The Perceptron Algorithm

Rune Sundgot*and Tor Gulliksen
Department of Mathematics, University of Oslo,
P.b. 1053 Blindern, N-0316 Oslo, Norway.

Abstract

The perceptron algorithm is an iterative method for solving a system of linear inequalities. We study the performance of this algorithm when used as a "learning" algorithm in a type of artificial neural networks called perceptrons. We show that the number of iterations needed to reach a solution can grow hyperexponentially in the dimensionality of the problem. We also extend an earlier result concerning the behaviour of this algorithm in the case where no solution exists.

1 Introduction

The perceptron algorithm is an iterative algorithm for finding a solution to a set of linear inequalities. Our interest in the algorithm stems from the important role it has played in the field of artificial neural networks, where it has been used for "training" certain pattern classification machines called *perceptrons* (see for instance [Rosenblatt, 1962] or [Minsky and Papert, 1969]).

The simplest form of a perceptron is a so called linear classifier. Given two sets of "patterns", i.e. two sets of vectors V^+ and V^- in R^{n-1}. By "training" such a perceptron one means running an algorithm whose goal is (if possible) to produce a *weight*-vector w in R^{n-1}, and a *threshold value* Θ in R such that

$$\begin{array}{ll} vw > \Theta & \text{for } v \in V^+ \\ vw < \Theta & \text{for } v \in V^-. \end{array} \tag{1}$$

For the sake of conveniance, we put

$$V = \{V^+ \times \{1\}\} \cup \{-V^- \times \{-1\}\}. \tag{2}$$

Using the relation $x = (w, -\Theta)$ one easily checks that $(w, -\Theta)$ is a solution of (1) if and only if x is a solution of

$$vx > 0 \qquad \text{for } v \in V. \tag{3}$$

In general, any subset V of R^n for which (3) has a solution x will be called *separable*, and x will be called a *separating vector* for V. We will only consider finite sets V. The perceptron algorithm is the following iterative method for computing a separating vector for V:

> Start with some initial vector x in R^n.
> As long as possible choose any v in V for which
> $vx \leq 0$, and replace x by $x + v$.

It is well known that this algorithm terminates with a separating vector for V whenever V is separable, see for instance [Minsky and Papert, 1969] or [Block and Levin, 1970]. In the non-separable case, it never terminates.

Let V be a finite set of vectors in R^n.

*Supported by a grant from the Norwegian Research Counsil.

DEFINITION 1 . *A V-chain is a sequence of vectors $x_{(0)}, x_{(1)}, \ldots$, which for all i satisfies*

$$x_{(i+1)} = x_{(i)} + v_{(i)} \qquad \textit{for some } v_{(i)} \in V \textit{ satisfying } x_{(i)}v_{(i)} \leq 0. \tag{4}$$

Observe that the succesive values of x being produced by the perceptron algorithm is a V-chain. The symbols $x_{(0)}, x_{(1)}, \ldots, x_{(i)}, \ldots$, will be used to denote these values. Thus the vector $x_{(i)}$ is the value reached after i iterations. The initial value $x_{(0)}$ can be any vector in R^n, but is usually set to 0.

When $x_{(0)} = 0$ and V is separable, let $i_{\max} = i_{\max}(V)$ and $i_{\min} = i_{\min}(V)$ denote, respectively, the maximum and the minimum number of iterations before the algorithm terminates. The actual number of iterations will depend on the choice of the vector v in each iteration. When V is given by V^+ and V^- as in (2), we call V and vectors in V *binary* if V^+ and V^- are subsets of $\{0,1\}^{n-1}$, and *bipolar* if V^+ and V^- are subsets of $\{-1,1\}^{n-1}$. Our main theorem's about separable sets are restricted to the binary case and the bipolar case, but several of the lemma's are valid for more general sets.

We first consider the case where V is separable. In section 2 we find a general upper bound of $i_{\max}(V)$. From this formula we get upper bounds in the binary case and in the bipolar case. Both these bounds grow hyperexponentially in n. In section 3 we find a lower bound of $i_{\min}(V)$, valid for a more restricted class of sets. We show that there exist sequences of both binary and bipolar sets from this class, for which also $i_{\min}$ grow hyperexponentially in n. As an easy consequence of these results, we show in section 4 that a conjecture in [Minsky and Papert, 1969] is false.

In [Block and Levin, 1970] a result called the Boundedness Theorem states that for each finite subset V of R^n, there exists some constant $K = K(V)$, such that $||x_{(i)}|| \leq ||x_{(0)}|| + K$. Although elegant, the proof is non-constructive. It does not give any indication of the size of K or how to compute K. In section 5 we give a new proof of this result, providing a recursive formula for computing K in terms of V.

It should be emphasized that other and somewhat more efficient algorithms exist for solving (3) (see for instance [Hand, 1981]). However, the computations involved in the perceptron algorithm are extremely simple and easily implementable in an artificial neural network.

SYMBOL	DESCRIPTION
0	the number 0, and also the 0-vector $(0,0,\ldots,0)$.
uv	euclidean inner product between vectors u and v.
$v^\perp$	the $n-1$ dimensional subspace of vectors orthogonal to v, $v \in R^n \setminus \{0\}$.
$v \perp G$	vector v is orthogonal to the subspace G.
$\text{proj}(v, G)$	orthogonal projection of v onto G.
$\angle(v, G)$	angle between vector v and subspace G, i.e. $\angle(v, \text{proj}(v, G))$.
I_k	$k \times k$ identity matrix.
E_k	the k-dimensional vector $(1, 1, \ldots, 1)$.
A^t	transpose of the matrix A.
$\text{Span}(V)$	space spanned by the vectors in the set V.
$\dim(V)$	dimension of $\text{Span}(V)$.
$\text{conv}(V)$	convex hull of V.
V_z	the subset of vectors in V orthogonal to $z \in R^n \setminus \{0\}$.
$\text{int}(B)$	interior of the set B.
∂B	the boundary of the set B.
[a,b]	the line segment $\text{conv}(\{a, b\})$, $a, b \in R^n$.
l_p^v	the line through p parallel to v, for v in V and p in $v^\perp$. Such lines will be called *V-lines*.

2 Upper bound of $i_{\max}$

In this section we first find a general upper bound of $i_{\max}$ for separable, finite, and nonempty subsets V of R^n. Then we specialize to the binary case and the bipolar case, and obtain upper bounds of $i_{\max}$ expressed as functions of n only. Observe that if V is a separable set, then 0 is not contained in V. To simplify notation we omit the condition $v \in V$ in $\min\{\cdot\}$ and $\max\{\cdot\}$ expressions.

For $p \in R^n$, let

$$\begin{aligned} B(p) &= \{x \in R^n : \|x - \frac{p}{2}\|^2 \leq \|\frac{p}{2}\|^2\} \\ &= \{x \in R^n : xx \leq xp\}, \end{aligned} \tag{5}$$

thus $B(p)$ is the ball with center $\frac{p}{2}$ and radius $\|\frac{p}{2}\|$.

LEMMA 2 . *If $V \subset B(p)$ then all V-chains starting with a vector in $B(p)$ will remain in $B(p)$.*

Proof. Let x and $x+v$ be two subsequent vectors in a V-chain. Assume $x \in B(p)$. We must proove that $x + v \in B(p)$. We know that $xv \leq 0$, and since $x \in B(p)$ and $v \in B(p)$ we have $xx \leq xp$ and $vv \leq vp$. Then

$$\begin{aligned} (x+v)(x+v) &= xx + 2xv + vv \\ &\leq xp + vp \\ &= (x+v)p \end{aligned}$$

so $x + v \in B(p)$. □

LEMMA 3 . *Let V be a separable, finite, and nonempty subset of R^n. If $V \subset B(p)$ then*

$$i_{\max} \leq \frac{pp}{\min\{vv\}}. \tag{6}$$

Proof. Let $x_{(0)}, x_{(1)}, \ldots,$ be a V-chain produced by the perceptron algorithm. Since $x_{(0)} = 0 \in B(p)$ we can apply Lemma 2 and conclude that $x_{(i)} \in B(p)$, in particular $x_{(i)}p \leq pp$. Since $x_{(i)}$ is a sum of i vectors from V, we also have

$$\begin{aligned} x_{(i)}p &\geq i \min\{vp\} \\ &\geq i \min\{vv\}. \end{aligned}$$

The algorithm must terminate before $i \min\{vv\}$ increases beyond pp, and that implies (6). □

Let V be as in Lemma 3, then there exists a vector p such that $V \subset B(p)$. Indeed, if x is a separating vector for V, $p = kx$, and $k \geq \max\{\frac{vv}{vx}\}$, then $V \subset B(p)$.

The best bound we can get from (6) is obtained when p is as small as possible. From the definition of $B(p)$ it follows that the set $S = \{p \in R^n : V \subset B(p)\}$ is both closed and convex (and nonempty since V is separable). Then there exists a unique smallest vector in S. Let p^* denote this vector. We get

$$i_{\max} \leq \frac{p^*p^*}{\min\{vv\}}. \tag{7}$$

Before further elaboration on this bound, we consider another bound of $i_{\max}$. In [Minsky and Papert, 1969] and [Block and Levin, 1970], bounds of $i_{\max}$ are found which can be rewritten or generalized to

$$i_{\max} \leq \frac{\max\{vv\}}{ww}, \tag{8}$$

where w is any separating vector for V satisfying $ww \leq \min\{vw\}$. In this case there exists a largest such vector w, which we denote w^* (this is in fact the unique smallest vector in $\text{conv}(V)$). The best bound obtainable from (8) is

$$i_{\max} \leq \frac{\max\{vv\}}{w^*w^*}. \tag{9}$$

To compare these bounds, let $f, g : R^n \to R$ be defined by $f(x) = xx/\min\{vv\}$ and $g(x) = \max\{vv\}/xx$. Put $C = \max\{vv\}/\min\{vv\}$. From (7) we have $i_{\max} \leq f(p^*)$ and from (9) $i_{\max} \leq g(w^*)$. Using that $w^*w^* = \min\{vw^*\}$, it is easy to check that the vector $p = (\frac{\max\{vv\}}{w^*w^*})w^*$ satisfies $V \subset B(p)$, and that

$$f(p^*) \leq f(p) = C \cdot g(w^*).$$

Similarly, the vector $w = (\frac{\min\{vv\}}{p^*p^*})p^*$ satisfies $ww \leq \min\{vw\}$, and

$$g(w^*) \leq g(w) = C \cdot f(p^*).$$

When V is a bipolar set then $C = 1$ and $f(p^*) = g(w^*)$, so in this case (7) and (9) give exactly the same bound of $i_{\max}$. When V is a binary set, all we know is that $C \leq n$. For some binary sets we have $f(p^*) < g(w^*)$, for other sets $f(p^*) > g(w^*)$. We will continue with finding an upper bound of $f(p^*)$. We have not found an upper bound of $g(w^*)$, other than C times an upper bound of $f(p^*)$.

Now, if $v_1, \ldots, v_n \in V$ are linearly independent, define $X = X(v_1, \ldots, v_n)$ to be the unique vector satisfying $v_i v_i = v_i X$ for $i = 1, \ldots, n$.

LEMMA 4 . *If* $\text{Span}(V) = R^n$ *we can find linearly independent vectors* $v_1, \ldots, v_n$ *in* V *such that*

$$||p^*|| \leq ||X(v_1, \ldots, v_n)||.$$

Proof. Let $V' = V \cap \partial B(p^*)$. Obviously $V' \neq \emptyset$. Let $\mathcal{H} = \text{Span}(\mathcal{V}')$ and $k = \dim(\mathcal{H})$. Then $1 \leq k \leq n$. We will first proove that $p^* \in \mathcal{H}$. Assume (for contradiction) that $p^* \notin \mathcal{H}$. Let $q = \text{proj}(p^*, \mathcal{H})$, and let l denote the line through p^* and q. Then l is orthogonal to $\mathcal{H}$ and $l \cap \text{int}(B(p^*)) \neq \emptyset$. For all $q' \in l$ we have $V' \subset \partial B(q')$. Since V is finite we can find $\epsilon > 0$ such that $V \setminus V' \subset B(q')$ for all $q' \in R^n$ satisfying $||q' - p^*|| < \epsilon$. Any $q' \in l \cap \text{int}(B(p^*))$ satisfying $||q' - p^*|| < \epsilon$ must then satisfy $V \subset B(q')$. But $||q'|| < ||p^*||$ for such a q'. This is the contradiction which proves that $p^* \in \mathcal{H}$. Now, choose n linearly independent vectors $v_1, \ldots, v_n$ in V, where $v_1, \ldots, v_k \in V'$ and $v_{k+1}, \ldots, v_n \in V \setminus V'$. Since p^* and $X(v_1, \ldots, v_n)$ have the same inner product with each of the vectors in V' it follows that $(X(v_1, \ldots, v_n) - p^*) \perp \mathcal{H}$, and since $p^* \in \mathcal{H}$ it follows that

$$\begin{aligned} ||p^*||^2 &\leq ||p^*||^2 + ||X(v_1, \ldots, v_n) - p^*||^2 \\ &= ||X(v_1, \ldots, v_n)||^2. \end{aligned}$$

□

From (7) and Lemma 4 it follows that if $\text{Span}(V) = R^n$, then we can find linearly independent vectors $v_1, \ldots, v_n$ in V such that

$$i_{\max} \leq \frac{||X(v_1, \ldots, v_n)||^2}{\min\{vv\}}. \tag{10}$$

If V_1 is a separable binary set and $\dim(V_1) = m < n$, we can always find another separable binary set V_2 where $\dim(V_2) = n$ and $V_1 \subset V_2$ (simply let V_2 contain an extra $n - m$ linearly independent binary vectors not in Span(V_1)). Any V_1-chain will then also be a V_2-chain, so $i_{\max}(V_1) \leq i_{\max}(V_2)$. It follows that any upper bound of $i_{\max}$ for separable binary sets which span R^n will also be an upper bound of $i_{\max}$ for *all* separable binary sets. The same is also true for bipolar sets.

THEOREM 5 . *If $V \subset R^n$ is a separable binary set then*

$$i_{\max} \leq n^5(n-1)^{n-1}. \tag{11}$$

If $V \subset R^n$ is a separable bipolar set then

$$i_{\max} \leq n^4(\frac{n-1}{4})^{n-1}. \tag{12}$$

Proof. From the preceeding discussion we can assume that Span(V) = R^n. Let $v_1, \ldots, v_n$ be n linearly independent vectors in V. From the definition of $X = X(v_1, \ldots, v_n)$ it follows that

$$X^t = A^{-1} \begin{bmatrix} v_1 v_1 \\ \vdots \\ v_n v_n \end{bmatrix}, \tag{13}$$

where A is the $n \times n$ matrix with $v_1, \ldots, v_n$ as row-vectors. To find an upper bound of $||X(v_1, \ldots, v_n)||$ we use that

$$A^{-1} = \frac{1}{\det(A)} \begin{bmatrix} C_{11} & C_{12} & \ldots & C_{1n} \\ C_{21} & C_{22} & \ldots & C_{2n} \\ \vdots & \vdots & & \vdots \\ C_{n1} & C_{n2} & \ldots & C_{nn} \end{bmatrix}^t, \tag{14}$$

where C_{ij} is $(-1)^{i+j}$ times the determinant of the $(n-1) \times (n-1)$ submatrix that remains after the i'th row and the j'th column are deleted from A. The C_{ij}'s are determinants of matrices whose rows are bounded in length by $(n-1)^{\frac{1}{2}}$, both when V is binary and when V is bipolar. The absolute values of the C_{ij}'s are then bounded by $(n-1)^{\frac{n-1}{2}}$. For $i = 1, \ldots, n$ we have $v_i v_i \leq n$, so from (10), (13) and (14) we then get

$$i_{\max} \leq \frac{n^5(n-1)^{n-1}}{\min\{vv\} \det^2(A)}.$$

When V is a binary set we have $\min\{vv\} \geq 1$ and $\det^2(A) \geq 1$, which gives the upper bound for binary sets. When V is a bipolar set we have $\min\{vv\} = n$. By adding the first row of A to each of the other rows we see that $\det(A)$ is a multiple of 2^{n-1}, so $\det^2(A) \geq 4^{n-1}$ and we get the upper bound for bipolar sets. □

Both the bounds in Theorem 5 grow hyperexponentially in n. In the next section we show that this must be the case for every upper bound of $i_{\max}$ for binary and bipolar sets.

3 Lower bound of $i_{\min}$

Consider, as an example, a set V consisting of exactly n linearly independent vectors $v_1, \ldots, v_n$ in R^n. Observe that such a set must be separable. Let A be the $n \times n$ matrix with $v_1, \ldots, v_n$

as row-vectors, and let Y be the unique vector satisfying $v_j Y = 1$ for $j = 1, \ldots, n$, in matrix notation $AY^t = E_n^t$. Assume that $v_1, \ldots, v_n$ all have integer components, and let

$$Y = a_1 v_1 + \ldots + a_n v_n$$

be the unique representation of Y as a linear combination of $v_1, \ldots, v_n$. Assume that V is *positive* by which we mean that $a_1, \ldots, a_n \geq 0$. Let $x_{(i)}$ be a separating vector for V, reached by the perceptron algorithm (with $x_{(0)} = 0$) after $i = i_{\min}(V)$ iterations. Since $x_{(i)}$ then also have integer components, it follows that $vx_{(i)} \geq 1$ for all v in V. It follows that

$$\begin{aligned} i_{\min}(V) &= x_{(i)} Y \\ &= x_{(i)}(a_1 v_1 + \cdots + a_n v_n) \\ &= (v_1 x_{(i)}) a_1 + \cdots + (v_n x_{(i)}) a_n \\ &\geq (v_1 Y) a_1 + \cdots + (v_n Y) a_n \\ &= Y(a_1 v_1 + \cdots + a_n v_n) \\ &= YY \\ &= \|Y\|^2. \end{aligned}$$

Inspired by this example, we will find binary and bipolar sets consisting of n linearly independent vectors in R^n, where the square length of the corresponding Y-vectors grow hyper-exponentially in n. These sets will not necessarily be positive sets, but the next lemma guarantees the existence of such sets. Whenever we have a set V or a matrix A as in this example, the corresponding Y-vector will be denoted $Y(V)$ or $Y(A)$.

LEMMA 6 . *Let V consist of n linearly independent vectors in R^n. If V is not a positive set, we can make V positive by changing sign of some of its vectors. This can be done so that the length of $Y(V)$ increases.*

Proof. Let V consist of n linearly independent vectors $v_1, \ldots, v_n$ in R^n, and assume $Y(V) = a_1 v_1 + \cdots + a_n v_n$, where for instance $a_n < 0$. Let $V' = \{v_1, \ldots, v_{n-1}, -v_n\}$. We will first show that $\|Y(V')\| > \|Y(V)\|$. Let $G = \text{Span}(v_1, \ldots, v_{n-1})$, $Y_G = \text{proj}(Y(V), G)$ and $Y_G^{\perp} = Y(V) - Y_G$. Now $v_j Y_{G^\perp} = 0$ for $j = 1, \ldots, n-1$. Since $a_n < 0$ it follows that v_n and $Y(V)$ lies on opposite sides of G, in particular $v_n Y_{G^\perp} < 0$. We can then find a constant $k > 0$ such that $(-v_n)(kY_{G^\perp}) = 2$. It follows that

$$\begin{aligned} Y(V') &= Y(V) + kY_{\perp} \\ &= Y_G + (1+k) Y_{G^\perp}, \end{aligned}$$

so $\|Y(V')\| > \|Y(V)\|$ since $Y_G \perp Y_{G^\perp}$. If V' is not a positive set, we can again change sign of one of its vectors such that the length of the corresponding Y-vector increases. Since only a finite number of sets can be reached by such sign changes, this process must eventually end up with a positive set. □

Changing sign of a binary vector produces a new binary vector, and binary vectors have integer components. It follows from the preceeding example and lemma, that if $V \subset R^n$ consists of n linearly independent binary vectors, then there also exists another such set $\bar{V}$, such that $i_{\min}(\bar{V}) \geq \|Y(\bar{V})\|^2 \geq \|Y(V)\|^2$. The same holds also for bipolar vectors.

We will now begin the construction of sets with large Y-vectors. In this construction we will use Hadamard matrices. A Hadamard matrix is a square matrix with orthogonal rows, and with components ± 1. Let $H_1 = [-1]$ and define recursively

$$H_{2k} = \begin{bmatrix} H_k & H_k \\ H_k & -H_k \end{bmatrix}.$$

It is easy to show by induction that this defines a $k \times k$ Hadamard matrix H_k for every $k = 2^l$, $l \geq 0$. Observe that the first row of these matrices consist entirely of -1's, so from the orthogonality all other rows consist of an equal number of 1's and -1's. Observe also that $H_k H_k^t = kI_k$.

LEMMA 7 . *Let $k = 2^l$, $l \geq 0$, and define the $2k \times 2k$ matrix*

$$M = \begin{bmatrix} I & H^t J \\ H & I \end{bmatrix},$$

where $I = I_k$, $H = H_k$, and J is the $k \times k$ matrix with 1's just below the diagonal and 0's elsewhere. Then M is invertible and the vector $Y = Y(M)$ satisfies

$$\|Y\| \geq k^k.$$

Proof. Let $D = I - HH^tJ$. Then

$$D = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ -k & 1 & 0 & \cdots & 0 & 0 \\ 0 & -k & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & 0 & \cdots & -k & 1 \end{bmatrix}$$

and

$$D^{-1} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ k & 1 & 0 & \cdots & 0 & 0 \\ k^2 & k & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ k^{k-2} & k^{k-3} & k^{k-4} & \cdots & 1 & 0 \\ k^{k-1} & k^{k-2} & k^{k-3} & \cdots & k & 1 \end{bmatrix}.$$

It is easy to check that

$$M^{-1} = \begin{bmatrix} I + H^tJD^{-1}H & -H^tJD^{-1} \\ -D^{-1}H & D^{-1} \end{bmatrix}.$$

Let Y_1 and Y_2 denote the vectors consisting of, respectively, the first k and the last k components of the vector $Y = Y(M)$. Since $Y^t = M^{-1}E_{2k}^t$ it follows that, in particular,

$$Y_2^t = D^{-1}(I - H)E_k^t.$$

Since the first row of H consists entirely of -1's, and the other rows have an equal number of -1's and 1's, we get

$$(I - H)E_k^t = \begin{bmatrix} k+1 \\ 1 \\ \vdots \\ 1 \end{bmatrix},$$

so

$$Y_2^t = \begin{bmatrix} k+1 \\ k^2 + k + 1 \\ \vdots \\ k^k + k^{k-1} + \cdots + 1 \end{bmatrix},$$

and $||Y|| \geq ||Y_2|| \geq k^k$. □

The rows of the matrix M in Lemma 7 are linearly independent since M is invertible, but they are neither binary nor bipolar since they have -1, 0 and 1-components in the same row. However, the matrix M will be very useful in obtaining the matrices we want.

THEOREM 8 . *Let $n = 4 \cdot 2^l$, $l \geq 0$. There exists a separable binary set $V' \subset R^n$ satisfying*

$$i_{\min}(V') \geq 4(\frac{n}{4})^{\frac{n}{2}}, \tag{15}$$

and a separable bipolar set $V'' \subset R^n$ satisfying

$$i_{\min}(V'') \geq (\frac{n}{4})^{\frac{n}{2}}. \tag{16}$$

Proof. Let $k = \frac{n}{4}$ and let M be the $2k \times 2k$ matrix of Lemma 7. Define M_0^-, M_0^+, M_-^0 and M_+^0 to be the matrices made from M by replacing all components of the upper index type by components of the lower index type. The indexes $-$, $+$ and 0 refer to -1, $+1$ and 0 respectively, so for instance M_0^- is made from M by replacing all -1's in M by 0's. Observe that $M_0^- + M_0^+ = M$ and $M_-^0 + M_+^0 = 2M$. Let

$$A' = \begin{bmatrix} M_0^- & B \\ M_0^+ & -B \end{bmatrix} \quad \text{and} \quad A'' = \begin{bmatrix} M_-^0 & C \\ M_+^0 & -C \end{bmatrix},$$

where B is an invertible matrix consisting of 0's and 1's and with only 1's in the last column, and C is an invertible matrix consisting of -1's and 1's. For instance, we can choose B equal to I_{2k} except for the last column, and C equal to H_{2k}. Observe that the rows of A' are binary vectors and the rows of A'' are bipolar vectors. Both A' and A'' are easily seen to be invertible since M is invertible. Consider the vector $Y = Y(A')$, which by definition satisfies

$$A'Y^t = E_n^t. \tag{17}$$

Let Y_1 and Y_2 consist of, respectively, the first and the last $2k$ components of Y. Then

$$M_0^- Y_1^t + BY_2^t = E_{2k}^t \tag{18}$$
$$M_0^+ Y_1^t - BY_2^t = E_{2k}^t. \tag{19}$$

Adding (18) and (19) together, then dividing by 2, we get $M(\frac{1}{2}Y_1^t) = E_{2k}^t$. From Lemma 7 it follows that $\frac{1}{2}||Y_1|| \geq k^k$, so

$$||Y|| \geq ||Y_1|| \geq 2k^k = 2(\frac{n}{4})^{\frac{n}{4}}.$$

Let $V = \{v_1, \ldots, v_n\}$ be the set consisting of the n row-vectors of A'. These vectors are linearly independent since A' is invertible, but V is not necessarily a positive set. However, as already noted, it is possible to change sign of some of the vectors in V to obtain a positive set V' which satisfies

$$i_{\min}(V') \geq ||Y(V')||^2 \geq ||Y(V)||^2 \geq 4(\frac{n}{4})^{\frac{n}{2}}.$$

This proves the binary case. The bipolar case follows from the matrix A'' in a similar way. □

4 A conjecture of Minsky and Papert

In [Minsky and Papert, 1969, page 156-158], special subsets U^+, U^- of $\{0,1\}^n$ are studied. These sets have the following property:

If x is a vector in R^n with integer components, satisfying

$$\begin{array}{ll} ux \geq 0 & \text{for } u \in U^+ \\ ux < 0 & \text{for } u \in U^-, \end{array}$$

then $\|x_i\| \geq f_i$, where f_i is the i'th Fibonacci number: $\{f_i\} = \{1,1,2,3,5,8,13,\ldots\}$.

It follows that the maximal component of x must then grow exponentially in n. Maximal here refers to the absolute value. Minsky and Papert conjecture that this is a worst case, i.e. that the maximal component never has to grow faster in n than in their example. We will in this section show that this conjecture is false. We will show that there exist examples where the maximal component must grow hyperexponentially in n.

From the proof of Theorem 8 we know that, for $n = 4\cdot 2^l$, $l \geq 0$, there exists a positive binary subset V of R^n, consisting of n linearly independent vectors $v_1,\ldots,v_n$, satisfying $\|Y(V)\|^2 \geq 4(\frac{n}{4})^{\frac{n}{2}}$. As before, $Y = Y(V)$ denote the unique vector satisfying $v_jY = 1$ for $j = 1,\ldots,n$. Recall that V is called positive if

$$Y = a_1v_1 + \cdots + a_nv_n$$

with non-negative coefficients $a_1,\ldots,a_n$. If the last component of Y is positive, then we change sign of all vectors in V. Then Y will change sign as well, and V will still be a positive set with the same coefficients $a_1,\ldots,a_n$ as before. Recall that $V = \{V^+ \times \{1\}\} \cup \{-V^- \times \{-1\}\}$. Let $V^{++} = V^+ \times \{1\}$ and $V^{--} = -V^- \times \{-1\}$. Let $\tilde{Y} = \tilde{Y}(V)$ denote the unique vector satisfying

$$\begin{array}{ll} v\tilde{Y} = 0 & \text{for } v \in V^{++} \\ v\tilde{Y} = 1 & \text{for } v \in V^{--}. \end{array}$$

It is easy to check that

$$\tilde{Y} = \frac{1}{2}Y - \frac{1}{2}(0,\ldots,0,1).$$

Indeed, if $v \in V^{++}$ then $\frac{1}{2}vY - \frac{1}{2}v(0,\ldots,0,1) = \frac{1}{2} - \frac{1}{2} = 0$, and if $v \in V^{--}$ then $\frac{1}{2}vY - \frac{1}{2}v(0,\ldots,0,1) = \frac{1}{2} + \frac{1}{2} = 1$. Now, let x be a vector in R^n with integer components, satisfying

$$\begin{array}{ll} vx \geq 0 & \text{for } v \in V^+ \times \{1\} \\ vx < 0 & \text{for } v \in V^- \times \{1\}. \end{array} \tag{20}$$

We will show that the maximal component of x must grow hyperexponentially in n, thereby disproving the conjecture of Minsky and Papert. Since x has integer components we have $vx \geq 0$ for $v \in V^{++}$ and $vx \geq 1$ for $v \in V^{--}$. We get

$$\begin{aligned} xY &= x(a_1v_1 + \cdots + a_nv_n) \\ &= (v_1x)a_1 + \cdots + (v_nx)a_n \\ &\geq (v_1\tilde{Y})a_1 + \cdots + (v_n\tilde{Y})a_n \\ &= \tilde{Y}(a_1v_1 + \cdots + a_nv_n) \\ &= \tilde{Y}Y \\ &= (\frac{1}{2}Y - \frac{1}{2}(0,\ldots,0,1))Y \\ &\geq \frac{1}{2}\|Y\|^2 \end{aligned}$$

It follows that $||x|| \geq \frac{1}{2}||Y|| = \frac{1}{2}(2(\frac{n}{4})^{\frac{n}{4}}) = (\frac{n}{4})^{\frac{n}{4}}$, which implies that the maximal component of x must grow hyperexponentially in n.

Remark. The inequalities (20) correspond to replacing the '>' in (1) with '$\geq$'. It is straightforward to modify the perceptron algorithm to solve this type of inequalities instead. The lower bounds of $i_{\max}$ in Theorem 8 would then have to be divided by 4.

5 Boundedness Theorem

In this section we assume that V is a finite subset of R^n and that $\mathrm{Span}(V) = R^n$. The initial vector $x_{(0)}$ in the perceptron algorithm can now be any vector in R^n, and V is no longer assumed to be separable.

THEOREM 9 (**Boundedness Theorem**). *Let V be a finite subset of R^n satisfying $\mathrm{Span}(V) = R^n$, and let $x_{(0)}, x_{(1)}, \ldots$ be a V-chain. For $1 \leq m \leq n-1$ let*

$$\Psi_m = \min\{\angle(v_0, \mathrm{Span}(v_1, \ldots, v_m))\}$$

where the minimum is taken over all choices of linearly independent vectors $v_0, v_1, \ldots, v_m$ in V. Let $K_0 = 0$, $K_1 = \max\{||v|| : v \in V\}$, and let

$$K_{m+1} = \sqrt{K_1^2 + 2\cos(\Psi_m)K_1K_m + K_m^2}/\sin(\Psi_m) \qquad \text{for } 1 \leq m \leq n-1. \tag{21}$$

Then, for all i,

$$||x_{(i)}|| \leq \begin{cases} K_n & \text{if } ||x_{(0)}|| \leq (K_n^2 - K_{n-1}^2)^{\frac{1}{2}} \\ (||x_{(0)}||^2 + K_{n-1}^2)^{\frac{1}{2}} & \text{if } ||x_{(0)}|| > (K_n^2 - K_{n-1}^2)^{\frac{1}{2}}, \end{cases} \tag{22}$$

in particular

$$||x_{(i)}|| \leq ||x_{(0)}|| + K_n. \tag{23}$$

As mentioned earlier, the Boundedness Theorem of [Block and Levin, 1970] only states the *existence* of a constant K_n satisfying (23). In our version of the Boundedness Theorem we provide a recursive expression for K_n in terms of V. This section consists of several lemmas which together proves Theorem 9.

Now let V, $\{K_m\}_{m=0}^n$, $\{\Psi_m\}_{m=1}^{n-1}$, and $x_{(0)}, x_{(1)}, \ldots$ be as in Theorem 9, with the extra assumption that $0 \notin V$. The case $0 \in V$ follows trivially from this case, since adding the 0-vector to V does not change the set of vectors which can occur in V-chains.

LEMMA 10 . *Let $C_0, C_1, \ldots, C_n$ be constants which satisfy $C_0 = 0$, $C_1 \geq \max\{||v|| : v \in V\}$, and*

$$C_{m+1} \geq \sqrt{C_1^2 + 2\cos(\Psi_m)C_1C_m + C_m^2}/\sin(\Psi_m) \qquad \text{for } 1 \leq m \leq n-1. \tag{24}$$

Let

$$\mathcal{B} = \bigcap_{z \in S^{n-1}} \{\, y \in R^n : yz \leq (C_n^2 - C_m^2)^{\frac{1}{2}} \text{ where } m = \dim(V_z)\,\}.$$

Then all V-chains starting in $\mathcal{B}$ will remain in $\mathcal{B}$.

Proof of Theorem 9 assuming Lemma 10. Observe that the constants in Lemma 10 satisfy $C_0 < C_1 < \cdots < C_n$. For $z \in S^{n-1}$ and $m = \dim(V_z)$ we have $0 \leq m \leq n-1$, so $(C_n^2 - C_{n-1}^2)^{\frac{1}{2}} \leq (C_n^2 - C_m^2)^{\frac{1}{2}} \leq C_n$ and

$$\{\, y \in R^n : ||y|| \leq (C_n^2 - C_{n-1}^2)^{\frac{1}{2}} \,\} \subset \mathcal{B} \subset \{\, y \in R^n : ||y|| \leq C_n \,\}. \tag{25}$$

To prove the first part of (22), assume that $||x_{(0)}|| \leq (K_n^2 - K_{n-1}^2)^{\frac{1}{2}}$ and choose $C_j = K_j$ for all $0 \leq j \leq n$. With these choices of $\{C_j\}_{j=0}^n$, the conditions in Lemma 10 are satisfied. From (25) it follows that $x_{(0)} \in \mathcal{B}$ and from Lemma 10 we get $x_{(i)} \in \mathcal{B}$, so $||x_{(i)}|| \leq C_n = K_n$. To prove the second part of (22), assume that $||x_{(0)}|| > (K_n^2 - K_{n-1}^2)^{\frac{1}{2}}$ and choose $C_j = K_j$ for $j = 0, 1, \ldots, n-1$ and $C_n = (||x_{(0)}|| + K_{n-1}^2)^{\frac{1}{2}}$. The conditions in Lemma 10 are again satisfied, and $||x_{(0)}|| = (C_n^2 - C_{n-1}^2)^{\frac{1}{2}}$ so $x_{(0)} \in \mathcal{B}$. It follows that $||x_{(i)}|| \leq C_n = (||x_{(0)}||^2 + K_{n-1}^2)^{\frac{1}{2}}$, thereby proving the second part of (22). Since $K_{n-1} < K_n$, the inequality (23) follows directly from (22). □

We will prove Lemma 10 by showing that $\mathcal{B}$ satisfies the conditions in the following lemma. Recall that for v in V and p in $v^\perp$, the V-line l_p^v is the line through p parallel to v, i.e. $l_p^v = \{p + rv : r \in R\}$.

LEMMA 11 . *If B is a convex subset of R^n and*

$$(p + v \notin B) \Rightarrow (l_p^v \cap B = \emptyset) \tag{26}$$

for every V-line l_p^v, then all V-chains starting in B will remain in B.

Proof. Let $B \subset R^n$ be as described and let x and $x + v$ be two subsequent vectors in a V-chain where $x \in B$. Then $v \in V$, $xv \leq 0$ and we must prove that $x + v \in B$. Let p be the unique vector in $v^\perp$ which satisfies $x \in l_p^v$. Then x, p, $x + v$ and $p + v$ all lie on the V-line l_p^v. Since $x \in l_p^v \cap B$ we know from (26) that $p + v \in B$. Since B is convex, the line segment $[x, p + v]$ is contained in B. Since $xv \leq (x + v)v \leq (p + v)v$ it follows that $x + v \in [x, p + v] \subset B$. □

Since $\mathcal{B}$ is an intersection of convex sets it follows that $\mathcal{B}$ itself is a convex set. It remains to prove that (26) is satisfied for $B = \mathcal{B}$. If $y \in R^n \setminus \mathcal{B}$ then there exists a vector z in S^{n-1} satisfying $yz > (C_n^2 - C_m^2)^{\frac{1}{2}}$, where $m = \dim(V_z)$. In this case we say that y is *removed* by z. The set $\mathcal{B}$ thus consists of all vectors y in R^n which are not removed by any vector in S^{n-1}. Observe that if $v \perp z$, then either z does not remove any vectors on the V-line l_p^v, or z removes all vectors on l_p^v. We will prove that if $p + v \notin \mathcal{B}$, then $p + v$ is removed by a vector z in S^{n-1} satisfying $v \perp z$, i.e. by a vector in $S^{n-1} \cap v^\perp$. We will need the following inequalities and couple of lemmas to prove this.

For $1 \leq m \leq n-1$ we easily obtain from (24) the following inequalities

$$(C_{m+1}^2 - C_1^2)^{\frac{1}{2}} \geq \frac{\cos(\Psi_m)C_1 + C_m}{\sin(\Psi_m)} \tag{27}$$

$$(C_{m+1}^2 - C_m^2)^{\frac{1}{2}} \geq \frac{C_1 + \cos(\Psi_m)C_m}{\sin(\Psi_m)}. \tag{28}$$

We also have, for $1 \leq m \leq n-1$, that

$$(C_{m+1}^2 - C_m^2)^{\frac{1}{2}} \geq \cos(\Psi_m)(C_{m+1}^2 - C_1^2)^{\frac{1}{2}} + \sin(\Psi_m)C_1. \tag{29}$$

To prove (29) let $\tilde{C}_m$ denote the righthandside of (24) and define $g_m : \langle C_m, \infty\rangle \rightarrow R$ by

$$g_m(t) = (t^2 - C_m^2)^{\frac{1}{2}} - \cos(\Psi_m)(t^2 - C_1^2)^{\frac{1}{2}} - \sin(\Psi_m)C_1.$$

It is easy to show that when $C_{m+1} = \tilde{C}_m$ we get equality in (27) and (28). These equalities in turn imply equality in (29), so $g_m(\tilde{C}_m) = 0$. Now,

$$g'_m(t) = t/(t^2 - C_m^2)^{\frac{1}{2}} - \cos(\Psi_m)t/(t^2 - C_1^2)^{\frac{1}{2}},$$

so $g'_m(t) > 0$ when $t > C_m$, i.e. $g_m(t)$ is a strictly increasing function. Since $C_{m+1} \geq \tilde{C}_m \overset{(24)}{>} C_m$ we get $g_m(C_{m+1}) \geq g_m(\tilde{C}_m) \geq 0$, and (29) follows. From (24) we have $C_1 < \sin(\Psi_{n-1})C_n$ and from (28) we have $C_1 \leq (C_{m+1}^2 - C_m^2)^{\frac{1}{2}}$ for $1 \leq m \leq n-1$, in particular $C_1 \leq (C_n^2 - C_{n-1}^2)^{\frac{1}{2}}$. These last inequalities will be used without further notice. We will also need the following two lemmas.

LEMMA 12 . *If $v \in V$, $p \in v^\perp$, and $p+v \notin \mathcal{B}$, then $p+v$ is removed by a vector z in S^{n-1} satisfying $p+v \in \mathrm{Span}(V_z, z)$.*

Proof. Since $p+v \notin \mathcal{B}$ there exists a vector z' in S^{n-1} satisfying $(p+v)z' > (C_n^2 - C_{m'}^2)^{\frac{1}{2}}$, where $m' = \dim(V_{z'})$. If $m' = n-1$ then $\mathrm{Span}(V_{z'}, z') = R^n$ and the proof is finished. Assume that $m' \leq n-2$. Let $z'' = \mathrm{proj}(z', \mathrm{Span}(V_{z'}, p+v))$. Since $(p+v)z' > 0$ it follows that $z'' \neq 0$. Let $z = z''/||z''||$ and $m = \dim(V_z)$. Since $z \perp \mathrm{Span}(V_{z'})$ it follows that $m \geq m'$. Since $||z'|| = 1$ then $||z''|| \leq 1$ and $(p+v)z \geq (p+v)z'' = (p+v)z' > (C_n^2 - C_{m'}^2)^{\frac{1}{2}} \geq (C_n^2 - C_m^2)^{\frac{1}{2}}$ so $p+v$ is removed by z. Since $z \in \mathrm{Span}(V_{z'}, p+v)$, but neither $p+v$ nor z is in $\mathrm{Span}(V_{z'})$, it follows that $p+v \in \mathrm{Span}(V_{z'}, z) \subset \mathrm{Span}(V_z, z)$. □

LEMMA 13 . *If $v \in V$, $p \in v^\perp$, $||p+v|| \leq C_n$, $z \in S^{n-1}$, $v \not\perp z$, $p+v \in \mathrm{Span}(V_z, z)$, and $p+v \in \mathrm{Span}(V_z, v)$, then $p+v$ is* not *removed by z.*

Proof. Assume that all seven conditions in Lemma 13 are satisfied, and let $m = \dim(V_z)$. If $m = 0$ then

$$(p+v)z \leq ||p+v|| \leq C_n = (C_n^2 - C_m^2)^{\frac{1}{2}},$$

so $p+v$ is not removed by z. Assume that $m \geq 1$, which also implies that $n \geq 2$. If $||p|| = 0$ then

$$(p+v)z = vz \leq ||v|| \leq C_1 \leq (C_{m+1}^2 - C_m^2)^{\frac{1}{2}} \leq (C_n^2 - C_m^2)^{\frac{1}{2}},$$

so $p+v$ is not removed by z. Assume that $||p|| > 0$. If $p+v \in \mathrm{Span}(V_z)$ then $(p+v)z = 0$ and $p+v$ is not removed by z. Assume that $p+v \notin \mathrm{Span}(V_z)$. Let $\alpha = \angle(v, \mathrm{Span}(V_z))$. Observe that $\alpha \geq \Psi_m \geq \Psi_{n-1}$. It follows that

$$\begin{aligned}
(p+v)z &= \cos(\angle(p,z))||p||||z|| + \cos(\angle(v,z))||v||||z|| \\
&\overset{(a)}{\leq} \cos(\alpha)(C_n^2 - ||v||^2)^{\frac{1}{2}} + \sin(\alpha)||v|| \\
&\overset{(b)}{\leq} \cos(\alpha)(C_n^2 - C_1^2)^{\frac{1}{2}} + \sin(\alpha)C_1 \\
&\overset{(c)}{\leq} \cos(\Psi_m)(C_n^2 - C_1^2)^{\frac{1}{2}} + \sin(\Psi_m)C_1 \\
&\overset{(d)}{\leq} (C_n^2 - C_m^2)^{\frac{1}{2}}.
\end{aligned}$$

(a): Since $p+v \in \mathrm{Span}(V_z, z)$, $p+v \in \mathrm{Span}(V_z, v)$ and $p+v \notin \mathrm{Span}(V_z)$ it follows that $\mathrm{Span}(V_z, z) = \mathrm{Span}(V_z, v)$, in particular $v \in \mathrm{Span}(V_z, z)$. Then $\angle(v, \mathrm{Span}(V_z)) = \angle(v, z^\perp)$ so $\alpha = \angle(v, z^\perp) = \angle(v^\perp, z)$. Since $p \in v^\perp$ and $p \neq 0$ we get $\alpha = \angle(v^\perp, z) \leq \angle(p, z)$ and $\cos(\alpha) \geq \cos(\angle(p,z))$. Since $||p+v|| \leq C_n$ we have $||p|| = (||p+v||^2 - ||v||^2)^{\frac{1}{2}} \leq (C_n^2 - ||v||^2)^{\frac{1}{2}}$. We also have $|\cos(\angle(v,z))| = \sin(\angle(v, z^\perp)) = \sin(\alpha)$, and $||z|| = 1$.

(b): The derivative of the preceeding expression with respect to $||v||$ is positive for $0 \leq ||v|| \leq \sin(\alpha)C_n$, and $||v|| \leq C_1 < \sin(\Psi_{n-1})C_n \leq \sin(\alpha)C_n$ (remember $n \geq 2$ and $\alpha \geq \Psi_m \geq \Psi_{n-1}$).

(c): Define $h : \langle 0, \pi/2] \to R$ by $h(\gamma) = \cos(\gamma)(C_n^2 - C_1^2)^{\frac{1}{2}} + \sin(\gamma)C_1$. We must show that $h(\alpha) \leq h(\Psi_m)$. We have $h'(\gamma) = -\sin(\gamma)(C_n^2 - C_1^2)^{\frac{1}{2}} + \cos(\gamma)C_1$. Since

$$h'(\gamma)(-\sin(\gamma)(C_n^2 - C_1^2)^{\frac{1}{2}} - \cos(\gamma)C_1) = \sin^2(\gamma)C_n^2 - C_1^2,$$

and $C_1 < \sin(\Psi_{n-1})C_n$ it follows that $h'(\gamma) \leq 0$ when $\gamma \geq \Psi_{n-1}$. We have $\alpha \geq \Psi_m \geq \Psi_{n-1}$, so when γ decreases from α to Ψ_m it follows that $h(\gamma)$ increases from $h(\alpha)$ to $h(\Psi_m)$, i.e. $h(\alpha) \leq h(\Psi_m)$.

(d): Consider the function g_m defined after (29). Since g_m is an increasing function, $g_m(C_{m+1}) \geq 0$, and $C_n \geq C_{m+1}$, it follows that $g_m(C_n) \geq 0$. This proves (d). $\square$

We are now ready to prove that (26) is satisfied for $B = \mathcal{B}$. Let $v \in V$, $p \in v^{\perp}$, and assume that $p + v \notin \mathcal{B}$. Then $||p+v|| \geq (C_n^2 - C_{n-1}^2)^{\frac{1}{2}}$, and since $||v|| \leq C_1 \leq (C_n^2 - C_{n-1}^2)^{\frac{1}{2}}$ it follows that $||p|| > 0$. We will prove that $p+v$ is removed by a vector from $S^{n-1} \cap v^{\perp}$. Consider first the case where $||p+v|| > C_n$. Let $z = p/||p||$ and $m = \dim(V_z)$. Then $v \perp z$ and $m \geq 1$. We have

$$(p+v)z = (p+v)\frac{p}{||p||} = ||p|| = (||p+v||^2 - ||v||^2)^{\frac{1}{2}} > (C_n^2 - C_1^2)^{\frac{1}{2}} \geq (C_n^2 - C_m^2)^{\frac{1}{2}}$$

so in this case is $p+v$ removed by a vector z in $S^{n-1} \cap v^{\perp}$, hence $l_p^v \cap \mathcal{B} = \emptyset$. Consider now the case where $||p+v|| \leq C_n$. From Lemma 12 it follows that $p+v$ is removed by a vector z in S^{n-1} satisfying $p+v \in \text{Span}(V_z, z)$. If $v \perp z$ the proof is finished. Assume that $v \not\perp z$. Six of the seven conditions in Lemma 13 are now satisfied, but the conclusion is false. It follows that the remaining condition must be false, i.e. $p+v \notin \text{Span}(V_z, v)$. Let again $m = \dim(V_z)$. It follows that $n \geq m+2$, so $n \geq 2$ and $m \leq n-2$. Let

$$\tilde{z} = (p+v) - \text{proj}(p+v, \text{Span}(V_z, v)).$$

Since $p+v \notin \text{Span}(V_z, v)$ it follows that $\tilde{z} \neq 0$, so $\tilde{z}/||\tilde{z}|| \in S^{n-1} \cap v^{\perp}$. We will show that $p+v$ is removed by $\tilde{z}/||\tilde{z}||$. Since $\tilde{z} \perp \text{Span}(V_z, v)$ it follows that $\dim(V_{\tilde{z}}) \geq \dim(V_z) + 1 = m+1$. Since $p+v \in \text{Span}(V_z, z)$, we have $p+v = rz + y$ for some $r \in R$ and $y \in \text{Span}(V_z)$. We have

$$\begin{aligned}
(p+v)\frac{\tilde{z}}{||\tilde{z}||} &\overset{(a)}{=} [\,||rz||^2 - ||\text{proj}(rz, \text{Span}(V_z, v))||^2\,]^{\frac{1}{2}} \\
&\overset{(b)}{>} [\,(C_n^2 - C_m^2) - ||\text{proj}(rz, \text{Span}(V_z, v))||^2\,]^{\frac{1}{2}} \\
&\overset{(c)}{\geq} [\,(C_n^2 - C_m^2) - (\frac{|(rz)v|}{||v|| \sin(\Psi_m)})^2\,]^{\frac{1}{2}} \\
&\overset{(d)}{>} [\,(C_n^2 - C_m^2) - (\frac{C_1 + \cos(\Psi_m)C_m}{\sin(\Psi_m)})^2\,]^{\frac{1}{2}} \\
&\overset{(e)}{\geq} (C_n^2 - C_{m+1}^2)^{\frac{1}{2}}.
\end{aligned}$$

(a): $(p+v)\frac{\tilde{z}}{||\tilde{z}||} = \frac{\tilde{z}\tilde{z}}{||\tilde{z}||} = ||\tilde{z}|| = [\,||(p+v) - \text{proj}(p+v, \text{Span}(V_z, v))||\,]^{\frac{1}{2}}$ $= [\,||rz - \text{proj}(rz, \text{Span}(V_z, v))||^2\,]^{\frac{1}{2}} = [\,||rz||^2 - ||\text{proj}(rz, \text{Span}(V_z, v))||^2\,]^{\frac{1}{2}}$.

(b): $p+v$ is removed by z so $(p+v)z > (C_n^2 - C_m^2)^{\frac{1}{2}}$, $(p+v)z = (rz+y)z = r$, and $||rz||^2 = r^2$.

(c): Let $e = v - \text{proj}(v, \text{Span}(V_z))$. It follows that $e \in \text{Span}(V_z, v)$ and $e \perp \text{Span}(V_z)$. Since also $z \perp \text{Span}(V_z)$ then

$$||\text{proj}(rz, \text{Span}(V_z, v))|| = ||\text{proj}(rz, \text{Span}(e))|| = ||\frac{(rz)e}{||e||}(\frac{e}{||e||})|| = \frac{|(rz)e|}{||e||}.$$

Now, $||e|| = ||v - \text{proj}(v, \text{Span}(V_z))|| = ||v|| \sin(\angle(v, \text{Span}(V_z))) \geq ||v|| \sin(\Psi_m)$, so $||\text{proj}(rz, \text{Span}(V_z, v))|| \leq \frac{|(rz)v|}{||v|| \sin(\Psi_m)}$.

(d): We have

$$\frac{|(rz)v|}{||v||} = |(p+v)\frac{v}{||v||} - y\frac{v}{||v||}| \leq |(p+v)\frac{v}{||v||}| + |y\frac{v}{||v||}|,$$

where $(p+v)\frac{v}{||v||} = ||v|| \leq C_1$. If $y \neq 0$ then

$$||y|| = (||p+v||^2 - ||rz||^2)^{\frac{1}{2}} < (C_n^2 - (C_n^2 - C_m^2))^{\frac{1}{2}} = C_m,$$

and then

$$|y\frac{v}{||v||}| = ||y|||\cos(\angle(v,y))| < C_m \cos(\Psi_m),$$

which holds also if $y = 0$. We get $\frac{|(rz)v|}{||v||} < C_1 + \cos(\Psi_m)C_m$.

(e): Follows from (28). This is in fact the inequality which tells us how to choose C_{m+1} as a function of C_1, C_m, and Ψ_m.

This ends the proof that (26) is satisfied for $B = \mathcal{B}$. This means that Lemma 10 now is proved, and thereby Theorem 9.

COROLLARY 14 . *We have, for all i,*

$$||x_{(i)}|| \leq ||x_{(0)}|| + \max\{||v|| : v \in V\} \cdot \xi$$

where

$$\xi = \begin{cases} 1 & \text{if } n = 1 \\ \sqrt{n} & \text{if } n \geq 2 \text{ and } \Psi_1 = \pi/2 \\ (\frac{2}{1-\sin(\Psi_1)})/\prod_{j=1}^{n-1}\sin(\Psi_j) & \text{if } n \geq 2 \text{ and } \Psi_1 < \pi/2. \end{cases} \tag{30}$$

Proof. Since $K_1 = \max\{||v|| : v \in V\}$ it follows from (23) that it is sufficient to show that $K_n \leq K_1\,\xi$. If $n = 1$ then $K_n = K_1$ so this case is trivial. Let $n \geq 2$. If $\Psi_1 = \pi/2$ then necessarily $\Psi_1 = \Psi_2 = \cdots = \Psi_{n-1} = \pi/2$, so by repeated use of (21) we get $K_n = K_1\sqrt{n}$. Assume that $\Psi_1 < \pi/2$. From (21) it follows that

$$K_{m+1} \leq (K_1 + K_m)/\sin(\Psi_m) \qquad \text{for } 1 \leq m \leq n-1$$

By repeated use of this inequality we get

$$\begin{aligned} K_n &\leq K_1(2 + \sin(\Psi_1) + \sin(\Psi_1)\sin(\Psi_2) + \cdots + \sin(\Psi_1)\sin(\Psi_2)\cdots\sin(\Psi_{n-2}))/\prod_{j=1}^{n-1}\sin(\Psi_j) \\ &< K_1(1 + 1 + \sin(\Psi_1) + \sin^2(\Psi_1) + \sin^3(\Psi_1) + \cdots)/\prod_{j=1}^{n-1}\sin(\Psi_j) \\ &= K_1(1 + \frac{1}{1-\sin(\Psi_1)})/\prod_{j=1}^{n-1}\sin(\Psi_j) \\ &< K_1(\frac{2}{1-\sin(\Psi_1)})/\prod_{j=1}^{n-1}\sin(\Psi_j) \end{aligned}$$

□

It follows from a result in [Minsky and Papert, 1969] called the Cycling Theorem, that for any $\varepsilon > 0$ there exists a number $N = N(\varepsilon, V)$ such that $||x_{(i)}|| \leq ||x_{(0)}|| + \varepsilon$ for all i, when $||x_{(0)}|| \geq N$. From Theorem 9 we get a more detailed picture.

COROLLARY 15 . *For any $\varepsilon > 0$ we have $||x_{(i)}|| \leq ||x_{(0)}|| + \varepsilon$, for all i, when $||x_{(0)}|| \geq N(\varepsilon, V)$, where*

$$N(\varepsilon, V) = \begin{cases} (K_n^2 - \varepsilon^2)/2\varepsilon & \text{for } \varepsilon \in \langle 0, K_n - (K_n^2 - K_{n-1}^2)^{\frac{1}{2}} \rangle \\ K_n - \varepsilon & \text{for } \varepsilon \in [K_n - (K_n^2 - K_{n-1}^2)^{\frac{1}{2}}, K_n \rangle \\ 0 & \text{for } \varepsilon \in [K_n, \infty \rangle. \end{cases} \tag{31}$$

Proof. It follows from (22) that if $||x_{(0)}|| \leq (K_n^2 - K_{n-1}^2)^{\frac{1}{2}}$ then

$$||x_{(i)}|| \leq ||x_{(0)}|| + (K_n - ||x_{(0)}||),$$

and if $||x_{(0)}|| > (K_n^2 - K_{n-1}^2)^{\frac{1}{2}}$ then

$$||x_{(i)}|| \leq ||x_{(0)}|| + ((||x_{(0)}||^2 + K_{n-1}^2)^{\frac{1}{2}} - ||x_{(0)}||).$$

Let

$$f(||x_{(0)}||) = \begin{cases} K_n - ||x_{(0)}|| & \text{when } ||x_{(0)}|| \leq (K_n^2 - K_{n-1}^2)^{\frac{1}{2}} \\ (||x_{(0)}||^2 + K_{n-1}^2)^{\frac{1}{2}} - ||x_{(0)}|| & \text{when } ||x_{(0)}|| > (K_n^2 - K_{n-1}^2)^{\frac{1}{2}}. \end{cases}$$

Then $||x_{(i)}|| \leq ||x_{(0)}|| + f(||x_{(0)}||)$ for any $x_{(0)}$. The function $f(\cdot)$ is strictly decreasing and $f(0) = K_n$. For $\varepsilon \geq K_n$ we have $||x_{(i)}|| \leq ||x_{(0)}|| + \varepsilon$ for arbitrary $x_{(0)}$. For $0 < \varepsilon < K_n$ we have $||x_{(i)}|| \leq ||x_{(0)}|| + \varepsilon$ when $||x_{(0)}|| \geq f^{-1}(\varepsilon)$, and that gives us (31). □

In Theorem 9 we assume that Span$(V) = R^n$. In the case where $m = \dim(\text{Span}(V)) < n$, it is straightforward to show that $||x_{(i)}|| \leq ||x_{(0)}|| + K_m$, by using that Span$(V)$ is isomorpfic to R^m.

References

[Rosenblatt, 1962] Rosenblatt,F. (1962), *Principles of Neurodynamics*, Spartan Books, New York.

[Minsky and Papert, 1969] Minsky,M. and Papert,S, (1969), *Perceptrons, an Introduction to Computational Geometry*, MIT Press, Cambridge MA.

[Block and Levin, 1970] Block,H.D. and Levin,S.A. (1970), *On the boundedness of an iterative procedure for solving a system of linear inequalities*, Proc. Amer. Math. Soc., vol 26, no 2, p.229-235.

[Hand, 1981] Hand,D.J. (1981), *Discrimination and Classification*, Wiley, New York.