ON THE BOUNDEDNESS OF AN ITERATIVE PROCE-
DURE FOR SOLVING A SYSTEM OF
LINEAR INEQUALITIES[1]

H. D. BLOCK AND S. A. LEVIN

ABSTRACT. In this paper it is proved that the perceptron error-correction procedure stays bounded, even when no solution to the system of linear inequalities exists. This supplements earlier papers by B. Efron and by M. Minsky and S. Papert.

1. **Introduction.** Let $F = \{\phi\}$ be a finite set of vectors in Euclidean n-space E_n, and consider the problem of finding a vector w in E_n such that

$$(1) \qquad w \cdot \phi > 0 \qquad \text{for all } \phi \text{ in } F.$$

Such a vector w is called a *solution* of the system of inequalities (1).

The *error-correction* procedure we consider is an iterative method for finding a solution if one exists:

Choose A_1 arbitrarily in E_n. If $A_1 \cdot \phi > 0$ for all ϕ in F, then A_1 is a solution, and the process terminates. Otherwise, select as ϕ_1 any ϕ in F such that $A_1 \cdot \phi_1 \leq 0$, and let $A_2 = A_1 + \phi_1$. In general,

$$(2) \qquad \phi_i \in F \quad \text{such that} \quad A_i \cdot \phi_i \leq 0, \quad \text{and} \quad A_{i+1} = A_i + \phi_i,$$
$$i = 1, 2, \cdots,$$

as long as the process continues. It terminates when, and only when, a solution is reached.

2. **Soluble case.** If the problem defined in (1) has a solution, then it is easy to show that the process (2) will terminate at a solution after a finite number of steps. This result is well known; but the following proof, basically due to W. C. Ridgeway, is shorter than any previously published, and is therefore presented for completeness:

Suppose w is a solution of (1). Set

$$a = \min_{\phi \in F} (w \cdot \phi), \qquad b = \max_{\phi \in F} \|\phi\|, \qquad \bar{w} = b^2 w / a.$$

Then from (2),

$$\|A_{i+1} - \bar{w}\|^2 = \|A_i - \bar{w}\|^2 + 2(A_i - \bar{w})\cdot\phi_i + \|\phi_i\|^2$$
$$\leqq \|A_i - \bar{w}\|^2 - 2\bar{w}\cdot\phi_i + b^2 \leqq \|A_i - \bar{w}\|^2 - b^2.$$

Iterating, we obtain

$$0 \leqq \|A_{i+1} - \bar{w}\|^2 \leqq \|A_1 - \bar{w}\|^2 - ib^2,$$

and so

$$i \leqq \|A_1 - \bar{w}\|^2/b^2,$$

which completes the proof.

3. **Insoluble case.** If no solution exists, then the process (2) does not terminate. Excluding the trivial case in which F contains the zero vector, the process cannot converge, since $\|A_{i+1}-A_i\| = \|\phi_i\|$ $\geqq \min_{\phi\in F}\|\phi\| > 0$. The interesting fact is that the process always stays bounded. This was conjectured by Nils Nilsson and independently by Terry Beyer, and a proof was offered by B. Efron in a technical report [1]. This report was quite difficult to read; and, since no version ever was published in a standard journal, the theorem in a sense lacked a convincing proof. Finally, in their book [2], Minsky and Papert published a reasonably clear analysis. There are some informal parts to their treatment that need to be, and can be, made rigorous; but there is one crucial gap which, it appears, cannot be easily bridged. (This is the assertion ([2, p. 187, 1.17]) "$\|C\| < M_{n-1}$," which does not follow, as implied, from the relation "$\|B_1+C\|$ $< \|B_1\| + M_{n-1}$.") The purpose of the present paper is to give a precise proof of the Boundedness Theorem (§5 below). In so doing, somewhat stronger form of boundedness is proved (Lemma 1). This is new and of interest in itself.

The novel ideas and cogent insights of Minsky and Papert [2] provide the basis for the proof below. Many parts of our proof are adapted directly from theirs.

4. **The fundamental lemma.** In this section we prove the basic result (Lemma 1) of this paper. It can be applied directly to the proof of Minsky and Papert to bridge the existing gap, or used in a much more straightforward way to yield the Boundedness Theorem (Theorem 1) as a corollary (§5).

DEFINITION. A sequence $A_1, A_2, \cdots, A_k$ of vectors which satisfy (2) for $i = 1, 2, \cdots, k-1$ is called an *F-chain*. It is called a *proper F-chain* if $\|A_i\| \geqq \|A_1\|$ for $i = 1, 2, \cdots, k$.

LEMMA 1. *For any finite set F of vectors in Euclidean n-space E_n, there exists a positive number $M_n(F)$ such that if $B_1, B_2, \cdots, B_j$ is any proper F-chain, then*

$$(3) \qquad \|B_j - B_1\| \leq M_n(F).$$

We defer the proof of this lemma until we establish certain conventions. We introduce the notation $\hat{z} = z/\|z\|$ for nonzero vectors z, and in general indicate that a vector is a unit vector by the circumflex. By $z^\perp$, we denote the set of vectors in E_n orthogonal to z. These form an $(n-1)$-dimensional hyperplane, which is of course isomorphic to E_{n-1}. The set $F \cap z^\perp$ of vectors in E_n lies in the subspace $z^\perp$, and by $F^*(z)$ we denote this (possibly vacuous) set regarded as a set of vectors in E_{n-1}. Finally, if $B \in E_n$ and $z \in E_n$, we denote by $B + z^\perp$ the set of vectors of the form $B + v$, where $v \in z^\perp$. If each of the vectors in an F-chain lies in $B + z^\perp$, we say that the F-chain *lies in* $B + z^\perp$.

LEMMA 2. *For any unit vector $\hat{z}$ in E_n, and any finite set F of vectors in E_n, $n \geq 2$, there exists a neighborhood $V(\hat{z})$ on the unit sphere in E_n and a number $K(\hat{z})$ such that if $B \in E_n$ with the properties that $\hat{B} \in V(\hat{z})$ and $\|B\| > K(\hat{z})$, then any proper F-chain beginning with B lies in $B + \hat{z}^\perp$.*

Again, we defer the proof of this lemma and prove first one preliminary result. We use the notation $\bar{y} = y - (y \cdot \hat{z})\hat{z}$ to denote the projection of an arbitrary E_n-vector y onto $\hat{z}^\perp$.

LEMMA 3. *Let $\hat{z} \in E_n$ and let $B, \cdots, C$ be a proper F-chain in E_n and lying in $B + \hat{z}^\perp$. Then the chain $\bar{B}, \cdots, \bar{C}$ is a proper $(F \cap \hat{z}^\perp)$-chain in E_n. Moreover, since the elements of this chain can be regarded as elements of E_{n-1}, the chain $\bar{B}, \cdots, \bar{C}$, regarded as a chain in E_{n-1}, is a proper $F^*(\hat{z})$-chain.*

PROOF. First, to show that $\bar{B}, \cdots, \bar{C}$ is an $(F \cap \hat{z}^\perp)$-chain, we consider any two consecutive elements u, $v = u + \phi$ of the chain $B, \cdots, C$. Then clearly $\phi \in F \cap \hat{z}^\perp$, and $\bar{v} - \bar{u} = \bar{\phi} = \phi$. Moreover $\bar{\phi} \cdot \bar{u} = \bar{\phi} \cdot u$, and hence $\bar{\phi} \cdot \bar{u} = \phi \cdot u \leq 0$. Hence the projections form an $(F \cap \hat{z}^\perp)$-chain.

To show that the projected chain is proper, we need only utilize the relation

$$\|\bar{u}\|^2 = \|u\|^2 - (u \cdot \hat{z})^2 = \|u\|^2 - (B \cdot \hat{z})^2.$$

Applying this result to B itself, and subtracting from the above, we obtain

$$\|\bar{u}\|^2 - \|\bar{B}\|^2 = \|u\|^2 - \|B\|^2 \geq 0$$

for any u in the chain $B, \cdots, C$. This completes the proof of Lemma 3.

We now prove, by an "alternating" procedure of induction on the dimension n of the Euclidean n-space in which the vectors of F lie, both Lemma 1 and Lemma 2. That is, Lemma 1 is trivial in E_1 (since we can simply choose $M_1(F) \geq \max_{\phi \in F} \|\phi\|$, or as an arbitrary positive number if F is empty), and we prove the lemmas therefore by proving the following two propositions:

PROPOSITION 1. *The truth of Lemma 1 in E_{n-1} implies the truth of Lemma 2 in E_n, for $n = 2, 3, \cdots$.*

PROPOSITION 2. *The truth of Lemma 2 in E_n and Lemma 1 in E_{n-1} implies the truth of Lemma 1 in E_n, for $n = 2, 3, \cdots$.*

Clearly, these suffice to establish the truth of both lemmas.

We first verify Proposition 1. Note that the proof is trivial if $F \subset \hat{z}^{\perp}$. Thus assume $F \not\subset \hat{z}^{\perp}$ and define $\delta = \frac{1}{5} \min_{\phi \notin \hat{z}^{\perp}, \phi \in F} |\phi \cdot \hat{z}| > 0$. Again, let $b = \max_{\phi \in F} \|\phi\|$ and let $\overline{M} = \max(b, M_{n-1}(F^*(\hat{z})))$. (Note that the case $F^*(\hat{z})$ empty is quite possible, but presents no special problems.) Take

$$V(\hat{z}) = \{\hat{x} : \|\hat{x} - \hat{z}\| < \delta/\overline{M}\}, \qquad K(\hat{z}) = \overline{M}^2/\delta.$$

It suffices to show that if $\hat{B} \in V(\hat{z})$ and $\|B\| > K(\hat{z})$, and if $B, \cdots, C$, $C + \phi$ is a proper F-chain (with possibly $B = C$) for which the sub-chain $B, \cdots, C$ (which is also a proper F-chain) lies entirely in $B + \hat{z}^{\perp}$, then $\phi \in \hat{z}^{\perp}$. But as we observed previously (Lemma 3), the projection $\tilde{B}, \cdots, \tilde{C}$ of the chain $B, \cdots, C$ onto $\hat{z}^{\perp}$ may be regarded as a proper $F^*(\hat{z})$-chain in E_{n-1}. Since the hypothesis of Proposition 1 is that Lemma 1 is true in E_{n-1}, there exists a number $M_{n-1}(F^*(\hat{z}))$ such that $\|\tilde{C} - \tilde{B}\| \leq M_{n-1}(F^*(\hat{z}))$. Moreover, $C - B = \tilde{C} - \tilde{B}$, and hence $\|C - B\| \leq M_{n-1}(F^*(\hat{z})) \leq \overline{M}$. Thus

$$\|\hat{C} - \hat{z}\| \leq \|\hat{C} - \hat{B}\| + \|\hat{B} - \hat{z}\| < \|C - B\|/(\|C\| \|B\|)^{1/2} + \delta/\overline{M}$$

$$\leq \overline{M}/\|B\| + \delta/\overline{M} \leq 2\delta/\overline{M}.$$

But

(4) $$0 \geq C \cdot \phi = \|C\| \hat{z} \cdot \phi + (C - \|C\| \hat{z}) \cdot \phi$$

and

$$|(C - \|C\| \hat{z}) \cdot \phi| \leq \|C\| \|\hat{C} - \hat{z}\| \|\phi\| < 2\delta \|C\|.$$

Now if $\phi \not\subset \hat{z}^{\perp}$, we would have $|\|C\| \hat{z} \cdot \phi| \geq 5\delta \|C\|$. It would then follow

from (4) that $\hat{z}\cdot\phi<0$ and $C\cdot\phi<-3\delta\|C\|$. Thus we would obtain, since $(C-B)\cdot\hat{z}=0$,

$$\|C+\phi\|^2-\|B\|^2 = \|\phi\|^2 + 2C\cdot\phi + \|C-B\|^2 + 2(C-B)\cdot(B-\|B\|\hat{z})$$
$$< \overline{M}^2 - 6\delta\|C\| + \overline{M}^2 + 2\overline{M}\|B\|\delta/\overline{M}$$
$$\leqq 2\overline{M}^2 - 4\delta\|B\| \leqq -2\overline{M}^2 < 0.$$

But this contradicts the assumption that $B, \cdots, C, C+\phi$ is proper. Hence $\phi\in\hat{z}^\perp$, which completes the proof of Proposition 1.

We now prove Proposition 2, using first the Heine-Borel covering theorem to select a finite set of neighborhoods $V(\hat{z}^{(1)}), \cdots, V(\hat{z}^{(N)})$ of unit vectors $\hat{z}^{(1)}, \cdots, \hat{z}^{(N)}$ such that the union of these neighborhoods covers the surface of the unit sphere in E_n. Set

$$K = \max_{1\leqq m\leqq N} K(\hat{z}^{(m)}).$$

We shall show that if $B_1, \cdots, B_j$ is any proper F-chain, then equation (3) holds with

$$M_n(F) = 2K + b + \max_{1\leqq m\leqq N} M_{n-1}(F^*(\hat{z}^{(m)})).$$

(as before, $b=\max_{\phi\in F}\|\phi\|$). This is clearly true if $\|B_j\|\leqq K$. If $\|B_j\|>K$, let t be the smallest integer for which $\|B_i\|\geqq K$ for all $i\geqq t$. Decompose the chain $B_t, \cdots, B_j$ into a sequence of *proper* subchains, each beginning with a vector of smaller norm than did the preceding subchain (by listing the vectors $B_t, \cdots, B_j$ in order and starting a new subchain whenever a vector is encountered of norm smaller than that of the initial vector of that subchain). Let the last such subchain be $B_s, \cdots, B_j$, and note that

$$K \leqq \|B_s\| \leqq \|B_t\|.$$

It is clear that $\|B_s-B_1\|\leqq 2K+b$, since either $\|B_1\|\geqq K$ and so $s=t=1$, or $\|B_1\|<K$, $\|B_s\|\leqq\|B_t\|<K+b$. Thus it suffices to show that

(5) $$\|B_j - B_s\| \leqq \max_{1\leqq m\leqq N} M_{n-1}(F^*(\hat{z}^{(m)})).$$

But since a hypothesis of Proposition 2 is that Lemma 2 is true in E_n, we know that, since $\hat{B}_s\in V(\hat{z}^{(l)})$ for some integer l in the set $1\leqq l\leqq N$ and $\|B_s\|\geqq K\geqq K(\hat{z}^{(l)})$, the chain $B_s, \cdots, B_j$ lies in $B_s +(\hat{z}^{(l)})^\perp$. Then, just as in the proof of Proposition 1, we can project the chain onto $(\hat{z}^{(l)})^\perp$ and employ Lemma 3 and the hypothesis that Lemma 1 holds in E_{n-1} to verify (5). Thus Proposition 2 is established.

5. Boundedness Theorem. From Lemma 1, we obtain the Boundedness Theorem as an easy corollary:

THEOREM 1. *Let $A_1, A_2, \cdots, A_j$ be a sequence of vectors generated by the error-correction procedure* (2). *Then $\|A_j\| \leq \|A_1\| + M$, where M is a constant depending on the set F, but not on A_1.*

If the chain $A_1, \cdots, A_j$ is proper, the proof is immediate from the lemma, since

$$\|A_j\| - \|A_1\| \leq \|A_j - A_1\| \leq M_n(F).$$

If it is not proper, it can be decomposed into a sequence of proper F-chains each beginning with a vector of smaller norm than did the preceding subchain, as in the proof of Proposition 2. If $A_i, \cdots, A_j$ is the last such subchain, then

$$\|A_j\| \leq \|A_j - A_i\| + \|A_i\| \leq M_n(F) + \|A_1\|.$$

This completes the proof of the theorem. Note that, in any case, we may choose $M = M_n(F)$, where n is the dimension of the n-dimensional Euclidean space E_n of which F is a subset.

6. Concluding remarks. By using Lemma 1, one can now complete the proof in Minsky and Papert [2, p. 182, Theorem 11.9] of their further result that for any $\epsilon > 0$, there exists an $N(\epsilon, F)$ such that if $B, \cdots, C$ is a proper F-chain with $\|B\| > N(\epsilon, F)$, then $\|C\| < \|B\| + \epsilon$.

Further, the Boundedness Theorem remains true if F consists of a finite set of vectors in a Hilbert space H. To see this, note that the set $A_1 \cup F$, where A_1 is the initial vector, is a finite set; and hence spans a finite dimensional subspace of H, isomorphic to E_n for some n, inside of which the whole process takes place. The theorem is then applied to the vectors of the set $A_1 \cup F$ regarded as lying in E_n.

Also, it is worth repeating the interesting observation of Minsky and Papert that Theorem 1 implies that if the finite set of vectors in F are constrained to have integer coordinates, then for each given initial vector, the process is finite state, i.e., only finitely many of the A_i given by (2) are distinct. This is also true of course if the coordinates are rational numbers, since F is finite.

It would be interesting to see an effective estimate of the bound M. It would also be interesting and useful to know if there is a method of selecting the ϕ_i from among those satisfying $A_i \cdot \phi_i \leq 0$ (e.g. the ϕ that has been waiting longest, or the one which maximizes $-A_i \cdot \phi$, or by cycling through the list of ϕ's in F in a fixed order) and a test on the vectors A_i (e.g. the rate of growth of $\|A_i\|$) by means of which one could decide as to whether or not a solution exists.

BIBLIOGRAPHY

1. B. Efron, *The perceptron correction procedure in nonseparable situations*, Rome Air Development Center Technical Documentary Report RADC-TDR-63-533, February 1964.

2. M. Minsky and S. Papert, *Perceptrons*, M.I.T. Press, Cambridge, Mass., 1969.

CORNELL UNIVERSITY, ITHACA, NEW YORK 14850