\varphi(\mathcal{H}, C) \subseteq \{v \in \mathbb{R}^l \ : \ \langle v, w \rangle > t\},$$

$$\varphi(\mathcal{H}, W_n \backslash C) \subseteq \{v \in \mathbb{R}^l \ : \ \langle v, w \rangle < t\}.$$

Now put $w'' := (w_1, \ldots, w_l, 0)$. Then we get

$$\varphi(\mathcal{H}'', C) \subseteq \{v'' \in \mathbb{R}^{l+1} \ : \ \langle v'', w'' \rangle > t\},$$

$$\varphi(\mathcal{H}'', W_n \backslash C) \subseteq \{v'' \in \mathbb{R}^{l+1} \ : \ \langle v'', w'' \rangle < t\}.$$

Thus, $\varphi(\mathcal{H}'', C)$ and $\varphi(\mathcal{H}'', W_n \backslash C)$ are linearly separable by the affine hyperplane

$$H' := \{v'' \in \mathbb{R}^{l+1} \ : \ \langle v'', w'' \rangle = t\}. \qquad \square$$

The next result shows that several subsets $C \subseteq W_n$ consisting of certain layers may be separated by some hyperplane arrangement which is induced by these layers in a canonical way.

Proposition 2.6 *Let $\mathcal{H} = (H_1, \ldots, H_l)$ denote some l-arrangement in $\mathbb{R}^n$ satisfying*

$$H_i^+ \cap W_n \subseteq H_{i+1}^+ \quad \text{for} \quad 1 \leq i \leq l-1. \tag{2.7}$$

Choose affine oriented hyperplanes H_0, H_{l+1} in $\mathbb{R}^n$ with $H_0^+ \cap W_n = \emptyset$ and $W_n \subseteq H_{l+1}^+$.
Let $C \subseteq W_n$ denote that subset of vertices of the n-cube such that for every i with $0 \leq i \leq l$ one has

$$W_n \cap (H_{i+1}^+ \backslash H_i^+) \subseteq \left\{ \begin{array}{ccc} C & \text{for} & i \equiv 1 \bmod 2 \\ W_n \backslash C & \text{for} & i \equiv 0 \bmod 2 \end{array} \right. . \tag{2.8}$$

Then $\mathcal{H}$ separates the set C.

Proof. By the assumptions of the proposition, for every $x \in W_n$ there exists some unique i with $0 \leq i \leq l$ satisfying $x \in H_i^- \cap H_{i+1}^+$. We get

$$\varphi(\mathcal{H}, x) = (\underbrace{-1, \ldots, -1}_{i}, \underbrace{1, \ldots, 1}_{l-i}), \tag{2.9}$$

and (2.8) implies

$$x \in \left\{ \begin{array}{ccc} C & \text{for} & i \equiv 1 \bmod 2 \\ W_n \backslash C & \text{for} & i \equiv 0 \bmod 2 \end{array} \right. . \tag{2.10}$$

Put

$$a := \left\{ \begin{array}{ccc} 0 & \text{for} & l \equiv 1 \bmod 2 \\ -1 & \text{for} & l \equiv 0 \bmod 2 \end{array} \right. ,$$

and define the linear map $f : \mathbb{R}^l \to \mathbb{R}$ by

$$f(v_1, \ldots, v_l) := \sum_{i=1}^{l} (-1)^{i+1} \cdot v_i.$$

Consider the affine oriented hyperplane G in $\mathbb{R}^l$ given by

$$\begin{array}{rcl} G & := & \{v \in \mathbb{R}^l \ : \ f(v) = a\}, \\ G^+ & := & \{v \in \mathbb{R}^l \ : \ f(v) > a\}, \\ G^- & := & \{v \in \mathbb{R}^l \ : \ f(v) < a\}. \end{array}$$

Then (2.9) and (2.10) imply

$$\varphi(\mathcal{H}, x) \in G^- \quad \text{for} \quad x \in C,$$

$$\varphi(\mathcal{H}, x) \in G^+ \quad \text{for} \quad x \in W_n \backslash C.$$

Thus $\varphi(\mathcal{H}, C)$ and $\varphi(\mathcal{H}, W_n \backslash C)$ are linearly separable by G. □

Remark: The affine oriented hyperplanes H_0 and H_{l+1} in the last result are of course only used for technical reasons.

One of the most important applications of Proposition 2.6 is to study the following

Problem 2.7 (Parity Problem) *For $n \geq 1$ put*[1]

$$C_P(n) := \{(x_1, \ldots, x_n) \in W_n \ : \ |\{i \ : \ x_i = -1\}| \equiv 1 \bmod 2\}. \tag{2.11}$$

Separate $C_P(n)$.

[1]Here – as in the sequel – $|A|$ denotes the cardinality of a finite set A.

The following theorem gives an upper bound for $h(C_P(n))$.

Theorem 2.8 *For all $n \geq 1$ one has*

$$h(C_P(n)) \leq n; \tag{2.12}$$

that is, $C_P(n)$ may be separated by some n-arrangement $(H_1, \ldots, H_n)$.

Proof. For $0 \leq i \leq n+1$ put

$$\begin{aligned} H_i &:= \left\{ (v_1, \ldots, v_n) \in \mathbb{R}^n \ : \ \sum_{j=1}^{n} v_j = n + 1 - 2i \right\}, \\ H_i^+ &:= \left\{ (v_1, \ldots, v_n) \in \mathbb{R}^n \ : \ \sum_{j=1}^{n} v_j > n + 1 - 2i \right\}, \\ H_i^- &:= \mathbb{R}^n \backslash (H_i \cup H_i^+). \end{aligned}$$

Then $H_0, H_1, \ldots, H_n, H_{n+1}$ and $C = C_P(n)$ fulfill the assumptions of Proposition 2.6; thus $\mathcal{H} = (H_1, \ldots, H_n)$ separates $C_P(n)$. □

3 Separations of Unions

In this section, we want to study unions of subsets of W_n and show that – under some certain supposition – separations of these subsets induce some separation of their union. Concerning the additional assumption, we state the following

Definition 3.1 *Assume $C \subseteq W_n$. An l-arrangement $\mathcal{H} = (H_1, \ldots, H_l)$ is called a* centered image separation *of C, if there exists some affine hyperplane G in $\mathbb{R}^l$, some affine map $f_G : \mathbb{R}^l \to \mathbb{R}$ with $G = f_G^{-1}(\{0\})$ as well as some $d > 0$ such that for $x \in W_n$ one has*

$$f_G(\varphi(\mathcal{H}, x)) = \begin{cases} d & \textit{for} \quad x \in C \\ -d & \textit{for} \quad x \in W_n \backslash C \end{cases}. \tag{3.1}$$

In other words, the following two conditions hold:
(i) *$\varphi(\mathcal{H}, C)$ and $\varphi(\mathcal{H}, W_n \backslash C)$ are linearly separable by G. (This means that $\mathcal{H}$ separates C.)*
(ii) *All points $\varphi(\mathcal{H}, x)$, $x \in W_n$, have the same distance to G.*

Examples 3.2
(i) *Assume $l = 1$; that is, C and $W_n \backslash C$ are linearly separable by some affine oriented hyperplane $H \subseteq \mathbb{R}^n$. Then the single hyperplane arrangement $\mathcal{H} = (H)$ is a centered image separation of C:*
If, say, $C \subseteq H^+$ and $W_n \backslash C \subseteq H^-$, we get

$$\varphi(\mathcal{H}, x) = \begin{cases} 1 & \textit{for} \quad x \in C \\ -1 & \textit{for} \quad x \in W_n \backslash C \end{cases}.$$

Thus (3.1) holds for $d=1$*,* $G=\{0\}\subseteq\mathbb{R}$ *and the identity map* $f_G:\mathbb{R}\to\mathbb{R}$*.*
(ii) *Assume* $l=2$*,* $C\subseteq W_n$*, and* $H_1,H_2\subseteq\mathbb{R}^n$ *are affine oriented hyperplanes satisfying*

$$C\subseteq H_1^+\cap H_2^-, \tag{3.2}$$

$$W_n\backslash C\subseteq (H_1^+\cap H_2^+)\cup(H_1^-\cap H_2^-). \tag{3.3}$$

Then $\mathcal{H}:=(H_1,H_2)$ *is a centered image separation of* C*:*
Put $G:=\{(x_1,x_2)\in\mathbb{R}^2\ :\ x_1-x_2=1\}$*, and define* $f_G:\mathbb{R}^2\to\mathbb{R}$ *by* $f_G(x_1,x_2):=x_1-x_2-1$*. Then we have* $G=f_G^{-1}(\{0\})$ *as well as*

$$f_G(\varphi(\mathcal{H},x))=\begin{cases}1 & \textit{for}\quad x\in C\\ -1 & \textit{for}\quad x\in W_n\backslash C\end{cases}. \tag{3.4}$$

Example 3.3 *Assume* $n=2$*, and put* $C:=\{(1,1)\}$ *as well as*

$$\begin{aligned}H_0 &:= \{(x_1,x_2)\in\mathbb{R}^2\ :\ x_1+x_2=1\},\\ H_1 &:= \{(x_1,x_2)\in\mathbb{R}^2\ :\ x_1=0\},\\ H_2 &:= \{(x_1,x_2)\in\mathbb{R}^2\ :\ x_2=0\}.\end{aligned}$$

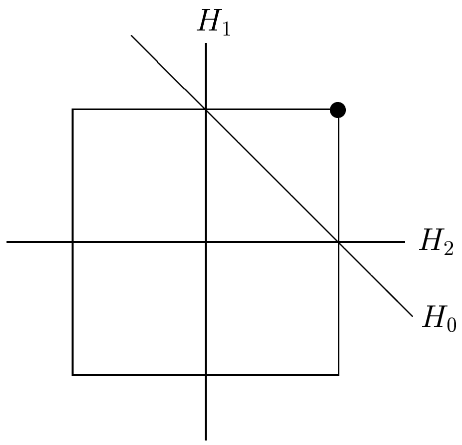

Figure 2: The two hyperplane arrangements $\mathcal{H}:=(H_1,H_2)$ and (H_0) of Example 3.3.

C *and* $W_2\backslash C$ *are linearly separable by* H_0*; thus, by Example 3.2 (i), the single hyperplane arrangement* (H_0) *is a centered image separation of* C*.*
Moreover, $\mathcal{H}:=(H_1,H_2)$ *separates* C*, too; however,* $\mathcal{H}$ *is not some centered image separation of* C*. Indeed,* $\varphi(\mathcal{H}):W_2\to W_2$ *is – without loss of generality – the identity map, and* $G:=H_0$ *is the unique affine hyperplane in* $\mathbb{R}^2$ *which linearly separates* C *from* $W_2\backslash C$ *such that the three points* $(1,1)$*,* $(-1,1)$ *and* $(1,-1)$ *have the same distance to* G*; however,* $(-1,-1)$ *has some larger distance to* $G=H_0$*.*

Now we can prove the following

Proposition 3.4 *Assume $C \subseteq W_n$, and $C_1, \ldots, C_m \subseteq W_n$ satisfy*

$$C = \bigcup_{i=1}^{m} C_i. \tag{3.5}$$

For every i with $1 \leq i \leq m$, assume that $\mathcal{H}_i = (H_1^i, \ldots, H_{l_i}^i)$ is some centered image separation of C_i. Then the composed hyperplane arrangement

$$\mathcal{H} := (H_1^1, \ldots, H_{l_1}^1, \ldots, H_1^m, \ldots, H_{l_m}^m)$$

separates the vertex set C.

Proof. By assumption, for every i with $1 \leq i \leq m$ there exists some $d_i > 0$ as well as some nonconstant affine map $f_i : \mathbb{R}^{l_i} \to \mathbb{R}$ satisfying

$$f_i(\varphi(\mathcal{H}_i, x)) = \begin{cases} d_i & \text{for} \quad x \in C_i \\ -d_i & \text{for} \quad x \in W_n \backslash C_i \end{cases}.$$

Now put $l := \sum_{i=1}^m l_i$, and define the affine map $f : \mathbb{R}^l \to \mathbb{R}$ by

$$f(a_1^{(1)}, \ldots, a_{l_1}^{(1)}, \ldots, a_1^{(m)}, \ldots, a_{l_m}^{(m)}) := \sum_{i=1}^{m} f_i(a_1^{(i)}, \ldots, a_{l_i}^{(i)}).$$

Put $d := \min\{d_1, \ldots, d_m\}$. Then we get:

$$f(\varphi(\mathcal{H}, x)) \geq 2d - \sum_{i=1}^{m} d_i \quad \text{for} \quad x \in C, \tag{3.6}$$

$$f(\varphi(\mathcal{H}, x)) = -\sum_{i=1}^{m} d_i \quad \text{for} \quad x \in W_n \backslash C. \tag{3.7}$$

Thus, $\varphi(\mathcal{H}, C)$ and $\varphi(\mathcal{H}, W_n \backslash C)$ are linearly separable by the affine hyperplane

$$H := \left\{ (a_1, \ldots, a_l) \in \mathbb{R}^l \ : \ f(a_1, \ldots, a_l) = d - \sum_{i=1}^{m} d_i \right\}. \quad \square$$

Remark: Unfortunately, the last result becomes wrong if we do not suppose that each $\mathcal{H}_i$ is some centered image separation of C_i but only assume that $\mathcal{H}_i$ separates C_i.
Consider once more Example 3.3, assume H_1, H_2 are as in this example, but now put $C' := \{(1,1), (-1,-1)\}$. The hyperplane arrangement $\mathcal{H} = (H_1, H_2)$ separates both of the sets $\{(1,1)\}$ and $\{(-1,-1)\}$. However, the composed hyperplane arrangement $\mathcal{H}' := (H_1, H_2, H_1, H_2)$ does not separate C', because otherwise Lemma 2.5 would imply that $\mathcal{H}$ separates C', too. But this is not the case.

As an important special case of Proposition 3.4, we get

Proposition 3.5 *Assume $C, C_1, \dots, C_m \subseteq W_n$ satisfy*

$$C = \bigcup_{i=1}^{m} C_i.$$

Moreover, suppose that for each i with $1 \leq i \leq m$, the sets C_i and $W_n \backslash C_i$ are linearly separable by some affine oriented hyperplane H_i in $\mathbb{R}^n$. Then $\mathcal{H} := (H_1, \dots, H_m)$ separates the set C.

Proof. This result is a trivial consequence of Example 3.2 (i) and Proposition 3.4 . □

We can now also prove that each subset $C \subseteq W_n$ may be separated by some l-arrangement for an appropriate number $l \in \mathbb{N}$. More precisely, we get the following

Theorem 3.6
(i) *For each $x \in W_n$, the sets $\{x\}$ and $W_n \backslash \{x\}$ are linearly separable.*
(ii) *Each subset $C \subseteq W_n$ may be separated by some l-arrangement consisting of $l \leq 2^{n-1}$ affine oriented hyperplanes; that is, one has*

$$h(C) \leq 2^{n-1}. \tag{3.8}$$

Proof.
(i) We write $x = (\varepsilon_1, \dots, \varepsilon_n)$ with $\varepsilon_i \in \{-1, 1\}$ for $1 \leq i \leq n$. Then the sets $\{x\}$ and $W_n \backslash \{x\}$ are linearly separable by the affine hyperplane

$$H := \left\{ (v_1, \dots, v_n) \in \mathbb{R}^n \;:\; \sum_{i=1}^{n} \varepsilon_i \cdot v_i = n - 1 \right\}.$$

(ii) By Remark (iii) following Definition 2.3, we have $h(C) = h(W_n \backslash C)$; therefore, we may assume $|C| \leq |W_n \backslash C|$ and thus $|C| \leq 2^{n-1}$. But then (3.8) follows trivially from (i) and Proposition 3.5. □

At the end of this section, we improve the inequality (3.8).

As a further consequence of Proposition 3.4, we prove

Proposition 3.7 *Assume $C, C_1, \dots, C_m \subseteq W_n$ satisfy*

$$C = \bigcup_{i=1}^{m} C_i.$$

Moreover, for $1 \leq i \leq m$ suppose that there exist affine oriented hyperplanes $G_i, H_i \subseteq \mathbb{R}^n$ as well as subsets $A_i, B_i \subseteq W_n$ satisfying

$$W_n = A_i \uplus B_i \uplus C_i, \tag{3.9}$$

$$\varphi((G_i, H_i), x) = \begin{cases} (-1,-1) & \text{for} \quad x \in A_i \\ (1,-1) & \text{for} \quad x \in C_i \\ (1,1) & \text{for} \quad x \in B_i \end{cases}. \tag{3.10}$$

Then the composed hyperplane arrangement $(G_1, H_1, \ldots, G_m, H_m)$ separates the set C.

Proof. In view of (3.9) and (3.10) we may conclude by Example 3.2 (ii) that for each i with $1 \leq i \leq m$, the pair (G_i, H_i) is a centered image separation of C_i. Thus Proposition 3.4 yields what we want. □

As a special case of Proposition 3.7, we want to point out the following

Proposition 3.8 *Suppose $C, C_1, \ldots, C_m \subseteq W_n$ satisfy*

$$C = \bigcup_{i=1}^{m} C_i.$$

Moreover, assume that for each i with $1 \leq i \leq m$ there exists some affine hyperplane K_i in $\mathbb{R}^n$ satisfying

$$W_n \cap K_i = C_i. \tag{3.11}$$

Then one has $h(C) \leq 2m$.

Proof. For $1 \leq i \leq m$, we may choose affine oriented hyperplanes G_i, H_i in $\mathbb{R}^n$ which are parallel to K_i such that the following conditions hold:

$$(G_i \cup H_i) \cap W_n = \emptyset,$$

$$G_i^- \cap H_i^+ = \emptyset,$$

$$G_i^+ \cap H_i^- \cap W_n = K_i \cap W_n = C_i.$$

Now we can apply Proposition 3.7 to the sets

$$A_i := W_n \cap G_i^- \cap H_i^-, \qquad B_i := W_n \cap G_i^+ \cap H_i^+$$

and conclude that the hyperplane arrangement $(G_1, H_1, \ldots, G_m, H_m)$ separates C. □

In the last part of this section, we want to improve – for all $C \subseteq W_n$ – the upper bound for $h(C)$ as stated in Theorem 3.6 (ii). For this purpose, we study so called frames which cover W_n. First of all, we recall the following

Definition 3.9 *The* Hamming distance *on W_n is the metric $d_H : W_n \times W_n \to \{0, 1, \ldots, n\}$ defined by*

$$d_H((x_1, \ldots, x_n), (x_1', \ldots, x_n')) \;:=\; |\{i \;:\; x_i \neq x_i'\}|. \tag{3.12}$$

Definition 3.10 *A subset $F \subseteq W_n$ is called a* frame *in W_n, if F consists of a distinguished element $y_0 \in W_n$, called the* root *of F, as well as all its neighbours with respect to d_H; that is*

$$F \;=\; \{y_0\} \cup \{y \in W_n \;:\; d_H(y, y_0) = 1\}. \tag{3.13}$$

Clearly, every frame F in W_n satisfies $|F| = n + 1$.
The following result shows why we are interested to study frames in W_n.

Proposition 3.11 *Assume $F_1, \ldots, F_m$ are frames in W_n which cover W_n; that means, one has*

$$W_n \;=\; \bigcup_{i=1}^{m} F_i. \tag{3.14}$$

Then for every $C \subseteq W_n$ we have

$$h(C) \;\leq\; \frac{3}{2} \cdot m. \tag{3.15}$$

Proof. Let $y_1, \ldots, y_m$ denote the roots of $F_1, \ldots, F_m$, respectively. By symmetry, we may assume that there exists some t with $\frac{m}{2} \leq t \leq m$ such that $y_1, \ldots, y_t \in C$ as well as $y_{t+1}, \ldots, y_m \in W_n \backslash C$. (If $t < \frac{m}{2}$, Remark (iii) following Definition 2.3 shows that we may exchange the roles of C and $W_n \backslash C$.)
Put $C_i := C \cap F_i$ for $1 \leq i \leq m$.
We prove that $C_1, \ldots, C_t$ may be separated by one single hyperplane and that $C_{t+1}, \ldots, C_m$ may be separated by some centered image separation consisting of two hyperplanes. Finally, we shall apply Proposition 3.4.
For $1 \leq i \leq m$, write $y_i = (\varepsilon_{i1}, \ldots, \varepsilon_{in})$, and for $1 \leq j \leq n$ let y_{ij} denote the unique vertex in F_i which differs from y_i exactly in the j-th component. Put

$$J_i \;:=\; \{j \in \{1, \ldots, n\} \;:\; y_{ij} \in C_i\}, \qquad j_i \;:=\; |J_i|,$$

$$K_i \;:=\; \{1, \ldots, n\} \backslash J_i, \qquad k_i \;:=\; |K_i|.$$

Assume first that $1 \leq i \leq t$, that means $y_i = (\varepsilon_{i1}, \ldots, \varepsilon_{in}) \in C_i$. In this case, define the linear map $f_i : \mathbb{R}^n \to \mathbb{R}$ by

$$f_i(v_1, \ldots, v_n) \;:=\; \sum_{j \in J_i} \varepsilon_{ij} \cdot v_j + 3 \cdot \sum_{j \in K_i} \varepsilon_{ij} \cdot v_j,$$

and define the affine oriented hyperplane H_i in $\mathbb{R}^n$ by

$$\begin{aligned} H_i &:= \{v \in \mathbb{R}^n \ : \ f_i(v) = n + 2 \cdot k_i - 3\}, \\ H_i^+ &:= \{v \in \mathbb{R}^n \ : \ f_i(v) > n + 2 \cdot k_i - 3\}, \\ H_i^- &:= \mathbb{R}^n \backslash (H_i \cup H_i^+). \end{aligned}$$

Then one has

$$\begin{aligned} f_i(y_i) &= f_i(\varepsilon_{i1}, \ldots, \varepsilon_{in}) \ = \ n + 2 \cdot k_i, \\ f_i(y_{ij}) &= n + 2 \cdot k_i - 2 \qquad \text{for } j \in J_i, \\ f_i(w) &< n + 2 \cdot k_i - 3 \qquad \text{for } w \in W_n \backslash C_i. \end{aligned}$$

Thus, we have $C_i \subseteq H_i^+$ and $W_n \backslash C_i \subseteq H_i^-$. In particular, the single hyperplane arrangement (H_i) is a centered image separation of C_i (cf. Example 3.2 (i).)

Now, suppose $t < i \leq m$, that means $y_i = (\varepsilon_{i1}, \ldots, \varepsilon_{in}) \in W_n \backslash C_i$. In this case, define the linear map $f_i : \mathbb{R}^n \to \mathbb{R}$ by

$$f_i(v_1, \ldots, v_n) \ := \ 3 \cdot \sum_{j \in J_i} \varepsilon_{ij} \cdot v_j + \sum_{j \in K_i} \varepsilon_{ij} \cdot v_j,$$

and define the affine oriented hyperplane H_i in $\mathbb{R}^n$ by

$$\begin{aligned} H_i &:= \{v \in \mathbb{R}^n \ : \ f_i(v) = n + 2 \cdot j_i - 3\}, \\ H_i^+ &:= \{v \in \mathbb{R}^n \ : \ f_i(v) > n + 2 \cdot j_i - 3\}, \\ H_i^- &:= \mathbb{R}^n \backslash (H_i \cup H_i^+). \end{aligned}$$

Moreover, define the affine oriented hyperplane G_i in $\mathbb{R}^n$ by

$$\begin{aligned} G_i &:= \left\{ (v_1, \ldots, v_n) \in \mathbb{R}^n \ : \ \sum_{j=1}^{n} \varepsilon_{ij} \cdot v_j = n - 3 \right\}, \\ G_i^+ &:= \left\{ (v_1, \ldots, v_n) \in \mathbb{R}^n \ : \ \sum_{j=1}^{n} \varepsilon_{ij} \cdot v_j > n - 3 \right\}, \\ G_i^- &:= \mathbb{R}^n \backslash (G_i \cup G_i^+). \end{aligned}$$

Then for $t < i \leq m$ one has

$$\begin{aligned} C_i &\subseteq H_i^- \cap G_i^+, \\ F_i \backslash C_i &\subseteq H_i^+ \cap G_i^+, \\ W_n \backslash F_i &\subseteq H_i^- \cap G_i^-. \end{aligned}$$

Thus, Example 3.2 (ii) shows that (H_i, G_i) is a centered image separation of C_i.

Altogether, Proposition 3.4 shows that

$$\mathcal{H} \ := \ (H_1, \ldots, H_t, H_{t+1}, G_{t+1}, \ldots, H_m, G_m)$$

separates

$$C \ = \ \bigcup_{i=1}^{m} C_i.$$

Since we could assume $t \geq \frac{m}{2}$, we obtain

$$h(C) \ \leq \ t + 2 \cdot (m - t) \ = \ 2\,m - t \ \leq \ \frac{3}{2} \cdot m$$

as claimed. □

We still have the problem to cover W_n by certain frames $F_1, \ldots, F_m$ for some m as small as possible. Of course, there can exist a covering of pairwise disjoint frames only in case $n + 1 = 2^r$ for some $r \in \mathbb{N}$. In this case, arguments from the theory of Linear Codes show that there exist indeed $\frac{2^n}{n+1}$ frames which cover W_n. First, we recall the following

Proposition 3.12 *Assume $\mathbb{F}$ is a finite field with q Elements, suppose $n, k, r \in \mathbb{N}$ satisfy $n = k + r$, and presume $3 \leq d \leq n$. Then the following two conditions are equivalent:*
(i) There exists some k-dimensional subspace U of the vector space $\mathbb{F}^n$ such that all $v, v' \in U$ with $v \neq v'$ differ in at least d coordinates.
(ii) There exists some subset A of the vector space $\mathbb{F}^r$ with $|A| = n$ such that every subset I of A with $|I| = d - 1$ is linearly independent.

Proof. This is Satz 12.2 in [1]. □

Now, we identify – of course – the vertex set $W_n = \{1, -1\}^n$ with the vector space ${\mathbb{F}_2}^n$ in the obvious way, where $\mathbb{F}_2 = \{1, 0\}$ denotes the field with 2 elements.
We can now prove

Proposition 3.13 *Assume $n \geq 3$ satisfies $n + 1 = 2^r$ for some $r \in \mathbb{N}$. Then there exist $\frac{2^n}{n+1} = 2^{n-r}$ pairwise disjoint frames in W_n which constitute a covering of W_n.*

Proof. We apply Proposition 3.12 for $k = n - r$ and $d = 3$. Put $A := {\mathbb{F}_2}^r \backslash \{0\}$; then every subset of A consisting of 2 elements is linearly independent over $\mathbb{F}_2$. Since $|A| = 2^r - 1 = n$, Proposition 3.12, (ii) $\Rightarrow$ (i), shows that there exists some k-dimensional subspace U of ${\mathbb{F}_2}^n$ such that all $v, v' \in U$ with $v \neq v'$ differ in at least 3 coordinates. This means – and that is the decisive conclusion – that all

of those frames in ${\mathbb{F}_2}^n$ whose roots lie in U are pairwise disjoint.
Moreover, we have

$$|U| \;=\; 2^k \;=\; 2^{n-r} \;=\; \frac{2^n}{n+1},$$

and this proves what we want, namely, that there exist $\frac{2^n}{n+1}$ pairwise disjoint frames in W_n. Since all of these frames have exactly $n+1$ vertices, they must of course cover W_n. □

We still have to consider coverings of W_n by frames in case $n+1$ is not a power of 2. But then we make use of the following simple

Lemma 3.14 *Assume $F_1, \ldots, F_m$ are frames in W_n with*

$$W_n \;=\; \bigcup_{i=1}^{m} F_i.$$

Then there exist $2\,m$ frames in W_{n+1} covering W_{n+1}.

Proof. Let $y_1, \ldots, y_m$ denote the roots of the frames $F_1, \ldots, F_m$, respectively. Then the frames in W_{n+1} exhibiting the roots

$$(y_1, 1), \ldots, (y_m, 1), (y_1, -1), \ldots, (y_m, -1)$$

satisfy what we want. □

For $x \in \mathbb{R}$, let $[x]$ denote the Gaussian integer; that is the largest $k \in \mathbb{Z}$ satisfying $k \leq x$.
We can now prove

Proposition 3.15 *Assume $n \geq 1$. Then there exist*

$$f_n \;:=\; 2^{n-[\log_2(n+1)]}$$

frames in W_n which cover W_n. Moreover, one has

$$f_n \;\leq\; \frac{2^{n+1}}{n+2}. \tag{3.16}$$

Proof. For $n = 1$ and $n = 2$, the assertions are obvious, because in these special cases, there exists a covering of W_n consisting of n frames.
Now assume $n \geq 3$. The first assertion is clear by Proposition 3.13, if $n+1$ is a power of 2. If, on the other hand, $2^r < n+1 < 2^{r+1}$ holds for some $r \in \mathbb{N}$, the first assertion follows from Proposition 3.13 and a repeated application of Lemma 3.14 for the values $n' = 2^r - 1$, $n' = 2^r$, …, $n' = n-1$. Note that

$$[\log_2(n+1)] \;=\; r$$

does not depend on n as long as $2^r \leq n+1 < 2^{r+1}$.
To verify (3.16), assume again that $r \in \mathbb{N}$ satisfies $2^r \leq n+1 \leq 2^{r+1}-1$. Then we get $2^{-r} \leq \frac{2}{n+2}$ and thus

$$f_n \;=\; 2^{n-r} \;\leq\; \frac{2^{n+1}}{n+2}. \quad \square$$

Now we can summarize Proposition 3.11 and Proposition 3.15 and obtain directly the following result, which is rather better than Theorem 3.6 (ii).

Theorem 3.16 *For every $n \in \mathbb{N}$ and $C \subseteq W_n$ we have*

$$h(C) \;\leq\; 3 \cdot 2^{n-1-[\log_2(n+1)]} \;\leq\; \frac{3}{n+2} \cdot 2^n. \tag{3.17}$$

In particular, one has $h(C) = \mathcal{O}\left(\frac{2^n}{n}\right)$. $\square$

Note that – in general – the second bound in (3.17) is of course slightly worse than the first bound; however, the second bound is more manageable.

4 A Worst Case Lower Bound for $h(C)$

In the last sections, we have been mainly interested in upper bounds for $h(C)$, $C \subseteq W_n$; Theorem 3.16 shows that $h(C)$ grows at most exponentially with n. In this section, we want to derive some lower bound for the number

$$h_n := \max\{h(C) \;:\; C \subseteq W_n\}. \tag{4.1}$$

We shall see that h_n grows at least exponentially with n. To this end, we shall use arguments concerning numbering of unordered pairs $\{C, W_n \backslash C\}$ for $C \subseteq W_n$ such that C and $W_n \backslash C$ are linearly separable. First, we state the following

Definition 4.1 *For $n \geq 1$ let $t(n)$ denote the number of unordered partitions $\{C, C'\}$ of W_n for which C and C' are linearly separable.*

Remarks:
(i) Since $|W_n| = 2^n$, there exist

$$\frac{1}{2} \cdot 2^{|W_n|} = 2^{(2^n-1)}$$

unordered partitions of W_n into two sets.
(ii) The partition $\{\emptyset, W_n\}$ has to be considered while computing $t(n)$.

Example: Assume $n = 2$. There exist 8 unordered partitions of W_2 into two sets. By Example 2.4, only one of these partitions does not contribute to the computation of $t(2)$; thus we have $t(2) = 7$.

For general $n \in \mathbb{N}$ we want to obtain nontrivial upper bounds for $t(n)$. First we recall the following

Proposition 4.2 *Assume $k > n \geq 1$, and in $\mathbb{R}^n$ there are given k points $y_1, \ldots, y_k$ in general position; that means, every subset Y' of $Y = \{y_1, \ldots, y_k\}$ with $|Y'| = n + 1$ is affinely independent. Let $s(n, k)$ denote the number of unordered partitions $\{Y_1, Y_2\}$ of Y such that Y_1 and Y_2 are linearly separable. Then one has*

$$s(n, k) = \sum_{j=0}^{n} \binom{k-1}{j}. \tag{4.2}$$

Proof. This result is shown in [14]. $\square$

Certainly, the vertices of W_n are far from being in general position; however, the next result relates the numbers $t(n)$ and $s(n, 2^n)$.

Proposition 4.3 *For every $n \in \mathbb{N}$ one has*

$$t(n) \leq s(n, 2^n). \tag{4.3}$$

Proof. Assume $H_1, \ldots, H_{t(n)}$ are affine hyperplanes in $\mathbb{R}^n$ which do not intersect W_n and such that any two distinct H_i, H_j, $1 \leq i < j \leq t(n)$, induce distinct unordered partitions of W_n. For any $x \in W_n$ we choose some open set U_x in $\mathbb{R}^n$ with $x \in U_x$ such that $U_x \cap H_i = \emptyset$ holds for all i with $1 \leq i \leq t(n)$ and $U_x \cap U_{x'} = \emptyset$ holds for all $x, x' \in W_n$ with $x \neq x'$. Now, for any set U_x, $x \in W_n$, we choose some $y(x) \in U_x$ such that the points $y(x)$, $x \in W_n$, are in general position. By our choice of the sets U_x, the affine hyperplanes $H_1, \ldots, H_{t(n)}$ induce $t(n)$ distinct unordered partitions of the set $Y := \{y(x) \,:\, x \in W_n\}$; this yields what we want. $\square$

Proposition 4.2 and Proposition 4.3 will yield an upper bound for $t(n)$. First, we prove

Lemma 4.4 *Assume $m, k \in \mathbb{N}$ satisfy $3m \leq k$. Then one has*

$$\sum_{j=0}^{m-1} \binom{k-1}{j} \leq \binom{k-1}{m}. \tag{4.4}$$

Proof. For fixed $k \in \mathbb{N}$, we proceed by induction on m. In case $m = 1$ we have $3 \leq k$ by the assumption of the lemma, and (4.4) states the even weaker inequality $1 \leq k - 1$.
Now assume $2 \leq m \leq \frac{k}{3}$, and we have already proved that

$$\sum_{j=0}^{m-2} \binom{k-1}{j} \leq \binom{k-1}{m-1}.$$

Then we get in view of $2m \leq k - m$:

$$\sum_{j=0}^{m-1} \binom{k-1}{j} \leq 2 \cdot \binom{k-1}{m-1} = \frac{2m}{k-m} \cdot \binom{k-1}{m} \leq \binom{k-1}{m}. \quad \Box$$

Now we obtain the following

Proposition 4.5 *For all $n \in \mathbb{N}$ with $n \geq 2$ we have*

$$t(n) \leq 2 \cdot \binom{2^n - 1}{n} + 1. \tag{4.5}$$

Proof. For $n = 2$ we have $t(2) = 7 = 2 \cdot \binom{3}{2} + 1$. For $n = 3$ we get by Proposition 4.2 and Proposition 4.3

$$t(3) \ \leq \ s(3,8) \ = \ \sum_{j=0}^{3} \binom{7}{j} \ = \ 1 + 7 + 21 + 35 \ = \ 64 \ < \ 71 \ = \ 2 \cdot \binom{7}{3} + 1.$$

For $n \geq 4$ we have $3n \leq 2^n$, and thus Proposition 4.2, Proposition 4.3 and Lemma 4.4 yield with $m = n$ and $k = 2^n$:

$$t(n) \ \leq \ \sum_{j=0}^{n} \binom{2^n - 1}{j} \ = \ \sum_{j=0}^{n-1} \binom{2^n - 1}{j} + \binom{2^n - 1}{n} \ \leq \ 2 \cdot \binom{2^n - 1}{n}$$

as claimed. $\Box$

Now we are able to prove the following main result of this section.

Theorem 4.6 *Assume $n \geq 2$, and choose $l = l_n \in \mathbb{N}$ such that* every *subset $C \subseteq W_n$ may be separated by some l'-arrangement for an appropriate number $l' \leq l$. Then one has*

$$l \ \geq \ -\frac{n^2}{2} + \sqrt{\frac{n^4}{4} + 2^n} \ > \ 2^{\frac{n}{2}} - \frac{n^2}{2}. \tag{4.6}$$

In other words, we have

$$h_n \ \geq \ -\frac{n^2}{2} + \sqrt{\frac{n^4}{4} + 2^n} \ > \ 2^{\frac{n}{2}} - \frac{n^2}{2}. \tag{4.7}$$

Proof. In view of (2.5) and (4.1), the inequality (4.7) is of course only a reformulation of the first assertion. To prove (4.6), we first note that Lemma 2.5 (iv) implies that every $C \subseteq W_n$ may be separated by some l-arrangement.
Let $\mathcal{H}_0$ denote some set of affine hyperplanes in $\mathbb{R}^n$ with $|\mathcal{H}_0| = t(n)$ and not intersecting W_n such that these $t(n)$ affine hyperplanes induce exactly the $t(n)$ distinct unordered partitions of W_n into two linearly separable sets.

Now, any of all the $2^{(2^n-1)}$ unordered partitions $\{C, C'\}$ of W_n is uniquely determined by (at least) some l-arrangement $\mathcal{H} = (H_1, \ldots, H_l)$ with $H_1, \ldots, H_l \in \mathcal{H}_0$ and some affine oriented hyperplane H in $\mathbb{R}^l$ which linearly separates $\varphi(\mathcal{H}, C)$ and $\varphi(\mathcal{H}, C')$. There exist $2^l \cdot (t(n))^l$ l-arrangements consisting of l oriented affine hyperplanes in $\mathcal{H}_0$; the factor 2^l arises from the orientations. Thus we get

$$2^l \cdot (t(n))^l \cdot t(l) \;\geq\; 2^{(2^n-1)}. \tag{4.8}$$

Note that the affine oriented hyperplane H in $\mathbb{R}^l$ causes the factor $t(l)$ instead of $2 \cdot t(l)$, because we consider unordered partitions $\{C, C'\}$ of W_n.

By the assumption of the theorem, we have $n \geq 2$ and thus also $l \geq 2$. Therefore, Proposition 4.5 and (4.8) yield

$$2^l \cdot \left(2 \cdot \binom{2^n-1}{n} + 1\right)^l \cdot \left(2 \cdot \binom{2^l-1}{l} + 1\right) \;\geq\; 2^{(2^n-1)}. \tag{4.9}$$

Furthermore, for $m \geq 2$ we have

$$2 \cdot \binom{2^m-1}{m} + 1 \;\leq\; 2^{(m^2-1)}. \tag{4.10}$$

This inequality is clear for $m = 2$, while for $m \geq 3$ we get

$$2 \cdot \binom{2^m-1}{m} + 1 \;\leq\; 2 \cdot \frac{(2^m)^m}{m!} \;<\; 2^{(m^2-1)}.$$

Now (4.9) and (4.10), applied to $m = n$ and $m = l$, yield

$$2^l \cdot 2^{(n^2-1)\cdot l} \cdot 2^{(l^2-1)} \;\geq\; 2^{(2^n-1)}.$$

Simplification of this inequality yields

$$2^{n^2 \cdot l + l^2} \;\geq\; 2^{(2^n)};$$

that is

$$l^2 + n^2 \cdot l - 2^n \;\geq\; 0$$

and thus

$$l \;\geq\; -\frac{n^2}{2} + \sqrt{\frac{n^4}{4} + 2^n} \;>\; 2^{\frac{n}{2}} - \frac{n^2}{2}$$

as claimed. $\square$

Note that the inequality (4.7) is trivial for $n = 1$. Thus, by summarizing Theorem 3.16 and Theorem 4.6 we obtain

Theorem 4.7 *For every $n \in \mathbb{N}$ one has*

$$2^{\frac{n}{2}} - \frac{n^2}{2} < -\frac{n^2}{2} + \sqrt{\frac{n^4}{4} + 2^n} \leq h_n \leq \frac{3}{n+2} \cdot 2^n. \quad \square \tag{4.11}$$

The two left terms in (4.11) are almost equal for large n, and differ considerably only for small n. Although the bounds of h_n specified in (4.11) differ quantitatively in some essential manner, we see yet that h_n grows exponentially with n.

5 Conclusions

With respect to a theory of feedforward networks the derived results, as stated in Theorem 4.7, are understood as a first step in a program which tries to make use of geometric techniques to solve open problems in this context. Here we addressed the problem of determining the minimal number of hidden neurons of a feedforward network, which should be able to solve any given binary classification problem for n inputs, i.e. to realize any Boolean function on n inputs. The derived upper bound (4.11), although it is better than the weaker bound 2^{n-1} or other known results reported in the literature, is still too high to be of practical relevance for real world applications of these networks. In fact, it is well known that many problems can be solved with much less neurons; for instance, the parity problem (Problem 2.7) for n inputs can always be solved with n hidden neurons. On the other hand, Theorem 4.7 states, that for a given n there always exists a class of binary classification problems for which a solution needs more than $(2^{\frac{n}{2}} - \frac{n^2}{2})$ hidden neurons. Of course, this lower bound gets effective only for large n. Thus, its main use is for asymptotic considerations. But, since h_n must grow exponentially with n, it also provides the discouraging insight that a large class of Boolean problems needs also very large networks for a solution.

From the viewpoint of these results the following questions may be of relevance: One may classify the problems according to the minimal number of hyperplanes a solution has to use. Although it might be difficult to decide, in which class a given problem has to be located, the cardinality of these classes is of interest. For n large, are most of the problems “trivial” in the sense that the minimal number of hyperplanes a solution needs is much less than the lower bound (4.11) for h_n? Or are most problems “complex” in the sense that the minimal number of hyperplanes a solution needs is larger than this lower bound?

Furthermore, many interesting problems, represented by a vertex set C, inherit a symmetry property like, for instance, the parity problem (Problem 2.7). Lower

and upper bounds of $h(C)$ of course will depend on this symmetry and might be effectively reduced for a known symmetry of the problem. The combination of the geometric techniques used in this paper with group theoretical aspects of binary classification problems will lead to more specific and much stronger results.
The strength of feedforward networks is their ability to "learn"; i.e. there exists a potential function and a gradient descend algorithm, called backpropagation, which, under certain conditions, is able to find solutions for a given problem [11]. These networks have to use smooth transfer functions instead of the step functions referred to in this paper. Our results also apply to these type of networks because, as outlined in the introduction, the hyperplanes used in our arguments still can be identified with the centers of graded neurons. On the other hand, there exists a conjecture, that for networks using sigmoidal (S-shaped) transfer functions the lower bounds for h_n should be further reducible.

References

[1] M. Aigner, *Diskrete Mathematik*, Vieweg, Berlin (1993).

[2] M. Arai, Bounds on the number of hidden units in binary-valued three-layer neural networks, *Neural Networks*, **6**, (1993), 855–860.

[3] A. R. Barron, Approximation and estimation bounds for artificial neural networks, in: Proceedings of the Fourth Annual Workshop on Computational Learning, University of California, Santa Cruz, Morgan Kaufman, San Mateo (1991), pp. 243-249.

[4] G. Cybenko, Approximation by superposition of a sigmoidal function, *Mathematics of Control, Signals, and Systems* **2**, (1989), 303–314.

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

[6] R. Hecht-Nielsen, Theory of the Backpropagation Neural Network, in: Proc. 1989 International Joint Conference on Neural Networks, Vol.I, IEEE Press, New York (1989), pp. 593–605.

[7] J. A. Hertz, A. Krogh, and R. G. Palmer, *Introduction to the Theory of Neural Computation*, Addison-Wesley, Redwood City (1991).

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

[9] K. Hornik, Approximation capabilities of multilayer feedforward networks, *Neural Networks*, **4**, (1991), 251–257.

[10] S. C. Huang, and Y. F. Huang, Bounds on the number of hidden neurons in multilayer perceptrons, *IEEE Transactions on Neural Networks*, **2**, (1991), 47–55.

[11] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, Learning internal representation by error propagation, in: D. E. Rumelhart, and J. L. McClelland (eds.), *Parallel Distributed Processing: Explorations in the Microstructures of Cognition*, Vol. 1, MIT Press, Cambridge M.A. (1986), pp. 318–362.

[12] M. Stinchcombe, and H. White, Universal approximation using feedforward networks with non-sigmoid hiddenlayer activation function, *Proceedings of the International Joint Conference on Neural Networks*, IEEE Press, New York (1989) pp. I.613–617.

[13] Z. Wang, C. di Massimo, M. T. Tham, and A. J. Morris, A procedure for determining the topology of multilayer feedforward neural networks, *Neural Networks*, **7**, (1994), 291–300.

[14] Zaslavsky, T. *Facing up to Arrangements: Face-Count Formulas for Partitions of Space by Hyperplanes*, American Mathematical Society, Providence (1975).