Working Paper

Exact Classification With Two-Layered Perceptrons

E.H.L. Aarts

P.J. Zwietering

J. Wessels

Exact Classification With
Two-Layered Perceptrons

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

Exact Classification With Two-Layered Perceptrons

P.J. Zwietering[1], E.H.L. Aarts[1,2] and J. Wessels[1,3]

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
3. International Institute for Applied Systems Analysis, A-2361 Laxenburg, Austria.

Abstract

We study the capabilities of two-layered perceptrons for classifying exactly a given subset. Both necessary and sufficient conditions are derived for subsets to be exactly classifiable with two-layered perceptrons that use the hard-limiting response function. The necessary conditions can be viewed as generalizations of the linear-separability condition of one-layered perceptrons and confirm the conjecture that the capabilities of two-layered perceptrons are more limited than those of three-layered perceptrons. The sufficient conditions show that the capabilities of two-layered perceptrons extend beyond the exact classification of convex subsets. Furthermore, we present an algorithmic approach to the problem of verifying the sufficiency condition for a given subset.

Key words: classification, multi-layered perceptrons, neural networks

1 Introduction

Classification is a basic capability of multi-layered perceptrons that plays an important role in the application of this type of neural network to other problem areas such as combinatorial optimization. In [14] we proved that any combinatorial optimization problem can be solved with a three-layered perceptron under some linearity constraints. This result is based on a necessary and sufficient condition for a subset to be exactly classifiable with a three-layered perceptron. Furthermore, we proved that subsets which can be classified exactly with an m-layered perceptron for some integer m, can also be classified exactly with a three-layered perceptron. So, from this point of view it is needless to use more than three layers. There may, however, be arguments for using more than three layers. For instance, using four layers might require a smaller total number of nodes in the network than using three layers.

One may pose the following question. Are the classification capabilities of two-layered perceptrons less than those of three-layered perceptrons, and, if there is a difference, how can it be characterized? This question is of practical interest since two-layered perceptrons are believed to learn much faster than three-layered perceptrons. Furthermore, the learning in two-layered perceptrons is more easy to analyze than in three-layered perceptrons.

In this paper we consider the problem of finding two-layered perceptrons that exactly classify a given subset. Furthermore, we restrict ourselves to perceptrons with a hard limiting response function. So far, few results have been reported on the exact classification capabilities of two-layered perceptrons using the hard limiting response function. In his introductory paper, Lippmann briefly discusses the classification capabilities of multi-layered perceptrons [8]. Lippmann states that a subset has to be *convex* for being classifiable with a two-layered perceptron. However, several authors have pointed out that the capabilities of two-layered perceptrons extend beyond the exact classification of convex subsets. Wieland and Leighton [13], 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 with a two-layered perceptron. We extend these results by presenting formal and more general results on the classification capabilities of two-layered perceptrons.

Some recent papers consider the approximate classification capabilities of two-layered perceptrons that use a sigmoidal response function. Cybenko [1], Funahashi [3], and Hornik et al. [5] show that two-layered perceptrons are capable of classifying a given subset within arbitrary precision. Although these results can also be used to obtain results about the exact classifiability of a given subset, as is shown by Cybenko [1] and Li [7], they do not say anything about the required number of hidden nodes. A first attempt to solve this problem is presented by Cybenko [2], who derives an upper-bound on the required number of hidden nodes. Our approach concentrates on finding the minimal number of hidden nodes for exactly classifying a given non-finite subset.

The remainder of the paper is organized as follows. In Section 2 we briefly summarize some basic definitions and known results. Section 3 presents necessary and sufficient conditions for a subset to be exactly classifiable with a two-layered perceptron. Furthermore, this section contains algorithms for the verification of the sufficiency condition. Section 4 presents a discussion and

some concluding remarks. The paper ends with an appendix that contains the proofs of the lemmas of Section 3.

2   Preliminaries

We consider the standard architecture of m-LPs, $m = 1, 2, \ldots$, i.e., multi-layered perceptrons with one output layer and $m - 1$ hidden layers; see also [8, 11, 14]. 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 $\theta \colon \mathbb{R}^N \to [0, 1]$. In this paper we only consider the hard limiting response function $\Theta \colon \mathbb{R}^N \to \{0, 1\}$ that satisfies $\Theta(\lambda) = 1$, if $\lambda \geq 0$, and $\Theta(\lambda) = 0$, if $\lambda < 0$. We consider the exact classification of non-finite subsets of $\mathbb{R}^N$ for some fixed $N \in \mathbb{N}$, which implies that the corresponding 2-LPs have N inputs. As usual $\mathbb{R}$ denotes the set of all real numbers, $\mathbb{N} = \{1, 2, 3, \ldots\}$ and $\mathbb{N}_0 = \{0, 1, 2, 3, \ldots\}$.

Let $R_{m,N,K}$ denote the set of all vector functions from $\mathbb{R}^N$ to $\{0, 1\}^K$ that can be associated with an m-LP having N inputs and K outputs, for $N, K \in \mathbb{N}$. We define $R_{m,N,K}$ iteratively. First $R_{1,N,K}$ is defined as

$$R_{1,N,K} = \{f \colon \mathbb{R}^N \to \{0, 1\}^K \mid f_i = \Theta \circ \tilde{f}_i, \ \tilde{f}_i \in \mathcal{A}_N, \ i = 1, \ldots, K\}, \tag{1}$$

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

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

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

$$R_{m+1,N,K} = \{f \colon \mathbb{R}^N \to \{0, 1\}^K \mid f = g \circ h, \ g \in R_{1,L,K}, \ h \in R_{m,N,L}, \ L \in \mathbb{N}\}. \tag{2}$$

Abbreviating $R_{m,N,1}$ as R_m, we then say that V can be classified with an m-LP if there exists an $f \in R_m$ such that $f(x) = 1$, if and only if $x \in V$. Furthermore, we define for each function $f \colon \mathbb{R}^N \to \{0, 1\}$, the set $\mathcal{J}(f) \subseteq \mathbb{R}^N$ by $\mathcal{J}(f) = \{x \in \mathbb{R}^N \mid f(x) = 1\}$. Now it follows that V can be classified with 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 with an m-LP is denoted C_m. It is obvious from the above that $C_m = \{\mathcal{J}(f) \mid f \in R_m\}$. Furthermore, one can easily verify the well-known fact that $C_1 = \{V \subseteq \mathbb{R}^N \mid V \text{ is a closed affine half-space}\} \cup \{\emptyset, \mathbb{R}^N\}$. This implies that if $V \neq \emptyset, \mathbb{R}^N$, then $V \in C_1 \Rightarrow V^* \notin C_1$, where V^* denotes the complement of the set V. Finally, we define the following four collections.

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

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

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

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

$\tilde{C}_1$ is the collection of open and closed affine half-spaces. P denotes the set of all polyhedra, where a polyhedron is defined as the intersection of a finite collection of closed affine half-spaces. $\tilde{P}$ can be viewed as the set of all *pseudo*-polyhedra, where a pseudo-polyhedron is the intersection of a finite collection of closed or open affine half-spaces. A polyhedron has all faces belonging to the set, whereas a pseudo-polyhedron can have faces belonging either to the set or to the complement of the set. U is the set of all subsets of $\mathbb{R}^N$ that can be represented as a union of a finite number of pseudo-polyhedra, which is identical to the set of all subsets that have piece-wise linear bounds. Note that $P \subseteq \tilde{P}$ and $\tilde{C}_1 \subseteq \tilde{P} \subseteq U$.

We use the following basic results; for proofs of these results see [14].

Proposition 1 *If $m \geq 2$, then $V \in C_m \Rightarrow V^* \in C_m$.*

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

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

Proposition 4 *Let $V \subseteq \mathbb{R}^N$, then V can be exactly classified with an m-LP only if $V \in U$.*

Combining Propositions 3 and 4 it follows that $C_m = U$ for all $m \geq 3$, which gives an necessary and sufficient condition for a subset to be classifiable with a m-LP, for all $m \geq 3$. A necessary and sufficient condition that formalizes the classification capabilities of 2-LPs is the subject of the next section.

3 Main results

We concentrate on the characterization of sets that are exactly classifiable by 2-LPs. From Propositions 2 and 4 we know that $\tilde{P} \subseteq C_2 \subseteq U$, a result already known for some time; see [8]. In Section 3.1 we give a necessary condition for a subset to be classifiable with a 2-LP, which will prove that $C_2 \subset U$. In Section 3.2 we present a sufficient condition for a subset to be classifiable with a 2-LP, which will prove that $\tilde{P} \subset C_2$. Finally, in Section 3.3 we introduce some algorithms for the verification of the sufficiency condition of Section 3.1.

3.1 A necessary condition for the existence of a 2-LP

In this section we demonstrate that the condition given in Proposition 4 is not sufficient for classifying a given subset with a 2-LP, i.e. there exist subsets of $\mathbb{R}^N$ that can be classified with a 3-LP but not with a 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 with a 2-LP. In a recent paper, Gibson & Cowan [4] have presented a similar result, which is tailored to the subset presented in Figure 1b. Although their result can be generalized, it is based on an approach that makes no

distinction between closed and open half-spaces in the classification, which is not suitable in our situation. Our formulation of the necessary condition for a subset to be classifiable with a 2-LP requires the existence of a hyperplane, two balls and some other notions, which are defined as follows.

The (open) ball $B_1(x_0, \delta)$, with center $x_0 \in \mathbb{R}^N$ and radius $\delta > 0$, is defined as the subset $\{x \in \mathbb{R}^N \mid \|x - x_0\| < \delta\}$. We use W° to denote the topological interior of a subset W. W° is defined as the set of all points x for which a ball B exists that satisfies $x \in B \subset W$. Furthermore, $\overline{W}$ denotes the topological closure of the subset W, which satisfies $\overline{W} = ((W^*)^\circ)^*$.

We now can express the necessary condition for a subset to be classifiable with a 2-LP by the following theorem.

Theorem 1 *Let $V \subseteq \mathbb{R}^N$, then V can be classified with a 2-LP only if there do not exist a $W \in C_1$ and two balls B_1, B_2, such that*

$$
\left. \begin{array}{ccccc} \emptyset & \neq & B_1 \cap W^\circ & \subseteq & V \\ \emptyset & \neq & B_1 \cap W^* & \subseteq & V^* \end{array} \right\} \land \left\{ \begin{array}{ccccc} \emptyset & \neq & B_2 \cap W^\circ & \subseteq & V^* \\ \emptyset & \neq & B_2 \cap W^* & \subseteq & V. \end{array} \right. \tag{7}
$$

Proof Assume that $V \in C_2$, $W \in C_1$ and two balls B_1, B_2 exist that satisfy (7), 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 $g \in R_{1,K,1}$, $h \in R_{1,N,K}$, $K \in \mathbb{N}$, and we may assume that $h_i \neq h_j$, for all $1 \leq i < j \leq K$. Let $g = \Theta \circ \tilde{g}$ with $\tilde{g}(w) = a \cdot w + b$ for some $a \in \mathbb{R}^K$, $b \in \mathbb{R}$ and define $V_i = \mathcal{J}(h_i)$, $i = 1, \ldots, K$.

From (7) it follows that $W \neq \emptyset, \mathbb{R}^N$, which means that W is a closed affine half-space. Let $\hat{W} = W \cap (W^\circ)^*$ be the hyperplane that bounds both W and $(W^\circ)^*$, and 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 $B_1 \cap \hat{W} \subseteq$ 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 (i) and (ii) lead to a contradiction. This implies that the same holds for (iii) and (iv), as they can be obtained from (ii) and (i), respectively, by interchanging V and V^*, and using Proposition 1. To prove that (i) and (ii) lead to a contradiction, we need the following lemma. As with all lemmas in this section, the reader is referred to the appendix for the proof of this lemma.

4

Lemma 1 *Let $V = \mathcal{J}(f)$, for some $f \in R_{2,N,1}$, $f = g \circ h$, with $g \in R_{1,K,1}$, $h \in R_{1,N,K}$, $K \in \mathbb{N}$, and $h_i \neq h_j$, for all $1 \leq i < j \leq K$. Let $g = \Theta \circ \tilde{g}$ with $\tilde{g}(w) = a \cdot w + b$ for some $a \in \mathbb{R}^K$, $b \in \mathbb{R}$ and define $V_i = \mathcal{J}(h_i)$, $i = 1, \ldots, K$. If there exists a $W \in C_1$ and a ball B such that*

$$\begin{cases} \emptyset & \neq & B \cap W & \subseteq & V \\ \emptyset & \neq & B \cap W^* & \subseteq & V^*, \end{cases} \tag{8}$$

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

(i). In this case (7) assumes the following form

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

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

(ii). Now (7) becomes:

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

Applying Lemma 1 to B_1 and W, we obtain $W = V_i$, for some $i \in \{1, ..., K\}$ and $a_i > 0$. Without loss of generality we may 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 Lemma 1 to V^*, W and B_2, we obtain $W = V_j$, for some $j \in \{1, ..., K\}$ and $-a_j > 0$. This would yield a pair $i \neq j$ with $V_i = V_j$, contradicting the assumptions. $\qquad \square$

In the above theorem the subset W corresponds to a closed affine half-space, i.e. $W = \{x \in \mathbb{R}^N \mid a \cdot x + b \geq 0\}$ for some $a \in \mathbb{R}^N \setminus \{0\}$ 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, belong to V or V^*. The conditions are only concerned with parts of the open affine half-spaces $W^\circ = \{x \mid a \cdot x + b > 0\}$ and $W^* = \{x \mid a \cdot x + b < 0\}$.

Suppose that $V \subseteq \mathbb{R}^N$ satisfies the conditions of Theorem 1 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 generalized exclusive-or problem. Theorem 1 proves that $f \notin C_2$. Hence, the condition in (7) can be viewed as a generalization of the condition of linear separability for a subset to be classifiable with a 1-LP, since this condition is responsible for the non-existence of a 1-LP for the exclusive-or problem. In Figures 1a and 1b we present two examples of subsets in $\mathbb{R}^2$ that do not belong to C_2. The subsets of $\mathbb{R}^2$ given in Figures 1a and 1b cannot be classified by a 2-LP since these subsets do not satisfy the conditions of Theorem 1; the circles correspond to the balls B_1 and B_2. Although the subset presented in Figure 1c satisfies
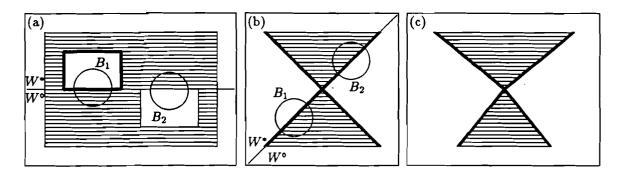
Figure 1: *Three subsets that cannot be classified exactly by a 2-LP. Note that solid boundary lines do and thin boundary lines do not belong to the presented sets.*

the conditions of Theorem 1, it can be shown that this subset cannot be classified exactly by a 2-LP, using a proof similar to that of Theorem 1; see also below.

As already mentioned in the introduction, the results of Cybenko [1] and others show that a subset $V \subseteq \mathbb{R}^N$ can be approximately classified with 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 $\|V \div V_\varepsilon\| < \varepsilon$, where $\div$ denotes the set theoretical symmetrical difference and $\|\cdot\|$ is a norm defined on the set of subsets of $\mathbb{R}^N$, for instance $\|V\| = \int_V dV$. 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 denotes the number of hidden nodes. If $V \in U$ satisfies the conditions of Theorem 1, then $V \notin C_2$ and it is intuitively clear that we must have $\lim_{\varepsilon \downarrow 0} L_\varepsilon = \infty$ in that case. For the two subsets in Figures 1a and 1b we have found approximating subsets $V_\varepsilon \in C_2$ with $L_\varepsilon \approx |\log \varepsilon|$ and $L_\varepsilon \approx 1/\varepsilon$, respectively. In Figure 3a, presented in Section 3.3, an example is given of a subset in C_2 that approximates the subset given in Figure 1b. The general upper-bound $L_\varepsilon = \mathcal{O}(\varepsilon^{-(N-1)/2})$ given by Cybenko [2] indicates that the result for the subset in Figure 1b can be improved upon.

Finally, we note that the condition given in Theorem 1 is not sufficient for a subset in U to be classifiable with a 2-LP. This follows from the fact that there exist subsets in U that satisfy the conditions of Theorem 1 but cannot be classified with a 2-LP. An example of such a subset is presented in Figure 1c. The proof that the subset in Figure 1c is not classifiable runs along the lines of the proof of Theorem 1. Unfortunately we have not been able to generalize it to a similar general condition as in (7). In the next section we therefore approach the problem of finding a characterization of C_2 from the other side by deriving a sufficient condition for a subset to be classifiable with a 2-LP.

3.2 A sufficient condition for the existence of a 2-LP

The examples given in the previous section show that $C_2 \subset U$, i.e., a strict inclusion. In this section we consider the problem of proving that $\bar{P} \subset C_2$. To the best of our knowledge the literature presents only few examples of subsets in C_2 that are not in $\bar{P}$; see for instance [6, 9, 13]. One such example [9] is shown in Figure 3a and can be proved to belong to C_2 by using

6

Theorem 2, which is given below; see also the next section. This theorem presents a sufficient condition for a subset to be classifiable with a 2-LP. It is the most general sufficient condition we found so far, based on the classifiability of the intersection of two classifiable subsets with a 2- LP. Moreover, it is the only sufficient condition for which we have found an algorithmic verification method. This verification method is presented in the next subsection.

In the following theorem we use $V_1 \backslash V_2 \backslash \cdots \backslash V_l$ as a shorthand notation for $V_1 \backslash (V_2 \backslash (\cdots (V_{l-1} \backslash V_l) \cdots))$.

Theorem 2 *Let $V = V_1 \backslash V_2 \backslash \cdots \backslash V_l$, for some $l \in \mathbb{N}$ and $V_1, V_2, \ldots, V_l \in \check{P}$, then $V \in C_2$.*

Proof For the proof we need the following two lemmas.

Lemma 2 *If $V \in \check{P}$, then there are $g = \Theta \circ \tilde{g} \in R_{1,k,1}$, $h \in R_{1,N,k}$ and $k \in \mathbb{N}$, such that $V = \mathcal{J}(g \circ h)$ and $\tilde{g}(h(x)) \in \{-k, -k+1, \ldots, 0\}$, for all $x \in \mathbb{R}^N$.*

Lemma 3 *Let $V_1 \in \check{P}$, $V_2 \in C_2$ and assume that $V_i = \mathcal{J}(g_i \circ h_i)$, $g_i = \Theta \circ \tilde{g}_i \in R_{1,k_i,1}$, $h_i \in R_{1,N,k_i}$, $\tilde{g}_1(h_1(x)) \in \{-k_1, -k_1+1, \ldots, 0\}$ and $\tilde{g}_2(h_2(x)) \in \{-\alpha, -\alpha+1, \ldots, \beta\}$, for some $k_1, k_2 \in \mathbb{N}$, $\alpha \in \mathbb{N}$, $\beta \in \mathbb{N}_0$, and for all $x \in \mathbb{R}^N$. Then $V_1 \backslash V_2 = \mathcal{J}(g \circ h)$, 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*

$$\tilde{g}(w_1, w_2) = \alpha \tilde{g}_1(w_1) - \tilde{g}_2(w_2) - 1, \tag{9}$$

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

Now we can prove the main result. Let $l \in \mathbb{N}$, $V_i \in \check{P}$, $i = 1, \ldots, l$, then, using Lemma 2, we may assume that $V_i = \mathcal{J}(g_i \circ h_i)$, $g_i = \Theta \circ \tilde{g}_i \in R_{1,k_i,1}$, $h_i \in R_{1,N,k_i}$ and $\tilde{g}_i(h_i(x)) \in \{-k_i, -k_i+1, \ldots, 0\}$ for some $k_i \in \mathbb{N}$ and all $x \in \mathbb{R}^N$. We complete the proof by showing that $V_1 \backslash V_2 \backslash \cdots \backslash V_l = \mathcal{J}(g \circ h)$, where $h \in R_{1,N,\Sigma k_i}$ and $g = \Theta \circ \tilde{g} \in R_{1,\Sigma k_i,1}$ are given by $h(x) = (h_1(x), \ldots, h_l(x))$ and

$$\tilde{g}(w_1, \ldots, w_l) = \sum_{i=1}^{l} (-1)^{i-1} \alpha_{l-i} \tilde{g}_i(w_i) - \frac{1}{2}((-1)^l + 1), \tag{10}$$

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

The proof is by induction. The result obviously holds for $l = 1$. Next, we assume that the result holds for some $l \in \mathbb{N}$, and show that it also holds for $l + 1$. If (10) holds for l, then by renumbering the variables we find that $V_2 \backslash V_3 \backslash \cdots \backslash V_{l+1} = \mathcal{J}(g \circ h)$, where $h \in R_{1,N,\Sigma k_{i+1}}$ and $g = \Theta \circ \tilde{g} \in R_{1,\Sigma k_{i+1},1}$ are given by $h(x) = (h_2(x), \ldots, h_{l+1}(x))$ and

$$\tilde{g}(w_2, \ldots, w_{l+1}) = \sum_{i=2}^{l+1} (-1)^i \alpha_{l+1-i} \tilde{g}_i(w_i) - \frac{1}{2}((-1)^l + 1), \tag{11}$$

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

7

Next, we apply Lemma 3 to $V_1\backslash V_2\backslash\cdots\backslash V_{l+1} = V_1\backslash(V_2\backslash\cdots\backslash V_{l+1})$, and find that $V_1\backslash V_2\backslash\cdots\backslash V_l = \mathcal{J}(e\circ f)$, where $f\in R_{1,N,\Sigma k_i}$ and $e = \Theta\circ\tilde{e}\in R_{1,\Sigma k_i,1}$ are given by $f(x) = (h_1(x),h(x)) = (h_1(x),h_2(x),\ldots,h_{l+1}(x))$ and

$$
\begin{aligned}
\tilde{e}(w_1,w_2,\ldots,w_{l+1}) &= \alpha_l\tilde{g}_1(w_1) - \tilde{g}(w_2,\ldots,w_{l+1}) - 1 \\
&= \alpha_l\tilde{g}_1(w_1) - \sum_{i=2}^{l+1}(-1)^i\alpha_{l+1-i}\tilde{g}_i(w_i) + \frac{1}{2}((-1)^l + 1) - 1 \\
&= \sum_{i=1}^{l+1}(-1)^{i-1}\alpha_{l+1-i}\tilde{g}_i(w_i) - \frac{1}{2}((-1)^{l+1} + 1)
\end{aligned}
$$

respectively, with $\tilde{e}(f(x)) \in \{-k_1\alpha_l - \alpha_{l-1},\ldots,\alpha_l - 1\}$, for all $x\in\mathbb{R}^N$. By defining $\alpha_{l+1} = k_1\alpha_l + \alpha_{l-1}$, we have proved the result for $l+1$. This completes the proof of the theorem. $\square$

From Theorem 2 it follows that the required number of hidden units of a 2-LP that classifies $V_1\backslash V_2\backslash\cdots\backslash V_l$ is at most $\sum_{i=1}^l k_i$, which equals the total number of half-spaces defining the subsets $V_1,\ldots,V_l$. If these subsets have defining half-spaces in common the required number of hidden units can be reduced accordingly. Although the proof of Theorem 2 contains an algorithm for the determination of a set of weights for a 2-LP that classifies a subset of the given form, the practical value of this algorithm is limited since these weights can become very large $(\alpha_l \approx k_1k_2\cdots k_l)$.

Verifying whether a subset V satisfies the conditions of Theorem 2 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 subsection.

3.3 A decomposition algorithm

In the previous subsection, a sufficient condition is given for a subset of $\mathbb{R}^N$ to be classifiable with a 2-LP; see Theorem 2. 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 $V_1,\ldots,V_l\in P$ such that $V = V_1\backslash V_2\backslash\cdots\backslash V_l$, if such a decomposition of V exists. The decomposition is based on the following ideas.

Assume that $V = V_1\backslash V_2$, for some *unknown* subsets $V_1,V_2\in\tilde{P}$. In the search for the unknown sets V_1 and V_2, we use that $V = V'\backslash(V'\backslash V)$, if and only if $V'\supseteq V$. Hence, we only need a $V'\supseteq V$ satisfying $V'\in\tilde{P}$ and $V'\backslash V\in\tilde{P}$. To find such a V' we exploit the property that all subsets in $\tilde{P}$ are convex, which implies that V' and $V'\backslash V$ must be convex. Using that $V'\backslash V = V'\backslash(V_1\backslash V_2) = (V'\backslash V_1)\cup(V'\cap V_2)$, we find that $V'\backslash V$ 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 $\text{conv}(V)$. Of course, the convexity of V' and $V'\backslash V$ does not necessarily guarantee that $V'\in\tilde{P}$ and $V'\backslash V\in\tilde{P}$. In Figure 2 below we give an example where the use of convex-hull does not suffice.
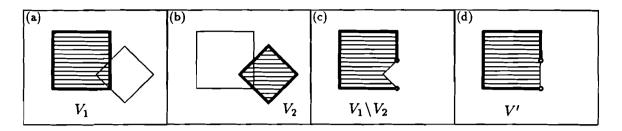
Figure 2: *Example demonstrating the necessity of using the closure of the convex-hull. Note that thin boundary lines and circles are not part of the given subsets.*

In Figures 2a and 2b, two subsets $V_1, V_2 \in P$ are given. Suppose that we want to decompose $V = V_1 \backslash V_2$, which is shown in Figure 2c. To this end we construct $V' = \text{conv}(V)$; see Figure 2d. Note that $V' \notin \tilde{P}$. The only way to ensure that $V' \in \tilde{P}$ is to take $V' = \overline{\text{conv}}(V)$, with $\overline{\text{conv}}(V)$ denoting the closure of the convex hull of V, in which case we have a stronger result, namely $V' \in P$. In Theorem 6 presented at the end of this section, we prove that $\overline{\text{conv}}(V) \in P$ for all $V \in U$. A disadvantage of using $\overline{\text{conv}}$ instead of conv is that we have to restrict ourselves to sets in $\mathbb{R}^N$ that have a decomposition consisting of subsets exclusively in P, for one can easily construct a $V = V_1 \backslash V_2$, with $V_1, V_2 \in \tilde{P}$, for which there do not exist $W_1, W_2 \in P$ such that $V = W_1 \backslash W_2$. In the following theorem the above ideas are used to prove the correctness of a DEComposition ALgorithm that calculates the decomposition of a given subset.

Theorem 3 *If for a given subset $V \subseteq \mathbb{R}^N$ there exists a decomposition of the form $V = W_1 \backslash W_2 \backslash \cdots \backslash W_k$ for some $k \in \mathbb{N}$ and $W_1, \ldots, W_k \in P$, then it can be found using the following DECAL-1 algorithm.*

```
program DECAL-1;
begin
    l := 0;
    T := V;
    while T ≠ ∅
    do
        V_{l+1} := conv(T);
        T := V_{l+1} \ T;
        l := l + 1
    od
end.
```

Proof Let $V \subseteq \mathbb{R}^N$ and $V = W_1 \backslash W_2 \backslash \cdots \backslash W_k$, for some $k \in \mathbb{N}$ and $W_1, \ldots, W_k \in P$. Then, using the equality $T = \overline{\text{conv}}(T) \backslash (\overline{\text{conv}}(T) \backslash T)$ discussed above, one can easily show that $V = V_1 \backslash V_2 \backslash \cdots \backslash V_l \backslash T$ is an invariant of DECAL-1. Furthermore, since $V = W_1 \backslash W_2 \backslash \cdots \backslash W_k$ and $W_1, \ldots, W_k \in P$, we have $V \in U$. Using the result of Theorem 6 below, which says that $\overline{\text{conv}}(T) \in P$, if $T \in U$, one can easily show that $T \in U$ and $V_l \in P$ are also invariants of DECAL-1. Hence, it remains to be proved that the algorithm terminates, which follows directly from the following lemma.

9

Lemma 4 *Let V be a subset of $\mathbb{R}^N$. If $V = W_1 \backslash W_2 \backslash \cdots \backslash W_k$ for some $k \in \mathbb{N}$ and $W_1, \ldots, W_k \in P$, then $l \leq k$ is an invariant of DECAL-1.*
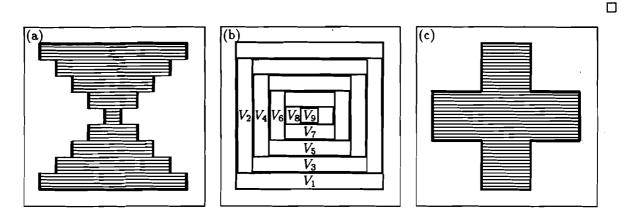
□

Figure 3: *In (b) the results of the decomposition algorithm are given, when it is applied to the subset in (a). In (c) a subset is given that can be classified with a 2-LP, for which the algorithm does not terminate because it has no decomposition of the described form.*

Figure 3b presents the intermediate steps of DECAL-1 when it is applied to the subset V of Figure 3a. We obtain $V = V_1 \backslash V_2 \backslash \cdots \backslash V_9$, where $V_1, V_2, \ldots, V_9$ are all rectangles obviously belonging to P. Since most of the example subsets presented in [6] have a decomposition of the proposed form, DECAL-1 can be used to find this decomposition for these subsets. In some sense, the use of Theorem 3 seems a bit paradoxical, because one can only compute the decomposition of a subset if it is known to exist, and the latter can only be guaranteed by giving this decomposition. However, DECAL-1 can also be used to find out whether a given subset $V \in U$ can be decomposed. If the algorithm terminates when applied to V, then it follows from the proof of Theorem 3 that a decomposition is found. If on the other hand the algorithm does not terminate, then Lemma 4 implies that there does not exist a decomposition of V.

The remaining problem is to determine whether the algorithm terminates. There is no general method known that does this, but we argue that in our case we can solve the problem. When the algorithm is applied to $V \in U$, then one can easily show that $V_{l+1} \subseteq V_l$ for all $l \in \mathbb{N}$; see the proofs of Lemma 4 and Theorem 5. We claim that the algorithm does not terminate, if and only if $V_{l+1} = V_l$, for some $l \in \mathbb{N}$, or in other words, the situation where $V_{l+1} \subset V_l$, for all $l \in \mathbb{N}$ cannot occur. The first part of this claim is expressed in the following theorem.

Theorem 4 *Let DECAL-1 be applied to a subset $V \in U$. Then the algorithm does not terminate if at some stage k in the execution of the algorithm V_{k+1} equals V_k.*

Proof Let T_l denote the set T at stage l of the algorithm, then $V_{l+1} = \overline{\text{conv}}(T_l)$ and $T_{l+1} = V_{l+1} \backslash T_l$. Assume that $V_{k+1} = V_k$ at stage k of the algorithm, then we have that

$$
\begin{aligned}
T_{k+1} &= V_{k+1} \backslash T_k \\
&= V_k \backslash (V_k \backslash T_{k-1})
\end{aligned}
$$

10

$$= V_k \cap T_{k-1}$$
$$= T_{k-1},$$

using that $V_k = \overline{\mathrm{conv}}(T_{k-1}) \supseteq T_{k-1}$. Since the algorithm reaches stage k, it follows that $T_l \neq \emptyset$ for all $l \leq k$. Hence, $T_{k+1} \neq \emptyset$ and the algorithm reaches stage $k+1$, where it calculates $V_{k+2} = \overline{\mathrm{conv}}(T_{k+1}) = \overline{\mathrm{conv}}(T_{k-1}) = V_k = V_{k+1}$. Using the same argument in an iterative way, we find that $T_{l+1} = T_{l-1} \neq \emptyset$ and $V_{l+1} = V_l$ for all $l \geq k$, which proves that the algorithm does not terminate $\qquad\square$

Although we believe that the opposite of Theorem 4 is also true, we have not been able to proof this. We therefore present it as a conjecture with a sketch of a possible proof.

Conjecture 1 *Let DECAL-1 be applied to a subset $V \in U$. Then the algorithm does not terminate only if at some stage k in the execution of the algorithm V_{k+1} equals V_k.*

A proof of the above conjecture might be constructed along the following lines. Since $V \in U$, V is defined using a number, say n, of affine half-spaces. Assume that the algorithm when applied to V does not terminate, then at stage $n+2$ we have that $V = V_1 \backslash V_2 \backslash \cdots \backslash V_{n+2} \backslash T$ with $V_1 \supseteq V_2 \supseteq \cdots \supseteq V_{n+2} \supseteq T$. To complete the proof we must show that $V_1 \supset V_2 \supset \cdots \supset V_{n+2}$ leads to a contradiction. The first step is to observe that if $V_i, V_{i+1} \in P$ and $V_i \supset V_{i+1}$, then one of the half-spaces that defines V_i is not needed for defining V_{i+1}. The essential part of the proof is then to show that this implies that $n+1$ distinct half-spaces can be chosen, one for every pair V_i, V_{i+1}, that are all needed for the definition of V.

If correct, the above argument would imply that if DECAL-1 does not terminate by itself, then after at most $n+1$ iterations of the loop the program can be stopped since one may conclude that it will never terminate. As an example of a subset for which the algorithm does not terminate we present the subset in Figure 3c. One can easily verify that for this subset we obtain $V_2 = V_1$, which implies that $V_2 = V_3 = V_4 = \ldots$, and proves that this subset has no decomposition of the proposed form. However, this subset can be classified with a 2-LP, which proves that the sufficient condition of Theorem 2 is not a necessary condition.

So far, we have not discussed the execution of the different steps in DECAL-1. Especially the calculation of $\overline{\mathrm{conv}}(T)$ and $V_{l+1} \backslash T$ are non-trivial and time consuming. We first concentrate on the calculation of $V_{l+1} \backslash T = V_{l+1} \cap T^*$. The determination of $\overline{\mathrm{conv}}(T)$ is discussed at the end of this subsection. The following theorem shows that the calculation of T^* on every iteration of the loop can be replaced by a single calculation of V^* at the beginning of the algorithm.

Theorem 5 *The DECAL-1 algorithm presented in Theorem 3 can be replaced by the following DECAL-2 algorithm.*

```
program DECAL-2;
begin
    l := 0;
    T := V;
    while T ≠ ∅
    do
```

11

$$V_{l+1} := \overline{\mathrm{conv}}(T);$$
$$T := V_{l+1} \cap (-1)^{l+1}V$$
$$l := l + 1$$
od
end.

where $+V = V$ and $-V = V^*$.

Proof First we show that $V_{l+1} \subseteq V_l$, for all $l \in \mathbb{N}$. Let $V_0 = \mathbb{R}^N$ then $T \subseteq V_l$ is an invariant of DECAL-1. From the proof of Theorem 3 we know that $V_l \in P$ is also an invariant. Combining these two invariants it follows that $V_{l+1} = \overline{\mathrm{conv}}(T) \subseteq \overline{\mathrm{conv}}(V_l) = V_l$ for all $l \in \mathbb{N}_0$. If we assume that $T = V_l \cap (-1)^l V$ for some $l \in \mathbb{N}_0$, which holds for $l = 0$, then

$$
\begin{aligned}
V_{l+1} \backslash T &= V_{l+1} \cap T^* \\
&= V_{l+1} \cap (V_l \cap (-1)^l V)^* \\
&= V_{l+1} \cap (V_l^* \cup (-1)^{l+1} V) \\
&= V_{l+1} \cap (-1)^{l+1} V,
\end{aligned}
$$

which proves that $T = V_l \cap (-1)^l V$ is an invariant of DECAL-1. $\qquad\square$

The last part of this section considers the calculation of $\overline{\mathrm{conv}}(T)$. Theorem 6 presents a systematic method for the calculation of $\overline{\mathrm{conv}}(V)$, for any subset $V \in U$. Before we can explain the different steps of this method, we need the following elementary results.

The convex-hull of a subset $V \subseteq \mathbb{R}^N$ is defined as

$$\mathrm{conv}(V) = \{x \in \mathbb{R}^N \mid x = \textstyle\sum_{i=1}^r \lambda_i v_i,\ v_i \in V,\ \lambda_i \geq 0, \sum_{i=1}^r \lambda_i = 1,\ r \geq 1\}. \tag{12}$$

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 [12]. Hence, we have that

$$\mathrm{cone}(V) = \{x \in \mathbb{R}^N \mid x = \textstyle\sum_{i=1}^r \lambda_i v_i,\ v_i \in V,\ \lambda_i \geq 0\}. \tag{13}$$

An elementary result in linear algebra states that every polyhedron can be written as the sum of a bounded convex-hull and a cone; see Lemma 5 below. We use $\{x_i\}_{i=1}^k$ to denote the set $\{x_1, \ldots, x_k\}$, and by definition we have $\mathrm{conv}(\emptyset) = \emptyset$ and $\mathrm{cone}(\emptyset) = \mathrm{cone}(\{0\}) = \{0\}$.

Lemma 5 *Let $V \subseteq \mathbb{R}^N$, then $V \in P$, if and only if*

$$V = \mathrm{conv}(\{x_i\}_{i=1}^k) + \mathrm{cone}(\{y_i\}_{i=1}^p),$$

for some $k, p \in \mathbb{N}_0$ 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 6 *Let $V \in \tilde{P}$, then $\overline{V} \in P$.*

We are now ready for the final result of this paper. Theorem 6 proves that the closure of the convex-hull of a subset in U is always a polyhedron. Furthermore, the theorem presents a method for the determination of this polyhedron. The method consists of the following 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 6 and, hence, using Lemma 5 this yields the second step. The third step follows by using Lemma 5 in the opposite way. The proof of the final result is straightforward, using that $\overline{\bigcup_{i=1}^{k} V_i} = \bigcup_{i=1}^{k} \overline{V_i}$, except that we have to be careful to note that there exist $W_1, W_2 \in P$ for which $\mathrm{conv}(W_1 \cup W_2) \notin P$.

Theorem 6 *Let $V \in U$, then*

    (i) $\overline{\mathrm{conv}}(V) \in P$,

    (ii) $V' = \overline{\mathrm{conv}}(V)$ *can be calculated using the following three steps*

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

        2. Determine $k_i, p_i \in \mathbb{N}_0$ and $x_{ij}, y_{ij} \in \mathbb{R}^N$ such that

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

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

Proof (i) Follows directly from (ii) and Lemma 5.

(ii) Let $V \in U$, then $V = \bigcup_{i=1}^{l} V_i$ for some $l \in \mathbb{N}$ and $V_i \in \tilde{P}$. By Lemma 6 we have that $\overline{V}_i \in P$, and hence, by Lemma 5, it follows that

$$\overline{V}_i = \mathrm{conv}(\{x_{ij}\}_{j=1}^{k_i}) + \mathrm{cone}(\{y_{ij}\}_{j=1}^{p_i}),$$

for some $k_i, p_i \in \mathbb{N}$ and $x_{ij}, y_{ij} \in \mathbb{R}^N$. By combining this with the fact that

$$\overline{\mathrm{conv}}(V) = \overline{\mathrm{conv}}(\bigcup_{i=1}^{l} V_i) = \overline{\mathrm{conv}}(\bigcup_{i=1}^{l} \overline{V}_i),$$

the result follows from the following lemma.

Lemma 7 *Let $V_1, V_2, \ldots, V_l \in P$ for some $l \in \mathbb{N}$ and $V_i = \mathrm{conv}(\{x_{ij}\}_{j=1}^{k_i}) + \mathrm{cone}(\{y_{ij}\}_{j=1}^{p_i})$, $i = 1, \ldots, l$. Then $\overline{\mathrm{conv}}(\bigcup_{i=1}^{l} V_i) = \mathrm{conv}(\{x_{ij}\}_{i=1,j=1}^{l,\,k_i}) + \mathrm{cone}(\{y_{ij}\}_{i=1,j=1}^{l,\,p_i}).$*

                                                                     $\square$

4 Discussion and concluding remarks

Suppose we want to solve a given combinatorial optimization problem with a two-layered perceptron. This means that we have to find a 2-LP that classifies the subset $V = V_1^{(i)}$ for fixed i; see [14] for a definition of $V_1^{(i)}$. This can be done along the following lines. First use DECAL-2 to find a decomposition $V = V_1 \setminus V_2 \setminus \cdots \setminus V_l$ of V for some $l \in \mathbb{N}$ if it exists. Note that we have an explicit expression for both $+V = V$ and $-V = V^*$ in this case; see [14]. If a decomposition is found, we can use the algorithm described in the proof of Theorem 2 to find the weights of the corresponding 2-LP. However, due to the exponential time complexity of the calculation of $\overline{\text{conv}}(V)$ as described by Theorem 6, the practical use of this approach is limited to 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 in a direct algebraic way.

We have used the above approach to find a 2-LP for solving the Dynamic Lot-sizing Problem introduced by Wagner & Whitin; see Zwietering, Aarts & Wessels [15]. For some time we believed that the classification problems corresponding to this combinatorial optimization problem could be decomposed in the form given by Theorem 2. When we implemented the decomposition algorithm described in Section 3.3 and applied it to the subsets corresponding to the classification problems, we observed that the decomposition algorithm did not terminate. This proves that the decomposition does not exist, but it does not necessarily implies that the problem cannot be solved with a 2-LP. However, a careful examination shows that the subsets corresponding to the Dynamic Lot-sizing Problem have a structure similar to the one presented in Figure 3c, which can be proved not to be classifiable with a 2-LP, as indicated in Section 3.1. Consequently, the Dynamic Lot-sizing Problem cannot be solved with a 2-LP.

The main part of this paper discussed the classification capabilities of 2-LPs. A detailed analysis was used to find necessary and sufficient conditions for a subset to be classifiable with a 2-LP. The necessary conditions can be used to show that a given problem cannot be solved exactly with a 2-LP. One such problem is the sorting problem discussed in [16]. Training a 2-LP to solve such a problem is bound to give a poor result, in the sense that the learning algorithm cannot converge to the optimal solution, and one should therefore consider using a 3-LP in this case. The sufficient conditions can be used to prove that a problem can be solved exactly with 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 variation in the size of the weights implies that the use of a learning algorithm is sometimes more useful for the determination of the weights.

Acknowledgement

References

[1] G. CYBENKO, *Approximation by superpositions of a sigmoidal function*, Tech. Rep. No. 856, Univ. of Illinois, 1989.

[2] G. CYBENKO, Complexity theory of neural networks and classification problems, *Proceedings of the 1990 EURASIP Workshop on Neural Networks*, Sesimbra, Portugal, 26–44, 1990.

[3] K. FUNAHASHI, On the approximate realization of continuous mappings by neural networks, *Neural Networks* 2, 183–192, 1989.

[4] G.J. GIBSON AND C.F.N. COWAN, On the decision regions of multilayer perceptrons, *Proceedings IEEE* 78, 1590–1594, 1990.

[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*, American Institute of Physics, New York, 387–396, 1988.

[7] L.K. LI, On computing decision regions with neural nets, *Journal of Computer and System Sciences* 43, 509–512, 1991.

[8] R.P. LIPPMANN, An Introduction to computing with neural nets, *IEEE Acoustics, Speech and Signal Processing Magazine* 4, 4–22, 1987.

[9] J. MAKHOUL, R. SCHWARTZ, AND A. EL-JAROUDI, Classification capabilities of two-layer neural nets, *Proceedings of IEEE Int. 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. SCHRIJVER, *Theory of Linear and Integer Programming*, Wiley, 1986.

[13] A. WIELAND AND R. LEIGHTON, Geometric analysis of neural network capabilities, *Proceedings of the 1st Int. Conference on Neural Networks*, IEEE, III, 385–393, 1987.

[14] P.J. ZWIETERING, E.H.L. AARTS AND J. WESSELS, The design and complexity of exact multi-layered perceptrons, *Int. Journal of Neural Systems* 2, 185–199, 1991.

[15] P.J. ZWIETERING, E.H.L. AARTS AND J. WESSELS, Neural networks and production planning, *Proceedings Neuro Nimes '92 Int. Conference on Neural Networks*, Nimes, 529–542, 1991.

[16] P.J. ZWIETERING, E.H.L. AARTS AND J. WESSELS, *The minimal number of layers of a perceptron that sorts*, Memorandum COSOR 92-06, Eindhoven Univ. of Technology, 1992. Accepted for publication in the Journal of Parallel and Distributed Processing.

Appendix

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

Lemma 1 *Let $V = \mathcal{J}(f)$, for some $f \in R_{2,N,1}$, $f = g \circ h$, with $g \in R_{1,K,1}$, $h \in R_{1,N,K}$, $K \in \mathbb{N}$, and $h_i \neq h_j$, for all $1 \leq i < j \leq K$. Let $g = \Theta \circ \tilde{g}$ with $\tilde{g}(w) = a \cdot w + b$ for some $a \in \mathbb{R}^K$, $b \in \mathbb{R}$ and define $V_i = \mathcal{J}(h_i)$, $i = 1, \ldots, K$. If there exists a $W \in C_1$ and a ball B such that*

$$
\begin{cases}
\emptyset & \neq & B \cap W & \subseteq & V \\
\emptyset & \neq & B \cap W^* & \subseteq & V^*,
\end{cases}
\tag{14}
$$

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

Proof Let $W = \{x \in \mathbb{R}^N \mid p \cdot x + q \geq 0\}$ for some $p \in \mathbb{R}^N \backslash \{0\}$ 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 $f(x) = \Theta(\tilde{g}(h(x))) = \Theta(0 h_+(x) + 0 h_-(x) + \sum_{i=1}^{K} a_i h_i(x) + b)$, we may assume without loss of generality that $h_1 = h_+$, $h_2 = h_-$ and $h_i \neq h_+, h_-$ for $i = 3, \ldots, K$. Let $h_i(x) = \Theta(c_i \cdot x + d_i)$ for some $c_i \in \mathbb{R}^N \backslash \{0\}$ and $d_i \in \mathbb{R}$, with $(c_i, d_i) \neq \lambda(p, q)$ for all $\lambda \in \mathbb{R}$ and $i = 3, \ldots, K$.

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

$$
B(x_i, \varepsilon_i) \subseteq B(x_{i-1}, \varepsilon_{i-1}) \cap \{x \in \mathbb{R}^N \mid c_i \cdot x + d_i \neq 0\}, \quad \text{for } i = 3, \ldots, K.
\tag{15}
$$

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 $i \in \{3, \ldots, K\}$, for which the ball $B(x_{i-1}, \varepsilon_{i-1})$ satisfies the above conditions. We then construct $B(x_i, \varepsilon_i)$ satisfying (15). 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 (15) 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 $i = 3, \ldots, K$. 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 (14), 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 $\tilde{g}(h(x_K)) \geq 0, \tilde{g}(h(y)) \geq 0$ and $\tilde{g}(h(z)) < 0$. Finally, by using $B(x_K, \varepsilon_K) \subseteq \bigcap_{i=3}^{K}\{x \in \mathbb{R}^N \,|\, c_i \cdot x + d_i \neq 0\}$ one can show that $h_i(x_K) = h_i(y) = h_i(z)$ for $i = 3, \ldots, K$. 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. □

Lemma 2 *If $V \in \tilde{P}$, then there are $g = \Theta \circ \tilde{g} \in R_{1,k,1}$, $h \in R_{1,N,k}$ and $k \in \mathbb{N}$, such that $V = \mathcal{J}(g \circ h)$ and $\tilde{g}(h(x)) \in \{-k, -k+1, \ldots, 0\}$, for all $x \in \mathbb{R}^N$.*

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

Lemma 3 *Let $V_1 \in \tilde{P}$, $V_2 \in C_2$ and assume that $V_i = \mathcal{J}(g_i \circ h_i)$, $g_i = \Theta \circ \tilde{g}_i \in R_{1,k_i,1}$, $h_i \in R_{1,N,k_i}$, $\tilde{g}_1(h_1(x)) \in \{-k_1, -k_1+1, \ldots, 0\}$ and $\tilde{g}_2(h_2(x)) \in \{-\alpha, -\alpha+1, \ldots, \beta\}$, for some $k_1, k_2 \in \mathbb{N}$, $\alpha \in \mathbb{N}$, $\beta \in \mathbb{N}_0$, 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*

$$\tilde{g}(w_1, w_2) = \alpha \tilde{g}_1(w_1) - \tilde{g}_2(w_2) - 1, \tag{16}$$

respectively. Furthermore $\tilde{g}(h(x)) \in \{-k_1\alpha - \beta - 1, \ldots, \alpha - 1\}$ 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 $x \notin V_2$.

It remains to show that $\tilde{g}(h(x)) \in \{-k_1\alpha - \beta - 1, \ldots, \alpha - 1\}$ for all $x \in \mathbb{R}^N$, which follows directly from (16). □

Lemma 4 *Let V be a subset of $\mathbb{R}^N$. If $V = W_1 \setminus W_2 \setminus \cdots \setminus W_k$ for some $k \in \mathbb{N}$ and $W_1, \ldots, W_k \in P$, then $l \leq k$ is an invariant of DECAL-1.*

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

$$
\begin{aligned}
Z_k &= \emptyset, \\
Z_i &= W_{i+1} \setminus Z_{i+1}, & (i = k-1, \ldots, 0), & \tag{17} \\
\tilde{V}_0 &= \mathbb{R}^N, \\
\tilde{V}_i &= \overline{\mathrm{conv}}(\tilde{V}_{i-1} \cap Z_{i-1}), & (i = 1, \ldots, k), & \tag{18} \\
T_0 &= V, \\
T_i &= \tilde{V}_i \setminus T_{i-1}, & (i = 1, \ldots, k). & \tag{19}
\end{aligned}
$$

17

Using (17) one can easily show that $\tilde{V}_{i-1} \cap Z_{i-1} = (\tilde{V}_{i-1} \cap W_i) \backslash W_{i+1} \backslash \cdots \backslash W_k$, for $i = 1, \ldots, k$. Hence, using (18), $\tilde{V}_0 \in P$, Theorem 6 and mathematical induction we find that $\tilde{V}_i \in P$ for $i = 0, \ldots, k$. This implies that we also have $\tilde{V}_i \cap \tilde{V}_{i-1} \cap W_i \in P$, for all $i = 1, \ldots, k$, which we use to derive:

$$\tilde{V}_{i-1} \cap Z_{i-1} = \tilde{V}_i \backslash Z_i, \tag{20}$$

for all $i = 1, \ldots, k$. This proof goes as follows. Using (18) and (17) we find that

$$\begin{aligned}
\tilde{V}_{i-1} \cap Z_{i-1} &= \tilde{V}_i \cap \tilde{V}_{i-1} \cap Z_{i-1} \\
&= (\tilde{V}_i \cap \tilde{V}_{i-1} \cap W_i) \backslash Z_i \\
&\subseteq \tilde{V}_i \cap \tilde{V}_{i-1} \cap W_i \\
&\subseteq \tilde{V}_i \\
&= \overline{\text{conv}}(\tilde{V}_{i-1} \cap Z_{i-1}),
\end{aligned} \tag{21}$$

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 $\tilde{V}_i \cap \tilde{V}_{i-1} \cap W_i = \overline{\text{conv}}(\tilde{V}_{i-1} \cap Z_{i-1}) = \tilde{V}_i$. Substituting this back into (21) we get (20).

Now we use (20) to show that:

$$T_i = \tilde{V}_i \cap Z_i, \tag{22}$$

for all $i = 0, 1, \ldots, k$. We start with $T_0 = V = \mathbb{R}^N \cap V = \tilde{V}_0 \cap Z_0$. Next, assume (22) holds for some $i \in \{0, \ldots, k-1\}$, then from (20) it follows that $T_i = \tilde{V}_{i+1} \backslash Z_{i+1}$, and hence:

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

hereby completing the proof of (22) by mathematical induction. From (22), (18) and (19) we conclude that $\tilde{V}_i$ and T_i satisfy:

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

for all $i = 0, \ldots, k-1$. 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 $T_k = \tilde{V}_k \cap Z_k = \emptyset$. $\qquad \square$

Lemma 6 *Let $V \in \tilde{P}$, then $\overline{V} \in P$.*

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 $a_i \in \mathbb{R}^N$, $b_i \in \mathbb{R}$ ($i = 1, \ldots, l$), and define:

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

Since $V \subseteq W$, we have $\overline{V} \subseteq \overline{W} = W$. To prove that $W \subseteq \overline{V}$ we take $x \in W \backslash V$ 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 $i \in I_3 \backslash I_2$. Let $\delta > 0$ be such that $a_i \cdot y + b_i > 0$ for all $i \in I_3 \backslash I_2$ 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 $i \in I_3 \backslash I_2$. $\square$

Lemma 7 *Let* $V_1, V_2, \ldots, V_l \in P$ *for some* $l \in \mathbb{N}$ *and* $V_i = \mathrm{conv}(\{x_{ij}\}_{j=1}^{k_i}) + \mathrm{cone}(\{y_{ij}\}_{j=1}^{p_i})$, $i = 1, \ldots, l$. *Then* $\overline{\mathrm{conv}}(\bigcup_{i=1}^l V_i) = \mathrm{conv}(\{x_{ij}\}_{i=1\,j=1}^{l\quad k_i}) + \mathrm{cone}(\{y_{ij}\}_{i=1\,j=1}^{l\quad p_i})$.

Proof If $l = 1$ the proof is trivial, we therefore assume $l \geq 2$.
Define the subsets $W, V' \subseteq \mathbb{R}^N$ by:

$$
\begin{aligned}
W &= \{x \in \mathbb{R}^N \mid x = \textstyle\sum_{i=1}^l \lambda_i v_i,\ v_i \in V_i,\ \lambda_i \geq 0, \sum_{i=1}^l \lambda_i = 1\}, \\
V' &= \mathrm{conv}(\{x_{ij}\}_{i=1\,j=1}^{l\quad k_i}) + \mathrm{cone}(\{y_{ij}\}_{i=1\,j=1}^{l\quad p_i}).
\end{aligned}
$$

Using that V_i is convex one easily shows that $\overline{\mathrm{conv}}(V) = \overline{W}$. Hence, the proof is completed by showing that $\overline{W} = V'$. Since $W \subseteq V'$ implies $\overline{W} \subseteq \overline{V'} \subseteq V'$, it remains to verify that $V' \subseteq \overline{W}$.

Let $x \in V'$, then from 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 $k_1 = k_2 = \cdots = k_l = 0$, 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 $k_1 \geq k_2 \geq \ldots \geq k_r > 0$, $k_{r+1} = \ldots = k_l = 0$ for some $1 \leq r \leq l$ 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:

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

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

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 the lemma. $\square$

19