

The classification capabilities of exact two-layered perceptrons

Citation for published version (APA):

Zwietering, P. J., Aarts, E. H. L., & Wessels, J. (1991). *The classification capabilities of exact two-layered perceptrons*. (Memorandum COSOR; Vol. 9109). Technische Universiteit Eindhoven.

Document status and date:

Published: 01/01/1991

Document Version:

Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.tue.nl/taverne

Take down policy

If you believe that this document breaches copyright please contact us at:

openaccess@tue.nl

providing details and we will investigate your claim.

EINDHOVEN UNIVERSITY OF TECHNOLOGY
Department of Mathematics and Computing Science

Memorandum COSOR 91-09

The Classification Capabilities Of
Exact Two-Layered Perceptrons

P.J. Zwietering
E.H.L. Aarts
J. Wessels

Eindhoven University of Technology
Department of Mathematics and Computing Science
P.O. Box 513
5600 MB Eindhoven
The Netherlands

ISSN: 0926-4493

Eindhoven, May 1991
The Netherlands

The Classification Capabilities Of Exact Two-Layered Perceptrons

P.J. Zwietering¹, E.H.L. Aarts^{1,2} and J. Wessels¹

1. Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, the Netherlands
2. Philips Research Laboratories, P.O. Box 80.000, 5600 JA Eindhoven, the Netherlands

Abstract

The present paper studies the problem of finding a two-layered perceptron that exactly classifies a given subset. Such a two-layered perceptron is called exact with respect to the given subset. We derive both a necessary and a sufficient condition for a given subset to be classifiable by an exact two-layered perceptron. The necessary condition can be viewed as a generalization of the linear-separability condition of the one-layered perceptron and confirms the conjecture that the capabilities of exact two-layered perceptrons are more limited than those of exact three-layered perceptrons. The sufficient condition shows that the capabilities of exact two-layered perceptrons extend beyond the exact classification of convex subsets. Furthermore, we present a systematic verification method for the given sufficient condition.

Keywords: Classification, Multi-Layered Perceptrons, Size Hidden Layer

1 Introduction

In [13], we studied the design and complexity of multi-layered perceptrons for exactly solving a given problem. We focussed on the class of combinatorial optimization problems and showed that such problems can be solved exactly by three-layered perceptrons if certain linearity constraints are satisfied. To this end a given combinatorial optimization problem was reduced to the problem of exactly classifying a given subset by a multi-layered perceptron. We derived a necessary condition for a given subset to be exactly classifiable by a multi-layered perceptron and proved that this condition is sufficient for exactly classifying the subset by a three-layered perceptron. The question whether this condition is sufficient for exactly classifying the subset by a two-layered perceptron was left as an open problem. In this paper we therefore examine the exact classification capabilities of two-layered perceptrons.

Why are we interested in the exact solving of combinatorial optimization problems by a multi-layered perceptron? The answer is that we are not. However, the idea is that *if* we have the network configuration of a multi-layered perceptron that is capable of exactly

solving a given problem *provided* it has the right weights, a good approximation of these weights can be found using one of the many existing learning algorithms, such that we obtain a multi-layered perceptron that approximately solves the problem. Moreover, our hope is that an exact network configuration for a given problem is capable of learning to solve problems that are close in some sense to the original problem. On the other hand, if we know that for instance a two-layered perceptron cannot exactly solve a certain problem, whatever the number of hidden nodes and the values of the weights are, trying to learn a two-layered perceptron with an arbitrary chosen number of hidden nodes to solve the problem is doomed to fail. An example of such a problem is the sorting problem, which can be proven not to be exactly solvable by a two-layered perceptron using the results of this paper (see [14]).

The next question that arises is why to bother about two-layered perceptrons at all, since the capabilities of exact two-layered perceptrons will not exceed the capabilities of exact three-layered perceptrons. The main reason for studying the classification capabilities of exact two-layered perceptrons is that these networks are expected to learn faster than three-layered perceptrons, provided that an exact two-layered perceptron for the problem exists and one uses a sufficient number of hidden nodes. Furthermore, the final weight-vector of a two-layered perceptron that has been obtained by some learning algorithm is analyzed more easily. These are probably also the reasons why two-layered perceptrons are most frequently used. Finally, studying the classification capabilities of exact two-layered perceptrons is of theoretical interest.

We use the term exact two-layered perceptron (in general exact multi-layered perceptron) to emphasize the fact that the two-layered perceptron exactly classifies a given subset. As in [13] we restrict ourselves to perceptrons using the hard limiting response function. So far, there have been few reports on the classification capabilities of exact two-layered perceptrons using the hard limiting response function. In his introductory paper, Lippmann presents a brief summary of the classification capabilities of exact multi-layered perceptrons [8]. However, Lippmann incorrectly states that a subset has to be *convex* for being classifiable by an exact two-layered perceptron. Since then, several authors have pointed out that the capabilities of exact two-layered perceptrons extend beyond the exact classification of convex subsets. Wieland and Leighton [12], Huang and Lippmann [6] and Makhoul et al. [9] demonstrate this by some hand crafted examples of non-convex subsets that can be exactly classified by a two-layered perceptron. We extend these results by presenting formal and more systematical results on the classification capabilities of exact two-layered perceptrons.

More recent papers consider the approximate classification capabilities of two-layered perceptrons that use a sigmoidal response function. Cybenko [2], Funahashi [4] and Hornik et al. [5] show that two-layered perceptrons are capable of classifying a given subset within arbitrary precision, but they do not give much insight in the required number of hidden nodes. A first attempt to solve this problem is presented in [3], where an upper-bound on the required number of hidden nodes is derived. Our approach is directed towards finding the minimal number of required hidden nodes for exactly classifying a given subset.

Obviously, this number is also a lower-bound for approximately classifying a given subset if the error goes to zero. Note that we do not consider the results by Baum [1] since he discusses finite subsets only.

The remainder of the paper is organized as follows. In Section 2 we briefly summarize the most important elements of the study presented in [13]. Section 3 contains the main results of this paper in three parts: Section 3.1 presents a necessary condition for a subset to be classifiable by an exact two-layered perceptron. Section 3.2 presents a sufficient condition for a subset to be classifiable by an exact two-layered perceptron and describes how the weights of such an exact two-layered perceptron can be calculated if the subset satisfies this condition. An algorithm for a systematic verification of the sufficient condition is presented in Section 3.3. The paper ends with some concluding remarks and references. The proofs of the lemmas and theorems presented in Section 3 are given in the Appendix.

2 Preliminaries

In this paper we consider the standard architecture of a multi-layered perceptron with m layers (m -LP) consisting of one output layer and $m - 1$ hidden layers (see [8, 11, 13]). The output of a node is the result of a computation determined by a summation of a bias and the weighted inputs of that node, which is then passed through a non-linear response function θ . In this paper we only consider the hard limiting response function Θ that satisfies $\Theta(\lambda) = 1$ if $\lambda \geq 0$ and $\Theta(\lambda) = 0$ if $\lambda < 0$. Throughout the paper we consider the classification of (non-finite) subsets of \mathbb{R}^N for some fixed $N \in \mathbb{N}$, which implies that we have to consider 2-LPs with N inputs only.

We let $R_{m,N,K}$ denote the set of all vector functions from \mathbb{R}^N to $\{0, 1\}^K$ that can be formed by constructing an m -LP with N inputs and K outputs. We define $R_{m,N,K}$ iteratively, first $R_{1,N,K}$:

$$R_{1,N,K} = \{f: \mathbb{R}^N \rightarrow \{0, 1\}^K \mid f_i = \Theta \circ \tilde{f}_i, \tilde{f}_i \in \mathcal{A}_N, i = 1, \dots, K\}, \quad (1)$$

where \mathcal{A}_N denotes the set of all affine functions from \mathbb{R}^N to \mathbb{R} :

$$\mathcal{A}_N = \{\tilde{f}: \mathbb{R}^N \rightarrow \mathbb{R} \mid \tilde{f}(x) = a \cdot x + b, x \in \mathbb{R}^N, a \in \mathbb{R}^N, b \in \mathbb{R}\}.$$

Next, since an $(m + 1)$ -LP can be constructed by putting a 1-LP on top of an m -LP, we define $R_{m+1,N,K}$ in terms of $R_{1,L,K}$ and $R_{m,N,L}$, where L denotes the number of hidden nodes in the m^{th} hidden layer:

$$R_{m+1,N,K} = \{f: \mathbb{R}^N \rightarrow \{0, 1\}^K \mid f = g \circ h, g \in R_{1,L,K}, h \in R_{m,N,L}, L \in \mathbb{N}\}. \quad (2)$$

Abbreviate $R_{m,N,1}$ as R_m , then V can be classified by an exact m -LP if there exists an $f \in R_m$ such that $f(x) = 1$ if and only if $x \in V$. Define for each function $f: \mathbb{R}^N \rightarrow \{0, 1\}$ the set $\mathcal{J}(f) \subseteq \mathbb{R}^N$ by $\mathcal{J}(f) = \{x \in \mathbb{R}^N \mid f(x) = 1\}$, then it follows that V can be classified by an m -LP if $V = \mathcal{J}(f)$ for some $f \in R_m$. The collection of subsets of \mathbb{R}^N

that can be classified by an exact m -LP is denoted with C_m . It is clear from the above that $C_m = \{\mathcal{J}(f) \mid f \in R_m\}$. Furthermore, one can easily verify the well-known fact that $C_1 = \{V \subseteq \mathbb{R}^N \mid V \text{ is a closed linear half-space}\}$, which implies that $V \in C_1 \Rightarrow V^* \notin C_1$ (V^* denotes the complement of the set V). Finally, we define the following four collections of subsets:

$$\tilde{C}_1 = \{V \subseteq \mathbb{R}^N \mid V \in C_1 \vee V^* \in C_1\}, \quad (3)$$

$$P = \{V \subseteq \mathbb{R}^N \mid V = \bigcap_{i=1}^k V_i, V_i \in C_1, k \in \mathbb{N}\}, \quad (4)$$

$$\tilde{P} = \{V \subseteq \mathbb{R}^N \mid V = \bigcap_{i=1}^k V_i, V_i \in \tilde{C}_1, k \in \mathbb{N}\}, \quad (5)$$

$$U = \{V \subseteq \mathbb{R}^N \mid V = \bigcup_{i=1}^l V_i, V_i \in \tilde{P}, l \in \mathbb{N}\}. \quad (6)$$

From the above it follows that \tilde{C}_1 is the collection of open and closed linear half-spaces and, hence, P is the collection of polyhedra. The collection \tilde{P} can be viewed as the collection of all *pseudo*-polyhedra. A genuine polyhedron has all faces belonging to the set, whereas a pseudo-polyhedron can have faces either belonging to the set or to the complement of the set. Finally, the collection U is the collection of all subsets of \mathbb{R}^N that can be represented as a union of a finite number of pseudo-polyhedra. U is the collection of subsets that have piece-wise linear bounds. Note that $P \subseteq \tilde{P}$ and $\tilde{C}_1 \subseteq \tilde{P} \subseteq U$. The following basic results have been proven in [13].

Lemma 1 *If $m \geq 2$ then $V \in C_m$ implies that $V^* \in C_m$.*

Lemma 2 *Let $\{V_i \mid i = 1, \dots, l\}$ be a collection of subsets with $V_i \in C_m$ or $V_i^* \in C_m$ for all i , then $\bigcap_{i=1}^l V_i \in C_{m+1}$.*

Theorem 1 *Let $V \subseteq \mathbb{R}^N$, then V can be classified by an exact m -LP if $V \in U$ and $m \geq 3$.*

Theorem 2 *Let $V \subseteq \mathbb{R}^N$, then V can be classified by an exact m -LP only if $V \in U$.*

3 Main results

3.1 A necessary condition for the existence of an exact 2-LP

In this section we demonstrate that the condition given in Theorem 2 is not sufficient for classifying a given subset by an exact 2-LP, i.e. there exist subsets of \mathbb{R}^N that can be classified by an exact 3-LP but cannot be classified by an exact 2-LP, which implies that C_2 is a true subset of C_3 . We show this by proving the necessity of a second condition for classifying a given subset by a 2-LP. This condition requires the existence of a hyperplane and two spheres.

The sphere $B_1(x_0, \delta)$ with centre $x_0 \in \mathbb{R}^N$ and radius $\delta > 0$ is the set $\{x \in \mathbb{R}^N \mid \|x - x_0\| < \delta\}$. Furthermore, we use W° to denote the interior of a subset W , that is defined as the

set of all points x for which a sphere B exists that satisfies $x \in B \subset W$. The necessary condition for a subset to be classifiable by an exact 2-LP can now be expressed as follows.

Theorem 3 *Let $V \in U$ be a subset for which there exist a $W \in C_1$ and two spheres B_1, B_2 such that*

$$\left. \begin{array}{l} \emptyset \neq B_1 \cap W^\circ \subseteq V \\ \emptyset \neq B_1 \cap W^* \subseteq V^* \end{array} \right\} \wedge \left\{ \begin{array}{l} \emptyset \neq B_2 \cap W^\circ \subseteq V^* \\ \emptyset \neq B_2 \cap W^* \subseteq V \end{array} \right. \quad (7)$$

then $V \notin C_2$.

In the above theorem the subset W corresponds to a closed linear half-space, i.e. $W = \{x \in \mathbb{R}^N \mid a \cdot x + b \geq 0\}$ for some $a \in \mathbb{R}^N$ and $b \in \mathbb{R}$. The conditions in (7) do not specify whether the hyperplane $W = \{x \mid a \cdot x + b = 0\}$ (or parts of it) belongs to V or V^* . The conditions are only concerned with parts of the open linear half-spaces $W^\circ = \{x \mid a \cdot x + b > 0\}$ and $W^* = \{x \mid a \cdot x + b < 0\}$.

Suppose that $V \in \mathbb{R}^N$ satisfies the conditions of Theorem 3 and $f: \mathbb{R}^N \rightarrow \{0, 1\}$ satisfies $\mathcal{J}(f) = V$, then $f(x) = 1$ for all $x \in B_1 \cap W^\circ$, $x \in B_2 \cap W^*$ and $f(x) = 0$ for all $x \in B_1 \cap W^*$, $x \in B_2 \cap W^\circ$. Thus f solves some kind of extended exclusive-or problem. Theorem 3 proves that $f \notin C_2$. Hence, the condition (7) can be viewed as a generalization of the condition of linear separability for a subset to be classifiable by a 1-LP, since this condition is responsible for the non-existence of a 1-LP for the exclusive-or problem.

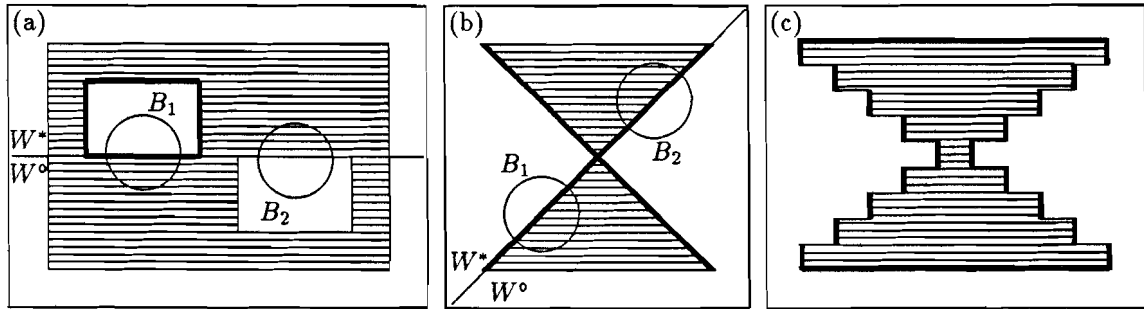


Figure 1: The subsets of \mathbb{R}^2 given in (a) and (b) cannot be classified by an exact 2-LP since they satisfy the conditions of Theorem 3; the circles correspond to the spheres B_1 and B_2 . The subset presented in (c) can be classified by an exact 2-LP since it satisfies the conditions of Corollary 1; see Section 3.2 and 3.3. Note that solid boundary lines do and thin boundary lines do not belong to the presented sets.

As already mentioned in the introduction, the results of Cybenko [2] and others show that a subset $V \subseteq \mathbb{R}^N$ can be approximately classified by a 2-LP with arbitrary precision. In our context this implies that for all $\varepsilon > 0$ there exists a $V_\varepsilon \in C_2$ such that $\|V - V_\varepsilon\| < \varepsilon$. Let $V_\varepsilon = \mathcal{J}(g_\varepsilon \circ h_\varepsilon)$ for some $g_\varepsilon \in R_{1,L_\varepsilon,1}$, $h_\varepsilon \in R_{1,N,L_\varepsilon}$ and $L_\varepsilon \in \mathbb{N}$, the latter denoting the number of hidden nodes. If $V \in U$ satisfies the conditions of Theorem 3 then $V \notin C_2$ and it follows that we must have $\lim_{\varepsilon \downarrow 0} L_\varepsilon = \infty$. For the two subsets in Figure 1a and 1b we have found approximating subsets $V_\varepsilon \in C_2$ with $L_\varepsilon = \mathcal{O}(|\log \varepsilon|)$ and $L_\varepsilon = \mathcal{O}(1/\varepsilon)$, respectively. In Figure 1c an example is given of a subset in C_2 that approximates the

subset given in Figure 1b (see the following sections for a proof that this set belongs to C_2). The general upper-bound $L_\varepsilon = \mathcal{O}(\varepsilon^{-(N-1)/2})$ given by Cybenko [3] indicates that the result for the subset in Figure 1b can be improved.

3.2 A sufficient condition for the existence of an exact 2-LP

The examples given in the previous section show that $C_2 \subset U$ (strict inclusion). Furthermore, from Lemma 2 it follows that $\tilde{P} \subseteq C_2$, which implies that every piece-wise linear and convex set can be classified by an exact 2-LP, a result already known for some time [8]. To the best of our knowledge the literature presents only the few examples of subsets in $C_2 \setminus P$ [6, 9, 12]. One such example that is for the first time reported in [9] is shown in Figure 1c and can be proven to belong to C_2 by using Corollary 1 below (see also the next section). This corollary presents a sufficient (but not necessary) condition for a subset to be classifiable by an exact 2-LP. It is the most general sufficient condition we found so far, based on the classifiability of intersections of two classifiable subsets by an exact 2-LP. Moreover, it is the only sufficient condition for which we have found a systematic verification method. This verification method is presented in the next section.

The basic result that we use to obtain Corollary 1 below is that $V_1 \setminus V_2 \in C_2$ for all $V_1 \in \tilde{P}$ and $V_2 \in C_2$. Before we prove this result we need some preliminaries.

Let $V \in \tilde{P}$, then $V = \bigcap_{i=1}^k V_i$, for some $k \in \mathbb{N}$ and $V_1, \dots, V_k \in \tilde{C}_1$. Let $0 \leq r \leq k$ be such that $V_i \in C_1$ for $i = 1, \dots, r$ and $V_i^* \in C_1$ for $i = r+1, \dots, k$. Furthermore, let $h \in R_{1,N,k}$ and $g = \Theta \circ \tilde{g} \in R_{1,k,1}$ be defined by $V_i = \mathcal{J}(h_i)$, $i = 1, \dots, r$, $V_i = \mathcal{J}^*(h_i)$, $i = r+1, \dots, k$ and $\tilde{g}(x) = \sum_{i=1}^r x_i - \sum_{i=r+1}^k x_i - r$, ($x \in \mathbb{R}^k$), respectively. Then one can easily verify that $V = \mathcal{J}(g \circ h)$ and $\tilde{g}(h(x)) \in \{-k, -k+1, \dots, 0\}$ for all $x \in \mathbb{R}^N$.

If $V \in C_2$ then $V = \mathcal{J}(g \circ h)$ for some $g = \Theta \circ \tilde{g} \in R_{1,k,1}$, $h \in R_{1,N,k}$ and $k \in \mathbb{N}$. Analogous as above we can assume that $\tilde{g}(h(x)) \in \{\alpha, \alpha+1, \dots, \beta\}$ for some $\alpha, \beta \in \mathbb{Z}$, $\alpha \leq \beta$ and all $x \in \mathbb{R}^N$. Proving that this representation can always be attained is an easy exercise. Without loss of generality we may further assume that $\alpha < 0 \leq \beta$, such that we arrive at the conditions of the following theorem.

Theorem 4 *Let $V_1 \in \tilde{P}$, $V_2 \in C_2$ and assume $V_i = \mathcal{J}(g_i \circ h_i)$, $g_i = \Theta \circ \tilde{g}_i \in R_{1,k_i,1}$, $h_i \in R_{1,N,k_i}$, $\tilde{g}_1(h_1(x)) \in \{-k_1, -k_1+1, \dots, 0\}$ and $\tilde{g}_2(h_2(x)) \in \{\alpha, \alpha+1, \dots, \beta\}$ for some $k_1, k_2 \in \mathbb{N}$, $\alpha, \beta \in \mathbb{Z}$, $\alpha < 0 \leq \beta$ and for all $x \in \mathbb{R}^N$.*

Then $V_1 \setminus V_2 = \mathcal{J}(g \circ h)$, with $h \in R_{1,N,k_1+k_2}$ and $g = \Theta \circ \tilde{g} \in R_{1,k_1+k_2,1}$ given by $h(x) = (h_1(x), h_2(x))$ and

$$\tilde{g}(x_1, x_2) = -\alpha \tilde{g}_1(x_1) - \tilde{g}_2(x_2) - 1, \quad (8)$$

respectively. Furthermore $\tilde{g}(h(x)) \in \{k_1\alpha - \beta - 1, \dots, -\alpha - 1\}$ for all $x \in \mathbb{R}^N$.

In the conditions of the above theorem it is essential that $\tilde{g}_1(h_1(x))$ has only one positive value and that $\tilde{g}_2(h_2(x))$ is bounded away from zero. The results of Theorem 4 can be

divided into two parts. The fundamental result is that $V_1 \setminus V_2 \in C_2$ for all $V_1 \in \tilde{P}$ and $V_2 \in C_2$, the computational result indicates how $f \in R_2$ is calculated such that $\mathcal{J}(f) = V_1 \setminus V_2$. These results are easily generalized to the results of Corollary 1 and Corollary 2, respectively. Note that we use $V_1 \setminus V_2 \setminus \dots \setminus V_l$ as a shorthand for $V_1 \setminus (V_2 \setminus (\dots (V_{l-1} \setminus V_l) \dots))$.

Corollary 1 *Let $V = V_1 \setminus V_2 \setminus \dots \setminus V_l$ for some $l \in \mathbb{N}$ and $V_1, V_2, \dots, V_l \in \tilde{P}$, then $V \in C_2$.*

Corollary 2 *Let $l \in \mathbb{N}$, $V_i \in \tilde{P}$, $i = 1, \dots, l$, and assume $V_i = \mathcal{J}(g_i \circ h_i)$, $g_i = \Theta \circ \tilde{g}_i \in R_{1, k_i, 1}$, $h_i \in R_{1, N, k_i}$ and $\tilde{g}_i(h_i(x)) \in \{-k_i, -k_i + 1, \dots, 0\}$ for some $k_i \in \mathbb{N}$ and all $x \in \mathbb{R}^N$.*

Then $V_1 \setminus V_2 \setminus \dots \setminus V_l = \mathcal{J}(g \circ h)$, with $h \in R_{1, N, \Sigma k_i}$ and $g = \Theta \circ \tilde{g} \in R_{1, \Sigma k_i, 1}$ given by $h(x) = (h_1(x), \dots, h_l(x))$ and

$$\tilde{g}(x_1, \dots, x_l) = \sum_{i=1}^l (-1)^i \alpha_{l-i} \tilde{g}_i(x_i) - \frac{(-1)^l + 1}{2}, \quad (9)$$

respectively. Furthermore, $\tilde{g}(h(x)) \in \{\alpha_l, \dots, -\alpha_{l-1} - 1\}$ for all $x \in \mathbb{R}^N$. The numbers α_i , $i = 0, 1, \dots, l$, are recursively defined by $\alpha_0 = -1$ and $\alpha_{i+1} = k_{l-i} \alpha_i + \alpha_{i-1}$ for $i = 0, \dots, l-1$ ($\alpha_{-1} = 0$).

Note that the above corollary also holds for $l = 1$, which serves as a starting point for a straightforward proof using mathematical induction and Theorem 4. From Corollary 2 it follows that the required number of hidden units of an exact 2-LP that classifies $V_1 \setminus V_2 \setminus \dots \setminus V_l$ is at most $\sum_{i=1}^l k_i$, which equals the total number of half-spaces defining the subsets V_1, \dots, V_l . If these subsets have defining half-spaces in common the required number of hidden units can be reduced accordingly. Although Corollary 2 gives an algorithm for the determination of a set of weights for an exact 2-LP, the practical value of this algorithm is limited since these weights can become very large ($\alpha_l \approx -k_1 k_2 \dots k_l$).

If a subset V satisfies the condition of Corollary 1 we obviously have $V \in U$, since $C_2 \subset U$ by Theorem 2. However, not every $V \in U$ satisfies the condition of Corollary 1, which follows from the fact that $U \not\subseteq C_2$. Hence, one can view this condition as an *additional* condition to be imposed on a set $V \in U$ in order to belong to C_2 . The next theorem clarifies this idea, by giving an alternative formulation of the condition of Corollary 1 in which the additional condition is explicitly shown.

Theorem 5 *Let $V \in \mathbb{R}^N$, then there exist an $l \in \mathbb{N}$ and $V_1, V_2, \dots, V_l \in \tilde{P}$ such that $V = V_1 \setminus V_2 \setminus \dots \setminus V_l$ if and only if*

$$V = (W_1 \cap W_2^*) \cup (W_3 \cap W_4^*) \cup \dots \cup (W_{2r-1} \cap W_{2r}^*)$$

for some $r \in \mathbb{N}$, $W_1, W_2, \dots, W_{2r} \in P$ and $W_1 \subseteq W_3 \subseteq \dots \subseteq W_{2r-1}$.

Verifying whether a subset V satisfies the conditions of Corollary 1 implies that we must find the appropriate decomposition of V . We have developed an algorithm to compute this decomposition for a given V if such a decomposition for V exists. This is the subject of the following section.

3.3 A decomposition algorithm

In the previous section a sufficient condition is given for a subset of \mathbb{R}^N to be classifiable by an exact 2-LP (see Corollary 1). In this section we derive a systematic verification method for this condition in a slightly restricted case: for a given subset $V \subseteq \mathbb{R}^N$ the presented algorithm finds $V_1, \dots, V_l \in P$ such that $V = V_1 \setminus V_2 \setminus \dots \setminus V_l$, if such a decomposition of V exists. The following ideas lie behind the introduced decomposition algorithm.

Assume $V = V_1 \setminus V_2$ for some *unknown* subsets $V_1, V_2 \in \tilde{P}$. In search for a solution of this equation we note that $V = V' \setminus (V' \setminus V)$ if and only if $V' \supseteq V$. Hence, if we can find a $V' \supseteq V$ satisfying $V' \in \tilde{P}$ and $V' \setminus V \in \tilde{P}$ we have finished. To find such a V' we exploit the property that all subsets in \tilde{P} are convex, which implies that V' and $V' \setminus V$ must be convex. Using $V' \setminus V = V' \setminus (V_1 \setminus V_2) = (V' \setminus V_1) \cup (V' \cap V_2)$ we find that $V' \setminus V$ is convex if V' is convex and $V_1 \supseteq V'$. Since V_1 can be any convex set satisfying $V_1 \supseteq V$, we see that V' has to be the smallest convex set with $V' \supseteq V$. This unique set is generally called the convex-hull of V and denoted by $\text{conv.hull}(V)$. Of course, the convexity of V' and $V' \setminus V$ does not necessarily guarantee that $V' \in \tilde{P}$ and $V' \setminus V \in \tilde{P}$. In Figure 2 below we give an example where using the convex-hull does not suffice.

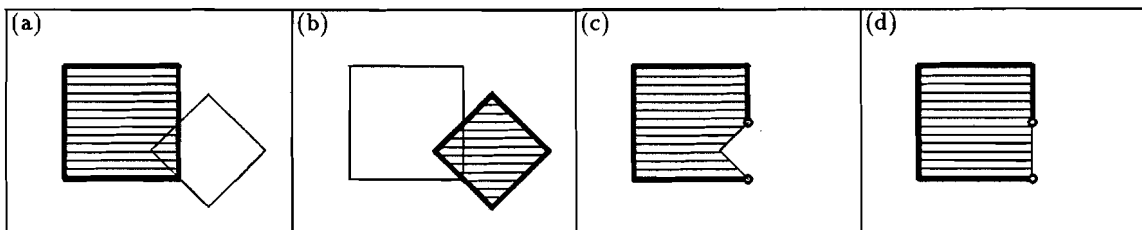


Figure 2: Example demonstrating the necessity of using the closure of the convex-hull.

In Figure 2a and 2b two subsets $V_1, V_2 \in P$ are given. Suppose we want to decompose $V = V_1 \setminus V_2$, which is presented in Figure 2c (thin boundary lines and circles are not part of the set). To this end we calculate $V' = \text{conv.hull}(V)$, the result being depicted in Figure 2d. It is not hard to see that $V' \notin \tilde{P}$. The only way to ensure that $V' \in \tilde{P}$ is to calculate $V' = \overline{\text{conv.hull}(V)}$, in which case we have a stronger result namely $V' \in P$. In Theorem 7 we prove that $\overline{\text{conv.hull}(V)} \in P$ for all $V \in U$. A disadvantage of using $\overline{\text{conv.hull}}$ instead of conv.hull is that we have to restrict ourselves to sets in \mathbb{R}^N that have a decomposition with only subsets in P , for one can easily construct a $V = V_1 \setminus V_2$, with $V_1, V_2 \in \tilde{P}$, for which there does not exist $W_1, W_2 \in P$ such that $V = W_1 \setminus W_2$.

Let V be a given subset of \mathbb{R}^N , then the decomposition $T = \overline{\text{conv.hull}(T)} \setminus (\overline{\text{conv.hull}(T)} \setminus T)$ found above is easily used to show that $V = V_1 \setminus V_2 \setminus \dots \setminus V_l \setminus T$ is an invariant of the DECAL-1 algorithm presented on the following page. This algorithm can be used to calculate a decomposition of a given subset.

DECAL-1

```

l := 0;
T := V;
while T ≠ ∅
do
  Vl+1 :=  $\overline{\text{conv.hull}(T)}$ ;
  T := Vl+1 \ T;
  l := l + 1
od;

```

If we assume that $V \in U$ then using Theorem 7 one can easily show that $T \in U$ is an invariant of the algorithm and $V_i \in P$, $i = 1, 2, 3, \dots$. This proves that the DECAL-1 algorithm finds a decomposition of $V \in U$ if the algorithm terminates. In the next theorem we prove that the algorithm terminates, and thus finds a decomposition of V , if a decomposition of V exists.

Theorem 6 *Let V be a subset of \mathbb{R}^N . If $V = W_1 \setminus W_2 \setminus \dots \setminus W_k$ for some $k \in \mathbb{N}$ and $W_1, \dots, W_k \in P$, then $l \leq k$ is an invariant of the DECAL-1 algorithm.*

The major part in the proof of Theorem 6 is the verification that $(l = 0) \vee (V_l \subseteq W_l)$ is an invariant of the DECAL-1 algorithm. For the remainder of the proof one can follow different lines. One alternative is to note that $V = W_1 \setminus W_2 \setminus \dots \setminus W_k \setminus W_{k+1}$ with $W_{k+1} = \emptyset$, which is then used to show that $(l \leq k) \wedge (T \neq \emptyset) \Rightarrow (l+1 \leq k)$. In the proof of Theorem 6 presented in the Appendix we follow a slightly different approach.

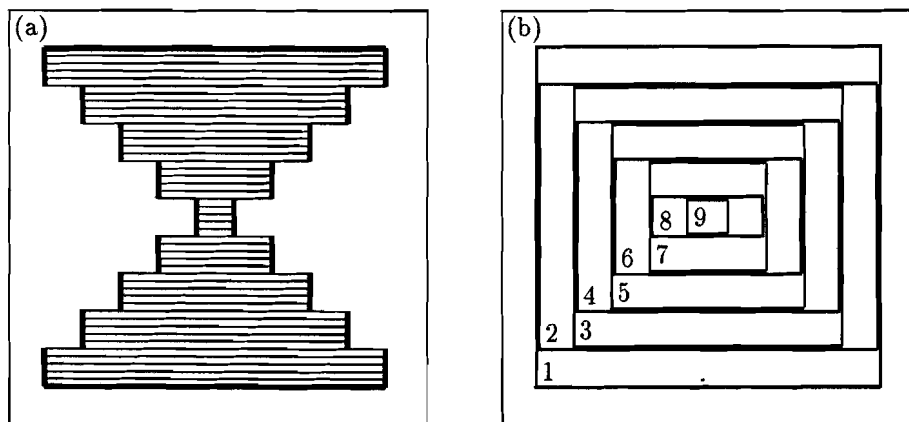


Figure 3: Example of the sequence of intermediate steps of the decomposition algorithm.

In Figure 3b we present the results of the DECAL-1 algorithm when it is applied to the subset V depicted in Figure 3a (which is a duplicate of Figure 1c). We obtain $V = V_1 \setminus V_2 \setminus \dots \setminus V_9$, with V_1, V_2, \dots, V_9 all rectangles obviously belonging to P , V_1 is the largest and V_9 is the smallest rectangle. The DECAL-1 algorithm can also be applied to most of the example subsets presented in [6].

So far, we have not discussed how the different steps in the DECAL-1 algorithm have to be executed. Especially the calculation of $\overline{\text{conv.hull}}(T)$ and $V_{l+1} \setminus T$ can cause considerable difficulty. We first give an alternative for the calculation of $V_{l+1} \setminus T$, the calculation of $\overline{\text{conv.hull}}(T)$ is dealt with below.

We show that we can replace the calculation of $V_{l+1} \setminus T$ in the DECAL-1 algorithm by the calculation of $V_{l+1} \cap (-1)^{l+1}V$, where $+V = V$ and $-V = V^*$. Let $V_0 = \mathbb{R}^N$ then one can verify that $T \subseteq V_l$ is an invariant of DECAL-1. Together with $V_i \in P$ for all i this implies that $V_l \subseteq V_{l-1}$ for all $l \geq 1$, which is easily used to show that $T = V_l \cap (-1)^l V$ is an invariant of DECAL-1. Hence, we find the following alternative for DECAL-1, which is especially convenient if we have a simple expression for both V and V^* . Note that this is the case for the subsets $V_1^{(i)}$ that arise in the solving of combinatorial optimization problems as presented in [13].

DECAL-2

```

l := 0;
T := V;
while T ≠ ∅
do
  Vl+1 :=  $\overline{\text{conv.hull}}(T)$ ;
  T := Vl+1 ∩ (-1)l+1V
  l := l + 1
od;
```

We now consider the calculation of $\overline{\text{conv.hull}}(T)$. In Theorem 7 below we give a systematic method for the calculation of $\overline{\text{conv.hull}}(V)$ for every subset $V \in U$. Before we can explain the different steps of this method we need to state some elementary results.

We use the following well-known expression for the convex-hull of a subset $V \subseteq \mathbb{R}^N$:

$$\text{conv.hull}(V) = \{x \in \mathbb{R}^N \mid x = \sum_{i=1}^r \lambda_i v_i, v_i \in V, \lambda_i \geq 0, \sum_{i=1}^r \lambda_i = 1, r \geq 1\}. \quad (10)$$

Indeed the convex-hull of the subset V is the smallest convex set containing V . Next, we define the cone of a subset $V \subseteq \mathbb{R}^N$ as the smallest convex cone containing V . A (convex) cone is a nonempty set of vectors C satisfying $x, y \in C \wedge \lambda, \mu \geq 0 \Rightarrow \lambda x + \mu y \in C$; see [7] for more information about the convex-hull and the cone. Hence, we have:

$$\text{cone}(V) = \{x \in \mathbb{R}^N \mid x = \sum_{i=1}^r \lambda_i v_i, v_i \in V, \lambda_i \geq 0\}. \quad (11)$$

An elementary results states that every polyhedron can be written as the sum of a bounded convex-hull and a cone, see Lemma 3 below. The proof of this lemma can be found in various books on linear algebra, one could use for instance [7] where also algorithms are presented for the calculation of x_i and y_i . Note that we use $\{x_i\}_{i=1}^k$ to denote the set $\{x_1, \dots, x_k\}$.

Lemma 3 ([7]) *Let $V \subseteq \mathbb{R}^N$, then $V \in P$ if and only if*

$$V = \text{conv.hull}(\{x_i\}_{i=1}^k) + \text{cone}(\{y_i\}_{i=1}^p)$$

for some $k, p \in \mathbb{N}$ and $x_i, y_i \in \mathbb{R}^N$.

Next, we show that the closure of a pseudo-polyhedron is a polyhedron. Recall that a pseudo-polyhedron is a polyhedron with a number of “missing” faces, which implies that the result is intuitively clear.

Lemma 4 *Let $V \in \tilde{P}$, then $\bar{V} \in P$.*

We are now ready for the final result of this paper. Theorem 7 proves that the closure of the convex-hull of a subset in U is always a polyhedron and gives a method for the determination of this polyhedron. The presented method consists of three steps. In the first step we apply the definition of U telling that every subset in U can be represented as a union of a finite number of pseudo-polyhedra. The closure of each of these pseudo-polyhedra is a polyhedron by Lemma 4 and hence, using Lemma 3 this explains the second step. The third step follows by using Lemma 3 in the opposite direction.

Theorem 7 *Let $V \in U$, then*

(i) $\overline{\text{conv.hull}}(V) \in P$,

(ii) $V' = \overline{\text{conv.hull}}(V)$ can be calculated using the following three steps:

1. Determine $l \in \mathbb{N}$ and $V_i \in \tilde{P}$ ($i = 1, \dots, l$) such that $V = \bigcup_{i=1}^l V_i$.

2. Determine $k_i, p_i \in \mathbb{N}$ and $x_{ij}, y_{ij} \in \mathbb{R}^N$ such that

$$\bar{V}_i = \text{conv.hull}(\{x_{ij}\}_{j=1}^{k_i}) + \text{cone}(\{y_{ij}\}_{j=1}^{p_i}).$$

3. Determine $V' \in P$ such that $V' = \text{conv.hull}(\{x_{ij}\}_{i=1}^l \{j=1}^{k_i}) + \text{cone}(\{y_{ij}\}_{i=1}^l \{j=1}^{p_i})$.

The method for the calculation of $\overline{\text{conv.hull}}(V)$ presented in the above theorem is not very efficient, which is mainly due to the second step. We are therefore searching a more efficient algorithm for the calculation of $\overline{\text{conv.hull}}(V)$. Until this more efficient algorithm is available we use the above method in the following way.

Suppose we want to solve a given combinatorial optimization problem with an exact 2-LP. This means that we have to find an exact 2-LP that classifies the subset $V = V_1^{(i)}$ for a fixed i (see [13] for a definition of $V_1^{(i)}$). We therefore try to find a decomposition $V = V_1 \setminus V_2 \setminus \dots \setminus V_l$ of V for some $l \in \mathbb{N}$, using the DECAL-2 algorithm. Since we have an explicit expression for both $+V = V$ and $-V = V^*$ in this case, it might be possible to find V_j iteratively by $V_0 = \mathbb{R}^N$, $V_{j+1} = \overline{\text{conv.hull}}(V_j \cap (-1)^j V)$, ($j = 1, \dots, l$) using the method of Theorem 7, for small values of the problem size N . Once, a general structure of the V_j 's is found one can then try to prove the correctness of the decomposition for general N directly.

4 Concluding remarks

We discussed the classification capabilities of exact two-layered perceptrons (2-LPs). A detailed analysis was used to obtain necessary and sufficient conditions for a subset to be classifiable by an exact 2-LP. The necessary conditions can be used to show that a given problem cannot be solved exactly by a 2-LP. One such problem is the sorting problem treated in [14]. Trying to learn a 2-LP to solve such a problem is bound to give a very poor result and one should therefore consider using a 3-LP in these cases. The sufficient conditions can be used to prove that a problem can be solved exactly by a 2-LP and its verification algorithm can be used to obtain the required number of hidden nodes. Although an exact set of weights can also be determined, the relatively large variance in the size of the weights implies that the use of a learning algorithm is sometimes more useful for the determination of the weights.

Acknowledgements

We thank C.A.J. Hurkens for his helpful comments about the calculation of the convex-hull of an arbitrary subset.

References

- [1] E.B. BAUM, On the Capabilities of Multilayer Perceptrons, *Journal of Complexity*, 4, 193–215, 1988.
- [2] G. CYBENKO, Approximation by Superpositions of a Sigmoidal function, *Tech. Rep. No. 856, Univ. of Illinois*, 1989.
- [3] G. CYBENKO, Complexity Theory of Neural Networks and Classification Problems, *Proceedings of the 1990 EURASIP Workshop on Neural Networks, Sesimbra, Portugal*, 26–44, 1990.
- [4] K. FUNAHASHI, On the Approximate Realization of Continuous Mappings by Neural Networks, *Neural Networks*, 2, 183–192, 1989.
- [5] K. HORNIK, M. STINCHCOMBE AND H. WHITE, Multilayer Feedforward Networks are Universal Approximators, *Neural Networks*, 2, 359–366, 1989.
- [6] W.Y. HUANG AND R.P. LIPPMANN, Neural Net and Traditional Classifiers, in (*D.Z. Anderson, ed.*) *Neural Information Processing Systems*, 387–396, 1987.
- [7] A. SCHRIJVER, *Theory of Linear and Integer Programming*, Wiley, 1986.
- [8] R.P. LIPPMANN, An Introduction to computing with neural nets, *IEEE ASSP Magazine*, 4, 4–22, 1987.
- [9] J. MAKHOUL, R. SCHWARTZ, AND A. EL-JAROUDI, Classification Capabilities of Two-Layer Neural Nets, *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, Glasgow, Scotland*, 635–638, 1989.
- [10] M. MINSKY AND S. PAPERT, *Perceptrons: An Introduction to Computational Geometry*, MIT Press, 1969.
- [11] D.E. RUMELHART, G.E. HINTON AND R.J. WILLIAMS, Learning Internal Representations by Error Propagation, in (*D.E. Rumelhart and J.L. McClelland, Eds.*) *Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Vol. 1: Foundations*, MIT Press, 318–362, 1986.
- [12] A. WIELAND AND R. LEIGHTON, Geometric Analysis of Neural Network Capabilities, *Proceedings of the 1st International Conference on Neural Networks*, IEEE, III, 385–393, 1987.
- [13] P.J. ZWIETERING, E.H.L. AARTS AND J. WESSELS, The Design and Complexity of Exact Multi-Layered Perceptrons, *Memorandum COSOR 91-08, Eindhoven Univ. of Techn.*, 1991.
- [14] P.J. ZWIETERING, E.H.L. AARTS AND J. WESSELS, Sorting With A Neural Net, *Memorandum COSOR 91-10, Eindhoven Univ. of Techn.*, 1991.

Appendix

This appendix contains the proofs of the lemmas and theorems presented in this paper.

Theorem 3 *Let $V \in U$ be a subset for which there exist a $W \in C_1$ and two spheres B_1, B_2 such that*

$$\left. \begin{array}{l} \emptyset \neq B_1 \cap W^\circ \subseteq V \\ \emptyset \neq B_1 \cap W^* \subseteq V^* \end{array} \right\} \wedge \left\{ \begin{array}{l} \emptyset \neq B_2 \cap W^\circ \subseteq V^* \\ \emptyset \neq B_2 \cap W^* \subseteq V, \end{array} \right. \quad (12)$$

then $V \notin C_2$.

Proof

Assume that $V \in C_2$, $W \in C_1$ and two spheres B_1, B_2 exist that satisfy (12), we show that this leads to a contradiction. Since $V \in C_2$ we have $V = \mathcal{J}(f)$, for some $f \in R_{2,N,1}$, $f = g \circ h$ with $g \in R_{1,K,1}$, $h \in R_{1,N,K}$ and $K \in \mathbb{N}$. Let $g = \Theta \circ \tilde{g}$ with $\tilde{g}(x) = a \cdot x + b$ for some $a \in \mathbb{R}^K$, $b \in \mathbb{R}$ and define $V_i = \mathcal{J}(h_i)$, $i = 1, \dots, k$, hereby assuming that $h_i \neq h_j$, ($i \neq j$).

From (12) it follows that $W \neq \emptyset, \mathbb{R}^N$, which means that W is a closed linear half-space. $\hat{W} = W \cap (W^\circ)^*$ is the hyperplane that bounds both W and $(W^\circ)^*$. Consider $B_1 \cap \hat{W}$, this intersection is not empty since $B_1 \cap W^\circ$ and $B_1 \cap W^*$ are not empty. Without loss of generality we assume that $B_1 \cap \hat{W} \subseteq V$ or $B_1 \cap \hat{W} \subseteq V^*$: if only a part of $B_1 \cap \hat{W} \subseteq V$ is a subset of V then we can shrink the size of B_1 such that one of the two statements becomes true, using that V has only a finite number of defining half-spaces h_i . The same argument holds for $B_2 \cap \hat{W}$, so that we obtain the following four cases:

- (i) $B_1 \cap \hat{W} \subseteq V$ and $B_2 \cap \hat{W} \subseteq V$,
- (ii) $B_1 \cap \hat{W} \subseteq V$ and $B_2 \cap \hat{W} \subseteq V^*$,
- (iii) $B_1 \cap \hat{W} \subseteq V^*$ and $B_2 \cap \hat{W} \subseteq V$,
- (iv) $B_1 \cap \hat{W} \subseteq V^*$ and $B_2 \cap \hat{W} \subseteq V^*$.

We show that in case (i) and (ii) we obtain a contradiction. This implies that the same holds for case (iii) and (iv), as they can be obtained from case (ii) and (i), respectively, by swapping V and V^* . To prove that case (i) and (ii) lead to a contradiction, we need the following Lemma.

Lemma

If

$$\left\{ \begin{array}{l} \emptyset \neq B \cap W \subseteq V \\ \emptyset \neq B \cap W^* \subseteq V^*, \end{array} \right. \quad (13)$$

for some sphere B , then $W = V_i$ for some $i \in \{1, \dots, k\}$ with $a_i > 0$ and if $V_j = (W^\circ)^*$ for some $j \neq i$ then $a_i > a_j$.

End of Lemma

We apply the result of this lemma to show that case (i) and (ii) lead to a contradiction. Case (i). In this case (12) becomes:

$$\left. \begin{array}{l} \emptyset \neq B_1 \cap W \subseteq V \\ \emptyset \neq B_1 \cap W^* \subseteq V^* \end{array} \right\} \wedge \left\{ \begin{array}{l} \emptyset \neq B_2 \cap W^\circ \subseteq V^* \\ \emptyset \neq B_2 \cap (W^\circ)^* \subseteq V. \end{array} \right.$$

Apply the Lemma to B_1 and W , then we obtain $W = V_i$ for some $i \in \{1, \dots, K\}$ and if $V_j = (W^\circ)^*$ then $a_i > a_j$. Apply the Lemma to B_2 and $(W^\circ)^*$, then we obtain $(W^\circ)^* = V_j$ for some $j \in \{1, \dots, K\}$ and hence $a_i > a_j$. However, since $V_i = W = (((W^\circ)^*)^*)^*$, we also find $a_j > a_i$, obviously obtaining a contradiction.

Case (ii). Then (12) becomes:

$$\left. \begin{array}{l} \emptyset \neq B_1 \cap W \subseteq V \\ \emptyset \neq B_1 \cap W^* \subseteq V^* \end{array} \right\} \wedge \left\{ \begin{array}{l} \emptyset \neq B_2 \cap W \subseteq V^* \\ \emptyset \neq B_2 \cap W^* \subseteq V. \end{array} \right.$$

Apply the Lemma to B_1 and W , then we obtain $W = V_i$ for some $i \in \{1, \dots, K\}$ and $a_i > 0$. Without loss of generality we assume $\tilde{g}(h(x)) \neq 0$ for all $x \in \mathbb{R}^N$, which implies that $V^* = \Theta \circ (-\tilde{g}) \circ h$. By applying the Lemma to V^* , W and B_2 , we then obtain $W = V_j$ for some $j \in \{1, \dots, K\}$ and $-a_j > 0$. This would yield a pair $i \neq j$ with $V_i = V_j$, contradicting the assumptions.

The proof of the theorem is now completed by giving the proof of the above Lemma.

Proof of Lemma

Let $W = \{x \in \mathbb{R}^N \mid p \cdot x + q \geq 0\}$ for some $p \in \mathbb{R}^N \setminus \{0\}$ and $q \in \mathbb{R}$. By defining $h_+, h_- \in R_1$ as $h_+(x) = \Theta(p \cdot x + q)$, $h_-(x) = \Theta(-p \cdot x - q)$, we have $W = \mathcal{J}(h_+)$ and $W^\circ = \mathcal{J}^*(h_-)$. Since we can write $f(x) = \Theta(\tilde{g}(h(x))) = \Theta(0h_+(x) + 0h_-(x) + \sum_{i=1}^K a_i h_i(x) + b)$, we may assume without loss of generality that $h_1 = h_+$, $h_2 = h_-$ and $h_i \neq h_+, h_-$ for $i = 3, \dots, K$. Let $h_i(x) = \Theta(c_i \cdot x + d_i)$ for some $c_i \in \mathbb{R}^N \setminus \{0\}$ and $d_i \in \mathbb{R}$, with $(c_i, d_i) \neq \lambda(p, q)$ for all $\lambda \in \mathbb{R}$ and $i = 3, \dots, K$.

We determine $x_2, x_3, \dots, x_K \in \hat{W} = \{x \in \mathbb{R}^N \mid p \cdot x + q = 0\}$ and $\varepsilon_2 > \varepsilon_3 > \dots, \varepsilon_K > 0$, with $B(x_2, \varepsilon_2) \subseteq B$ and

$$B(x_i, \varepsilon_i) \subseteq B(x_{i-1}, \varepsilon_{i-1}) \cap \{x \in \mathbb{R}^N \mid c_i \cdot x + d_i \neq 0\}, \quad \text{for } i = 3, \dots, K. \quad (14)$$

First, since $B \cap W \neq \emptyset$ and $B \cap W^* \neq \emptyset$, we must have $B \cap \hat{W} \neq \emptyset$. Hence, there exists an $x_2 \in B \cap \hat{W}$. Then obviously $B(x_2, \varepsilon_2) \subseteq B$ for some $\varepsilon_2 > 0$.

Next, assume x_{i-1} and ε_{i-1} have been determined for certain $i \in \{3, \dots, K\}$, for which $B(x_{i-1}, \varepsilon_{i-1})$ satisfies the above conditions. We then construct $B(x_i, \varepsilon_i)$ satisfying (14). Since $(c_i, d_i) \neq \lambda(p, q)$ for all $\lambda \in \mathbb{R}$, we cannot have $c_i \cdot x + d_i = 0$ for all $x \in B(x_{i-1}, \varepsilon_{i-1}) \cap \hat{W}$. Hence, there exists an $x_i \in B(x_{i-1}, \varepsilon_{i-1}) \cap \hat{W}$ with $c_i \cdot x + d_i \neq 0$. This implies that $B(x_i, \varepsilon_i) \subseteq \{x \in \mathbb{R}^N \mid c_i \cdot x + d_i \neq 0\}$ and $B(x_i, \varepsilon_i) \subseteq B(x_{i-1}, \varepsilon_{i-1})$ for some $\varepsilon_i > 0$.

Using (14) we see that $x \in B(x_K, \varepsilon_K)$ implies that $x \in B$ and $c_i \cdot x + d_i \neq 0$ for $i = 3, \dots, K$. Take $y, z \in B(x_K, \varepsilon_K)$ with $p \cdot y + q > 0$ and $p \cdot z + q < 0$, which is possible

since $p \cdot x_K + q = 0$. It then follows that $h_1(x_K) = h_1(y) = h_2(x_K) = h_2(z) = 1$, $h_1(z) = h_2(y) = 0$. Furthermore, by using (13), we have $x_K, y \in B \cap \mathcal{J}(h_+) = B \cap W \subseteq V$ and $z \in B \cap \mathcal{J}^*(h_+) = B \cap W^* \subseteq V^*$, which implies that $\tilde{g}(x_K) \geq 0$, $\tilde{g}(y) \geq 0$ and $\tilde{g}(z) < 0$. Finally, by using $B(x_K, \varepsilon_K) \subseteq \bigcap_{i=3}^K \{x \in \mathbb{R}^N \mid c_i \cdot x + d_i \neq 0\}$ one can show that $h_i(x_K) = h_i(y) = h_i(z)$ for $i = 3, \dots, K$. Hence,

$$a_1 = a_1 h_1(x_K) + a_2 h_2(x_K) - a_1 h_1(z) - a_2 h_2(z) = \tilde{g}(h(x_K)) - \tilde{g}(h(z)) > 0,$$

and

$$a_1 - a_2 = a_1 h_1(y) + a_2 h_2(y) - a_1 h_1(z) - a_2 h_2(z) = \tilde{g}(h(y)) - \tilde{g}(h(z)) > 0,$$

which completes the proof of the Lemma and hence of the Theorem. \square

Theorem 4 Let $V_1 \in \tilde{P}$, $V_2 \in C_2$ and assume $V_i = \mathcal{J}(g_i \circ h_i)$, $g_i = \Theta \circ \tilde{g}_i \in R_{1,k_i,1}$, $h_i \in R_{1,N,k_i}$, $\tilde{g}_1(h_1(x)) \in \{-k_1, -k_1 + 1, \dots, 0\}$ and $\tilde{g}_2(h_2(x)) \in \{\alpha, \alpha + 1, \dots, \beta\}$ for some $k_1, k_2 \in \mathbb{N}$, $\alpha, \beta \in \mathbb{Z}$, $\alpha < 0 \leq \beta$ and all $x \in \mathbb{R}^N$.

Then $V_1 \setminus V_2 = \mathcal{J}(g \circ h)$, with $h \in R_{1,N,k_1+k_2}$ and $g = \Theta \circ \tilde{g} \in R_{1,k_1+k_2,1}$ given by $h(x) = (h_1(x), h_2(x))$ and

$$\tilde{g}(x_1, x_2) = -\alpha \tilde{g}_1(x_1) - \tilde{g}_2(x_2) - 1, \quad (15)$$

respectively. Furthermore $\tilde{g}(h(x)) \in \{k_1\alpha - \beta - 1, \dots, -\alpha - 1\}$ for all $x \in \mathbb{R}^N$.

Proof

If $x \notin V_1$ then $\tilde{g}_1(h_1(x)) \leq -1$ and hence, $\tilde{g}(h(x)) \leq \alpha - \tilde{g}_2(h_2(x)) - 1 \leq -1$.

If $x \in V_1$ then $\tilde{g}_1(h_1(x)) = 0$ and hence, $\tilde{g}(h(x)) = -\tilde{g}_2(h_2(x)) - 1 \geq 0$ if and only if $x \notin V_2$. \square

Theorem 5 Let $V \in \mathbb{R}^N$, then there exist an $l \in \mathbb{N}$ and $V_1, V_2, \dots, V_l \in P$ such that $V = V_1 \setminus V_2 \setminus \dots \setminus V_l$ if and only if

$$V = (W_1 \cap W_2^*) \cup (W_3 \cap W_4^*) \cup \dots \cup (W_{2r-1} \cap W_{2r}^*)$$

for some $r \in \mathbb{N}$, $W_1, W_2, \dots, W_{2r} \in P$ and $W_1 \subseteq W_3 \subseteq \dots \subseteq W_{2r-1}$.

Proof

(\Rightarrow) Assume that $V = V_1 \setminus V_2 \setminus \dots \setminus V_l$ for some $l \in \mathbb{N}$ and $V_1, V_2, \dots, V_l \in P$. Since $V_l = V_l \setminus \emptyset$ and $\emptyset \in P$, we assume without loss of generality that $l = 2r$ for some $r \in \mathbb{N}$. Using $V_1 \setminus (V_2 \setminus W) = (V_1 \cap V_2^*) \cup (V_1 \cap W)$ and mathematical induction one can show that:

$$V_1 \setminus V_2 \setminus \dots \setminus V_l = (W_1 \cap W_2^*) \cup (W_3 \cap W_4^*) \cup \dots \cup (W_{2r-1} \cap W_{2r}^*),$$

where $W_{2i-1} = V_1 \cap V_3 \cap \dots \cap V_{2i-1}$ and $W_{2i} = V_{2i}$. Obviously $W_1, W_2, \dots, W_{2r} \in P$ and $W_1 \supseteq W_3 \supseteq \dots \supseteq W_{2r-1}$.

(\Leftarrow) Assume that $V = (W_1 \cap W_2^*) \cup (W_3 \cap W_4^*) \cup \dots \cup (W_{2r-1} \cap W_{2r}^*)$ for some $r \in \mathbb{N}$, $W_1, W_2, \dots, W_{2r} \in P$ and $W_1 \supseteq W_3 \supseteq \dots \supseteq W_{2r-1}$, then we also have:

$$V = (W_1 \cap W_2^*) \cup (W_1 \cap W_3 \cap W_4^*) \cup \dots \cup (W_1 \cap W_3 \cap \dots \cap W_{2r-1} \cap W_{2r}^*),$$

which implies that $V = W_1 \setminus W_2 \setminus \dots \setminus W_{2r}$, see the (\Rightarrow) part. \square

Theorem 6 *Let V be a subset of \mathbb{R}^N . If $V = W_1 \setminus W_2 \setminus \dots \setminus W_k$ for some $k \in \mathbb{N}$ and $W_1, \dots, W_k \in P$, then $l \leq k$ is an invariant of the DECAL-1 algorithm.*

Proof

Suppose $V = W_1 \setminus W_2 \setminus \dots \setminus W_k$ with $W_i \in P$ ($i = 1, \dots, k$). Define the sets T_i , \tilde{V}_i and Z_i ($i = 0, \dots, k$) by:

$$\begin{aligned} Z_k &= \emptyset, \\ Z_i &= W_{i+1} \setminus Z_{i+1}, \quad (i = k-1, \dots, 0), \end{aligned} \quad (16)$$

$$\begin{aligned} \tilde{V}_0 &= \mathbb{R}^N, \\ \tilde{V}_i &= \overline{\text{conv.hull}(\tilde{V}_{i-1} \cap Z_{i-1})}, \quad (i = 1, \dots, k), \end{aligned} \quad (17)$$

$$\begin{aligned} T_0 &= V, \\ T_i &= \tilde{V}_i \setminus T_{i-1}, \quad (i = 1, \dots, k). \end{aligned} \quad (18)$$

Using (16) one can easily show that $\tilde{V}_{i-1} \cap Z_{i-1} = (\tilde{V}_{i-1} \cap W_i) \setminus W_{i+1} \setminus \dots \setminus W_k$, for $i = 1, \dots, k$. Hence, using (17), $\tilde{V}_0 \in P$, Theorem 7 and mathematical induction we find that $\tilde{V}_i \in P$ for $i = 0, \dots, k$. This implies that we also have $\tilde{V}_i \cap \tilde{V}_{i-1} \cap W_i \in P$, for all $i = 1, \dots, k$, which we use to derive:

$$\tilde{V}_{i-1} \cap Z_{i-1} = \tilde{V}_i \setminus Z_i, \quad (19)$$

for all $i = 1, \dots, k$. This proof goes as follows. Using (17) and (16) we find that

$$\begin{aligned} \tilde{V}_{i-1} \cap Z_{i-1} &= \tilde{V}_i \cap \tilde{V}_{i-1} \cap Z_{i-1} \\ &= (\tilde{V}_i \cap \tilde{V}_{i-1} \cap W_i) \setminus Z_i \\ &\subseteq \tilde{V}_i \cap \tilde{V}_{i-1} \cap W_i \\ &\subseteq \tilde{V}_i \\ &= \overline{\text{conv.hull}(\tilde{V}_{i-1} \cap Z_{i-1})}, \end{aligned} \quad (20)$$

and hence, since $\tilde{V}_i \cap \tilde{V}_{i-1} \cap W_i \in P$ implies $\tilde{V}_i \cap \tilde{V}_{i-1} \cap W_i$ closed and convex, we have $\tilde{V}_i \cap \tilde{V}_{i-1} \cap W_i = \overline{\text{conv.hull}(\tilde{V}_{i-1} \cap Z_{i-1})} = \tilde{V}_i$. Substituting this back into (20) we get (19).

Now we use (19) to show that:

$$T_i = \tilde{V}_i \cap Z_i, \quad (21)$$

for all $i = 0, 1, \dots, k$. We start with $T_0 = V = \mathbb{R}^N \cap V = \tilde{V}_0 \cap Z_0$. Next, assume (21) holds for some $i \in \{0, \dots, k-1\}$, then from (19) it follows that $T_i = \tilde{V}_{i+1} \setminus Z_{i+1}$, and hence:

$$\begin{aligned} T_{i+1} &= \tilde{V}_{i+1} \setminus T_i \\ &= \tilde{V}_{i+1} \setminus (\tilde{V}_{i+1} \setminus Z_{i+1}) \\ &= \tilde{V}_{i+1} \cap Z_{i+1}, \end{aligned}$$

hereby completing the proof of (21) by mathematical induction. From (21), (17) and (18) we conclude that \tilde{V}_i and T_i satisfy:

$$\begin{aligned}\tilde{V}_{i+1} &= \overline{\text{conv.hull}}(T_i), \\ T_{i+1} &= \tilde{V}_{i+1} \setminus T_i,\end{aligned}$$

for all $i = 0, \dots, k-1$. Since $T_0 = V$, this proves that $T = T_l$ and $V_l = \tilde{V}_l$ are invariants of the DECAL-1 algorithm. The result now follows from $T_k = \tilde{V}_k \cap Z_k = \emptyset$. \square

Lemma 4 *Let $V \in \tilde{P}$, then $\bar{V} \in P$.*

Proof

Let $V \in \tilde{P}$, then $V = \bigcap_{i=1}^l V_i$, for some $l \in \mathbb{N}$ and $V_i \in \tilde{C}_1$. Assume $V_i = \{x \mid a_i \cdot x + b_i \geq 0\}$, $i \in I_1$, and $V_i = \{x \mid a_i \cdot x + b_i > 0\}$, $i \in I_2$, for some $a_i \in \mathbb{R}^N$, $b_i \in \mathbb{R}$ ($i = 1, \dots, l$), and define:

$$W = \bigcap_{i=1}^l \{x \mid a_i \cdot x + b_i \geq 0\}.$$

Since $V \subseteq W$, we have $\bar{V} \subseteq \bar{W} = W$. To prove that $W \subseteq \bar{V}$ we take $x \in W \setminus V$ and $\varepsilon > 0$. Then $a_i \cdot x + b_i = 0$ for $i \in I_3 \subseteq I_2$ and $a_i \cdot x + b_i > 0$ for $i \in I_3 \setminus I_2$. Let $\delta > 0$ be such that $a_i \cdot y + b_i > 0$ for all $i \in I_3 \setminus I_2$ and $y \in \mathbb{R}^N$ with $\|x - y\| < \delta$.

We may assume that $V \neq \emptyset$ (otherwise the proof is trivial), which implies that $z \in V$ for some $z \in \mathbb{R}^N$. Let $\lambda = \min(\varepsilon, \delta, 1)(\|x\| + \|z\| + 1)^{-1}$ and define $y = (1 - \lambda)x + \lambda z$. Then $\|x - y\| < \varepsilon$ and it remains to show that $y \in V$.

Firstly, since $x, z \in W$, $\lambda \in (0, 1]$ and W is convex we find that $y \in W$, which implies that $a_i \cdot y + b_i \geq 0$ for all $i \in I_1$. Secondly, we have that $a_i \cdot y + b_i = (1 - \lambda)(a_i \cdot x + b_i) + \lambda(a_i \cdot z + b_i) = \lambda(a_i \cdot z + b_i) > 0$ for all $i \in I_3$. Finally, $\|x - y\| < \delta$ implies that $a_i \cdot y + b_i > 0$ for all $i \in I_3 \setminus I_2$. \square

Theorem 7 *Let $V \in U$, then*

(i) $\overline{\text{conv.hull}}(V) \in P$,

(ii) $V' = \overline{\text{conv.hull}}(V)$ can be calculated using the following three steps:

1. Determine $l \in \mathbb{N}$ and $V_i \in \tilde{P}$ ($i = 1, \dots, l$) such that $V = \bigcup_{i=1}^l V_i$.
2. Determine $k_i, p_i \in \mathbb{N}$ and $x_{ij}, y_{ij} \in \mathbb{R}^N$ such that $\bar{V}_i = \text{conv.hull}(\{x_{ij}\}_{j=1}^{k_i}) + \text{cone}(\{y_{ij}\}_{j=1}^{p_i})$.
3. Determine $V' \in P$ such that $V' = \text{conv.hull}(\{x_{ij}\}_{i=1}^l \{j=1}^{k_i}) + \text{cone}(\{y_{ij}\}_{i=1}^l \{j=1}^{p_i})$.

Proof

(i) Follows directly from (ii) and Lemma 3.

(ii) Let $V \in U$, then $V = \bigcup_{i=1}^l V_i$ for some $l \in \mathbb{N}$ and $V_i \in \tilde{P}$. If $l = 1$ the proof is trivial, we therefore assume $l \geq 2$. By Lemma 4 we have that $\bar{V}_i \in P$, and hence, by Lemma 3, it follows that:

$$\bar{V}_i = \text{conv.hull}(\{x_{ij}\}_{j=1}^{k_i}) + \text{cone}(\{y_{ij}\}_{j=1}^{p_i}),$$

for some $k_i, p_i \in \mathbb{N}$ and $x_{ij}, y_{ij} \in \mathbb{R}^N$. Define the subsets $W, V' \subseteq \mathbb{R}^N$ by:

$$W = \{x \in \mathbb{R}^N \mid x = \sum_{i=1}^l \lambda_i v_i, v_i \in \bar{V}_i, \lambda_i \geq 0, \sum_{i=1}^l \lambda_i = 1\},$$

$$V' = \text{conv.hull}(\{x_{ij}\}_{i=1}^l \{y_{ij}\}_{i=1}^l) + \text{cone}(\{y_{ij}\}_{i=1}^l \{y_{ij}\}_{i=1}^l).$$

Then the proof is completed by showing that (a) $\overline{\text{conv.hull}}(V) = \bar{W}$, and (b) $\bar{W} = V'$.

(a) Follows straightforwardly by:

$$\overline{\text{conv.hull}}(V) = \overline{\text{conv.hull}}(\bigcup_{i=1}^l V_i) = \overline{\text{conv.hull}}(\bigcup_{i=1}^l \bar{V}_i) = \bar{W},$$

using that \bar{V}_i is convex in the last step.

(b) One can easily verify that $W \subseteq V'$, which implies that $\bar{W} \subseteq \bar{V}' \subseteq V'$. Hence, it remains to verify that $V' \subseteq \bar{W}$.

Let $x \in V'$, then from the definition of V' it follows that:

$$x = \sum_{i=1}^l \sum_{j=1}^{k_i} \mu_{ij} x_{ij} + \sum_{i=1}^l \sum_{j=1}^{p_i} \tau_{ij} y_{ij},$$

for some $\mu_{ij} \geq 0, \tau_{ij} \geq 0$ with $\sum_{i=1}^l \sum_{j=1}^{k_i} \mu_{ij} = 1$. If $k_1 = k_2 = \dots = k_l = 0$, then

$$x = \sum_{i=1}^l \frac{1}{l} \sum_{j=1}^{p_i} (l\tau_{ij}) y_{ij} \in W,$$

otherwise we assume without loss of generality that $k_1 \geq k_2 \geq \dots \geq k_r > 0, k_{r+1} = \dots = k_l = 0$ for some $1 < r \leq l$ and $\mu_{11} > 0$.

Take $\varepsilon > 0$. Let $\delta = \min(\frac{1}{2}\mu_{11}, \varepsilon(\sum_{i=1}^r \|x_{i1}\| + 1)^{-1}) > 0$ and define:

$$y = \sum_{i=1}^l \lambda_i \left(\sum_{j=1}^{k_i} \tilde{\mu}_{ij} x_{ij} + \sum_{j=1}^{p_i} \tilde{\tau}_{ij} y_{ij} \right),$$

where $\lambda_1 = \sum_{j=1}^{k_1} \mu_{1j} - \delta, \lambda_i = \sum_{j=1}^{k_i} \mu_{ij} + \frac{\delta}{l-1}, (i = 2, \dots, l), \tilde{\mu}_{11} = (\mu_{11} - \delta)/\lambda_1, \tilde{\mu}_{i1} = (\mu_{i1} + \frac{\delta}{l-1})/\lambda_i, (i = 2, \dots, r), \tilde{\mu}_{ij} = \mu_{ij}/\lambda_i, (i = 1, \dots, r, j = 2, \dots, k_i)$ and $\tilde{\tau}_{ij} = \tau_{ij}/\lambda_i$. This implies that $y \in W$ since $\lambda_i > 0, \sum_{i=1}^l \lambda_i = 1$ and $\sum_{j=1}^{k_i} \tilde{\mu}_{ij} = 1 (i = 1, \dots, r)$.

Finally, we have that $\|x - y\| = \|\delta x_{11} - \sum_{i=2}^r \frac{\delta}{l-1} x_{i1}\| \leq \delta(\sum_{i=1}^r \|x_{i1}\|) < \varepsilon$, which completes the proof of (b) and, hence, of the theorem. \square

EINDHOVEN UNIVERSITY OF TECHNOLOGY
Department of Mathematics and Computing Science
PROBABILITY THEORY, STATISTICS, OPERATIONS RESEARCH
AND SYSTEMS THEORY
P.O. Box 513
5600 MB Eindhoven, The Netherlands

Secretariate: Dommelbuilding 0.03
Telephone : 040-473130

-List of COSOR-memoranda - 1991

<u>Number</u>	<u>Month</u>	<u>Author</u>	<u>Title</u>
91-01	January	M.W.I. van Kraaij W.Z. Venema J. Wessels	The construction of a strategy for manpower planning problems.
91-02	January	M.W.I. van Kraaij W.Z. Venema J. Wessels	Support for problem formulation and evaluation in manpower planning problems.
91-03	January	M.W.P. Savelsbergh	The vehicle routing problem with time windows: minimizing route duration.
91-04	January	M.W.I. van Kraaij	Some considerations concerning the problem interpreter of the new manpower planning system formasy.
91-05	February	G.L. Nemhauser M.W.P. Savelsbergh	A cutting plane algorithm for the single machine scheduling problem with release times.
91-06	March	R.J.G. Wilms	Properties of Fourier-Stieltjes sequences of distribution with support in $[0,1)$.
91-07	March	F. Coolen R. Dekker A. Smit	Analysis of a two-phase inspection model with competing risks.
91-08	April	P.J. Zwietering E.H.L. Aarts J. Wessels	The Design and Complexity of Exact Multi-Layered Perceptrons.
91-09	May	P.J. Zwietering E.H.L. Aarts J. Wessels	The Classification Capabilities of Exact Two-Layered Peceptrons.