KYBERNETIKA ČÍSLO 2, ROČNÍK 5/1969

The Convergence of One Group of Correction Training Procedures

SVATOPLUK BLÁHA

The most of all training procedures for single threshold unit have a common feature: the change of weight vector depends directly on the pattern vector. The convergence condition can be formulated uniformly for all methods. There it is shown, that convergence of training procedures depends on the angle between the weight vector and the solution vector. A special conditions can be obtained by simple application of common principle on the individual methods. The principle can be used also when dead-zone of threshold element cannot be omitted.

1. INTRODUCTION*

Let $\mathscr{X}$ be a set of patterns each of them characterized by the d real numbers x_1, $x_2, \ldots, x_d$. We can take the individual numbers for components of the pattern vector $\mathbf{X}$ or point $\mathbf{X}$ in d-dimensional Euclidean space E_d. Let each pattern $\mathbf{X} \in \mathscr{X}$ belong to one of two subsets $\mathscr{X}_1$ and $\mathscr{X}_2$. We shall suppose, that these subsets are linearly separable, i.e. there exists such scalar linear function of vector $\mathbf{X}$ (discriminant function)

$$(1.1) \qquad g(\mathbf{X}) = x_1 w_1 + x_2 w_2 + \ldots + x_d w_d + w_{d+1}$$

depending also on real parameters $w_1, \ldots, w_{d+1}$, that

$$(1.2) \qquad \begin{aligned} g(\mathbf{X}) > 0 \quad &\text{for all} \quad \mathbf{X} \in \mathscr{X}_1 , \\ g(\mathbf{X}) < 0 \quad &\text{for all} \quad \mathbf{X} \in \mathscr{X}_2 . \end{aligned}$$

It is usefull to augment the set of components of $\mathbf{X}$ by a $(d + 1)$st component, whose value is always equal to $+1$. We shall denote this augmented pattern vector

* More details about the problems in Introduction can be found in [3].

by the symbol **Y**. The components of vector **Y** will be given by the equations

$$(1.3) \qquad y_i = x_i\,; \quad i = 1, 2, \ldots, d\,,$$
$$y_D = +1\,; \quad D = d + 1\,.$$

The set of augmented pattern vectors will be denoted by $\mathscr{Y}$ and two subsets by the symbols $\mathscr{Y}_1$ and $\mathscr{Y}_2$.

We can consider the parameters $w_1, \ldots, w_{d+1}$ for the components of weight vector **W**. The discriminant function $g(\mathbf{X})$ can be written as a dot product of the vectors **Y** and **W**:

$$(1.4) \qquad g(\mathbf{X}) = \mathbf{Y} \,.\, \mathbf{W}\,.$$

The equation

$$(1.5) \qquad \mathbf{Y} \,.\, \mathbf{W} = 0$$

is for fixed weight vector **W** the equation of hyperplan, which is normal to the weight vector and is called the decision hyperplan and will be denoted by symbol W. The equation (1.5) is for the fixed pattern vector the equation of hyperplan normal to the pattern vector. This hyperplan is called the pattern hyperplan and will be signed by Y. Both hyperplanes pass throught the origin and each divides space E_D into two half-spaces, the positive and the negative one, in dependence on the signum of product **Y**. **W**. The denomination positive or negative half-space of weight vector (pattern vector) will be used. It is clear from (1.2), that all patterns **Y**, belonging to the subset $\mathscr{Y}_1$, are the vectors of positive half-space of the solution weight vector **W** and others are the vectors of the negative half-space. The solution weight vector is any weight vector, for which the function $g(\mathbf{X}) = \mathbf{Y} \,.\, \mathbf{W}$ has the quality (1.2).

Let the set $\mathscr{Y}'_2$ be a set of the negatives of the vectors in $\mathscr{Y}_2$. The negative of vector **Y** will be denoted $-\mathbf{Y}$ and has the opposite direction to the vector **Y**. The union of the subset $\mathscr{Y}_1$ and $\mathscr{Y}'_2$ is the adjusted training set $\mathscr{Y}'$ and all members of $\mathscr{Y}'$ are lying in the positive half-space of solution vector **W**. Two relations (1.2) can be replaced by one:

$$(1.6) \qquad \mathbf{Y}' \,.\, \mathbf{W} > 0 \quad \text{for all} \quad \mathbf{Y}' \in \mathscr{Y}'\,.$$

A classifying problem for two given subsets will be solved by generating a weight vector sequence $S_W = \{\mathbf{W}_1, \mathbf{W}_2, \ldots, \mathbf{W}_k, \ldots\}$ such, that beginning with some index r the vectors $\mathbf{W}_r = \mathbf{W}_{r+1} = \ldots$ satisfies the inequality (1.6). The initial weight vector $\mathbf{W}_1$ is arbitrary. The weight vector sequence is recursively generated from a training sequence $S_{Y'} = \{\mathbf{Y}'_1, \mathbf{Y}'_2, \ldots, \mathbf{Y}'_k, \ldots\}$ whose each vector is the member of set $\mathscr{Y}'$ and every element of $\mathscr{Y}'$ occurs infinitely often in the training sequence. A weight vector is corrected by using a correction rule, when the pattern was classified incorrectly. If we omit in the training set all correctly classified patterns and in the weight vector sequence all corresponding vectors, we shall obtain a reduced training sequence

$$S_{\hat{Y}} = \{\hat{\mathbf{Y}}_1, \hat{\mathbf{Y}}_2, \ldots, \hat{\mathbf{Y}}_k, \ldots\} \tag{1.7}$$

and a reduced weight vector set

$$S_{\hat{W}} = \{\mathbf{W}_1, \mathbf{W}_2, \ldots, \mathbf{W}_k, \ldots\}. \tag{1.8}$$

In this paper will be considered the adjusted training set $\mathcal{Y}'$ and reduced sequences $S_{\hat{Y}}$, $S_{\hat{W}}$ only.

2. THE PRINCIPLE OF THE TRAINING PROCEDURES

In the k-th step the pattern vector $\hat{\mathbf{Y}}_k$ is, by assumption, incrorrectly classified using the weight vector i.e., the weight vector $\hat{\mathbf{W}}_k$ lies on the negative side of the pattern hyperplan Y_k. The point $\hat{\mathbf{W}}_k$ must be moved to the positive side of the hyperplan Y_k, if the correct classification is requested after the correction. It is clear, that the optimal trajectory for moving point $\hat{\mathbf{W}}_k$, is the line normal to the hyperplane Y_k, i.e., perpendicular to the vector $\hat{\mathbf{Y}}_k$. A correction rule can be expressed for these cases by the equation

$$\hat{\mathbf{W}}_{k+1} = \hat{\mathbf{W}}_k + c_k\hat{\mathbf{Y}}_k, \tag{2.1}$$

where c_k is a positive number called correction increment. Individual training methods differ one from the other just by the determination of correction increment.

3. DISTINCT TRAINING PROCEDURES

There are many variants of the mentioned training principle known from literature. Let us mention the typical ones of them.

The fixed increment correction procedure

The correction increment of this procedures is

$$c_k = \lambda; \quad \lambda = \text{const}. \tag{3.1}$$

This procedure represents the first group of the training methods, which are characterized by the dependence of change of weight vector $\mathbf{W}_k$ on the pattern vector $\mathbf{Y}_k$ only. Other features of this procedure will be derived in chapter 6.

The fractional correction procedure

This procedure, which is also known as relaxation method, represents the second group of training methods. This group is characterized by the dependence of the weight vector change on the distance of the point $\hat{\mathbf{W}}_k$ from pattern hyperplane corresponding to the vector $\hat{\mathbf{Y}}_k$. It is not too difficult to verify that this distance δ is [3]

$$\delta = \frac{|\hat{\mathbf{W}}_k \cdot \hat{\mathbf{Y}}_k|}{|\hat{\mathbf{Y}}_k|}. \tag{3.2}$$

The correction increment is given by the relation

$$c_k = \lambda \frac{|\hat{\mathbf{W}}_k \cdot \hat{\mathbf{Y}}_k|}{\hat{\mathbf{Y}} \cdot \hat{\mathbf{Y}}_k}; \quad \lambda = \text{const.} > 0\,. \tag{3.3}$$

The fractional procedure for the classifier with the dead-zone

If a dead-zone of threshold element must be considered, the correction increment will be given by the relation

$$c_k = \lambda \frac{d|\hat{\mathbf{Y}}_k| + |\hat{\mathbf{W}}_k . \hat{\mathbf{Y}}_k|}{\hat{\mathbf{Y}}_k \cdot \hat{\mathbf{Y}}_k}; \quad d > 0\,, \tag{3.4}$$

where d represents the width of the dead-zone. The pattern $\hat{\mathbf{Y}}_k$ is correctly classified if $|\hat{\mathbf{W}}_k \cdot \hat{\mathbf{Y}}_k| > d$.

4. GEOMETRIC ARRANGEMENT OF VECTOR SPACE

All pattern vectors from a training set, which is adjusted by the way described above, lie in the positive half-space $\mathscr{W}^+$ of a solution vector $\mathbf{W}$. Let us consider some weight vector $\mathbf{W}_k$, $k = 1, 2, \ldots$, and a pattern vector $\hat{\mathbf{Y}}_k$. As both vectors are the members of reduced sequences, their scalar product is negative i.e., the pattern vector lies in the negative half-space of vector $\hat{\mathbf{W}}_k$. Let us denote it by $\mathscr{W}_k^-$. Generally each pattern vector from the reduced training sequence lies in some region $\mathscr{R}_k$, which is a penetrating of the positive half-space $\mathscr{W}^+$ and the negative half-space $\mathscr{W}_k^-$. This idea is illustrated for three-dimensional space E_3 in two perpendicular projections on Fig. 4.1. The new point $\hat{\mathbf{W}}_{k+1}$ will lie in the region $\overline{\mathscr{R}}_k$, which will be obtained by the translation of the region $\mathscr{R}_k$. This translation is determined by the vector $\hat{\mathbf{W}}_k$.

The solution vector $\mathbf{W}$ and the intersection of the decission hyperplanes W and W_k (determined by vectors $\mathbf{W}$ and $\mathbf{W}_k$) determine the hyperplane P, which divide the space E_D into two half-spaces $\mathscr{P}_1$ and $\mathscr{P}_2$. If the weight vector $\hat{\mathbf{W}}_k$ lies in $\mathscr{P}_1$, then region $\mathscr{R}_k$ is always in $\mathscr{P}_2$ and similarly $\overline{\mathscr{R}}_k$ is always in $\overline{\mathscr{P}}_2$.

In every step of the training procedure the convergence of the process depends on the positions of three vectors: $\mathbf{W}$, $\hat{\mathbf{W}}_k$ and $\hat{\mathbf{Y}}_k$. We shall denote the angles between

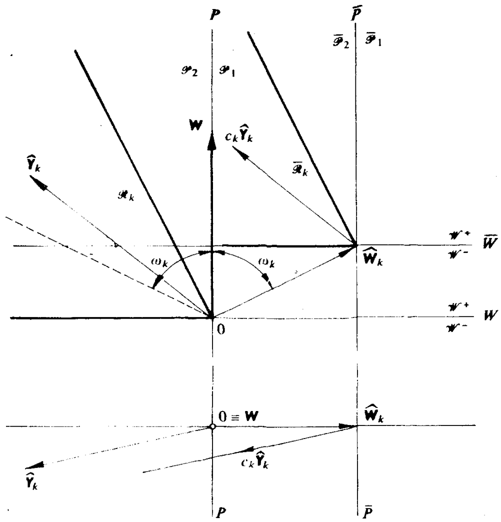

Fig. 4.1.

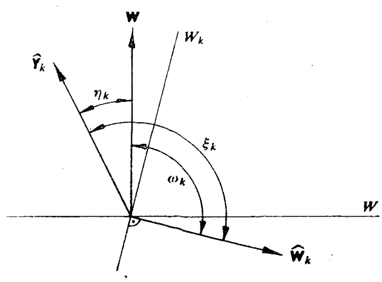

Fig. 4.2.

them (compare Fig. 4.2)

(4.1)
$$\begin{aligned} \sphericalangle\, \mathbf{W}, \hat{\mathbf{W}}_k &= \omega_k , \\ \sphericalangle\, \mathbf{W}, \hat{\mathbf{Y}}_k &= \eta_k , \\ \sphericalangle\, \hat{\mathbf{W}}_k, \hat{\mathbf{Y}}_k &= \xi_k . \end{aligned}$$

There is evident, that for any $D = 2, 3, \ldots$ is always valid in E_D:

$$(4.2) \qquad \begin{aligned} 0 &\leqq \omega_k \leqq \pi\,, \\ 0 &\leqq \eta_k \leqq \tfrac{1}{2}\pi\,, \\ \tfrac{1}{2}\pi &< \xi_k \leqq \pi\,. \end{aligned}$$

The training process will terminate, if such hyperplane W will be found, that all pattern vectors $\mathbf{Y} \in \mathscr{Y}'$ will be in the half-space $\mathscr{W}^+$. The hyperplane W pass through the origin and is determined by the solution vector $\mathbf{W}$. The absolute value of the solution vector is evidently irrelevant for this purpose.

It is obvious, that there exists more than one solution of the training process. Each of them lies in the solution region $\mathscr{W}_s$, the form of which is a polyhedral cone with vertex in the origin. The assumption that only one solution vector $\mathbf{W}$ exists is the strictest assumption and will be used in the next chapters. All, that will be said about one solution vector, will be valid, of course, for any other vector from the same solution region.

5. GENERAL CONVERGENCE CONDITIONS

Let's suppose there exists a solution weight vector $\mathbf{W}$ and an adjusted training set $\mathscr{Y}'$, which contains the patterns of two different classes. The changes of the weight vector $\mathbf{W}_k$, $k = 1, 2, \ldots$ are determined by any rule, the mathematical expression of which can be equation (2.1). As the patterns are classified only by signum of discriminant function the following definition of the training method convergence can be used:

A training method converges, if for angle ω_k between weight vector $\mathbf{W}_k$ and solution vector $\mathbf{W}$ is

$$(5.1) \qquad \lim_{k \to \infty} \omega_k = 0\,.$$

This equation can be satisfied only if it is possible to find such value of correction increment c_k that in every step (after every change of weight vector) is valid:

$$(5.2) \qquad \omega_{k+1} < \omega_k\,.$$

Since the scalar product of vectors is

$$(5.3) \qquad \mathbf{W} \,.\, \mathbf{W}_k = |\mathbf{W}|\,|\mathbf{W}_k| \cos \omega_k\,,$$

the relation (5.2) can be rewritten also using functions $\cos \omega$:

$$(5.4) \qquad \left(\frac{\mathbf{W} \,.\, \hat{\mathbf{W}}_k}{|\mathbf{W}|\,|\hat{\mathbf{W}}_k|}\right)^2 < \left(\frac{\mathbf{W} \,.\, \hat{\mathbf{W}}_{k+1}}{|\mathbf{W}|\,|\mathbf{W}_{k+1}|}\right)^2 \quad \text{for} \quad 0 \leqq \omega_k < \tfrac{1}{2}\pi$$

and is valid conversely for $\frac{1}{2}\pi < \omega_k \leqq \pi$. Using the equation (2.1) is easy to arrange the inequality (5.4) to the form

$$(5.5)\qquad \frac{(\mathbf{W}\,.\,\hat{\mathbf{W}}_k)^2}{\hat{\mathbf{W}}_k\,.\,\hat{\mathbf{W}}_k} < \frac{(\mathbf{W}\,.\,\hat{\mathbf{W}}_k)^2 + 2c_k(\mathbf{W}\,.\,\hat{\mathbf{W}}_k)(\mathbf{W}\,.\,\hat{\mathbf{Y}}_k) + c_k^2(\mathbf{W}\,.\,\hat{\mathbf{Y}}_k)^2}{\hat{\mathbf{W}}_k\,.\,\hat{\mathbf{W}}_k + 2c_k\hat{\mathbf{W}}_k\,.\,\hat{\mathbf{Y}}_k + c_k^2\hat{\mathbf{Y}}_k\,.\,\hat{\mathbf{Y}}_k}\,.$$

It is easy to verify, that the inequality (5.5) will be fulfilled when

$$(5.6)\qquad \frac{2c_k(\mathbf{W}\,.\,\hat{\mathbf{W}}_k)(\mathbf{W}\,.\,\hat{\mathbf{Y}}_k) + c_k^2(\mathbf{W}\,.\,\mathbf{Y}_k)^2}{(\mathbf{W}\,.\,\mathbf{W}_k)^2} > \frac{2c_k\hat{\mathbf{W}}_k\,.\,\hat{\mathbf{Y}}_k + c_k^2\hat{\mathbf{Y}}_k\,.\,\hat{\mathbf{Y}}_k}{\hat{\mathbf{W}}_k\,.\,\hat{\mathbf{W}}_k}$$

and from this relation we shall obtain with assumption that $c_k > 0$:

$$(5.7)\qquad c_k[|\hat{\mathbf{Y}}_k|^2(\mathbf{W}\,.\,\hat{\mathbf{W}}_k)^2 - |\hat{\mathbf{W}}_k|^2(\mathbf{W}\,.\,\hat{\mathbf{Y}}_k)^2] <$$
$$< 2[|\hat{\mathbf{W}}_k|^2(\mathbf{W}\,.\,\hat{\mathbf{W}}_k)(\mathbf{W}\,.\,\hat{\mathbf{Y}}_k) - (\mathbf{W}\,.\,\mathbf{W}_k)^2(\hat{\mathbf{W}}_k\,.\,\hat{\mathbf{Y}}_k)]\,.$$

If $|\hat{\mathbf{Y}}_k|^2(\mathbf{W}\,.\,\hat{\mathbf{W}}_k)^2 - |\hat{\mathbf{W}}_k|^2(\mathbf{W}\,.\,\hat{\mathbf{Y}}_k)^2 > 0$ (it is equivalent to $\cos^2\omega_k - \cos^2\eta_k > 0$), then will be evidently

$$(5.8)\qquad c_k < 2\frac{|\hat{\mathbf{W}}_k|^2(\mathbf{W}\,.\,\hat{\mathbf{W}}_k)(\mathbf{W}\,.\,\hat{\mathbf{Y}}_k) - (\mathbf{W}\,.\,\hat{\mathbf{W}}_k)^2(\hat{\mathbf{W}}_k\,.\,\hat{\mathbf{Y}}_k)}{|\hat{\mathbf{Y}}_k|^2(\mathbf{W}\,.\,\hat{\mathbf{W}}_k)^2 - |\hat{\mathbf{W}}_k|^2(\mathbf{W}\,.\,\mathbf{Y}_k)^2}$$

and in the opposite case the inequality (5.8) will be valid conversely. We shall remind equations (4.1) and multiply numerator and denominator of (5.8) by $(|\mathbf{W}|\,|\hat{\mathbf{W}}_k|\,|\hat{\mathbf{Y}}_k|)^2$ and arrange (5.8) to the form

$$(5.9)\qquad c_k < \frac{2|\hat{\mathbf{W}}_k|}{|\hat{\mathbf{Y}}_k|}\,\frac{\cos\omega_k(\cos\eta_k + |\cos\xi_k|\cos\omega_k)}{\cos^2\omega_k - \cos^2\eta_k}\,.$$

For economy of notation we define

$$(5.10)\qquad \Lambda_k = \frac{\cos\omega_k(\cos\eta_k + |\cos\xi_k|\cos\omega_k)}{\cos^2\omega_k - \cos^2\eta_k}\,.$$

The inequality (5.9) is correct, of course, only for $0 \leqq \omega_k < \frac{1}{2}\pi$ (i.e. for $\cos\omega_k > 0$) and for $\cos^2\omega_k - \cos^2\eta_k > 0$. The inequality (5.9) is valid conversely if $\cos^2\omega_k - \cos^2\eta_k < 0$. The relations for c_k, if $\frac{1}{2}\pi < \omega_k \leqq \pi$ can be calculated using this way and considering signum of denominator of Λ_k. Eventualy the relations for c_k can be written for clearness in table (5.11).

The region of possible angles ω_k and η_k, which is denoted by $\mathscr{A}$ consist of four subregions $\mathscr{A}_1, \ldots, \mathscr{A}_4$ defined also by relations (5.11). These regions are also illustrated on Fig. 5.1. Let us now investigate the properties of Λ_k in regions $\mathscr{A}_1, \ldots, \mathscr{A}_4$.

(5.11)

	$\cos^2 \omega_k - \cos^2 \eta_k > 0$	$\cos^2 \omega_k - \cos^2 \eta_k < 0$
$0 \leqq \omega_k < \frac{1}{2}\pi$ ($\cos \omega_k > 0$)	$\mathscr{A}_1$ $c_k < \dfrac{2\lvert\hat{\mathbf{W}}_k\rvert}{\lvert\hat{\mathbf{Y}}_k\rvert} \Lambda_k$	$\mathscr{A}_2$ $c_k > \dfrac{2\lvert\hat{\mathbf{W}}_k\rvert}{\lvert\hat{\mathbf{Y}}_k\rvert} \Lambda_k$
$\frac{1}{2}\pi < \omega_k \leqq \pi$ ($\cos \omega_k < 0$)	$\mathscr{A}_4$ $c_k > \dfrac{2\lvert\hat{\mathbf{W}}_k\rvert}{\lvert\hat{\mathbf{Y}}_k\rvert} \Lambda_k$	$\mathscr{A}_3$ $c_k < \dfrac{2\lvert\hat{\mathbf{W}}_k\rvert}{\lvert\hat{\mathbf{Y}}_k\rvert} \Lambda_k$

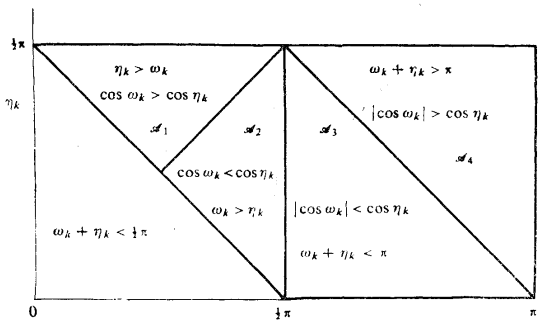

Fig. 5.1.

The region $\mathscr{A}_1$

At the end of training process the angles ω_k, η_k and ξ_k converge to these values

(5.12)
$$\lim_{k\to\infty} \omega_k = 0 \quad \text{so that} \quad \lim_{k\to\infty} \cos \omega_k = 1\,,$$
$$\lim_{k\to\infty} \eta_k = \tfrac{1}{2}\pi_{(-)} \quad \text{so that} \quad \lim_{k\to\infty} \cos \eta_k = 0\,,$$
$$\lim_{k\to\infty} \xi_k = \tfrac{1}{2}\pi_{(+)} \quad \text{so that} \quad \lim_{k\to\infty} \lvert\cos \xi_k\rvert = 0\,.$$

Using this limits we shall compute, that

(5.13)
$$\lim_{k\to\infty} \Lambda_k = \lim_{k\to\infty} \frac{\cos \omega_k \cos \eta_k}{\cos^2 \omega_k - \cos^2 \eta_k} + \lim_{k\to\infty} \frac{\cos^2 \omega_k \lvert\cos \xi_k\rvert}{\cos^2 \omega_k - \cos^2 \eta_k} = 0\,.$$

 As in region $\mathscr{A}_1$ must be

$$0 < c_k < \frac{2|\hat{\mathbf{W}}_k|}{|\hat{\mathbf{Y}}_k|} \Lambda_k ,$$

it means, that the training process converge if the change of the weight vector can be sufficiently small. The function $\Lambda_k = \Lambda(\omega_k, \eta_k, \xi_{k\max})$ is illustrated on Fig. 5.2.

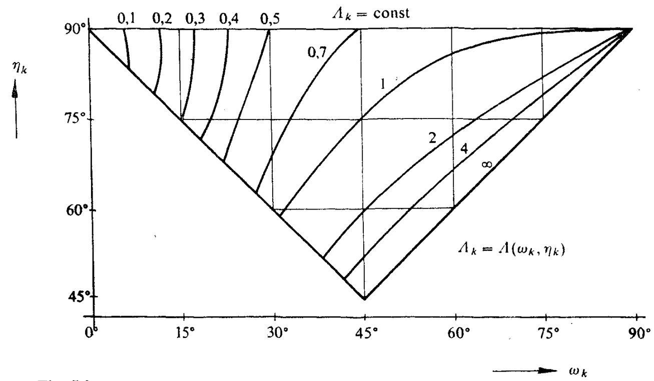

Fig. 5.2.

The region $\mathscr{A}_2$

In the region $\mathscr{A}_2(\omega_k + \eta_k > \frac{1}{2}\pi;\ \frac{1}{4}\pi \leqq \omega_k < \frac{1}{2}\pi)$ is Λ_k negative i.e., the respective convergence condition (5.11) is fulfilled for any positive c_k.

The region $\mathscr{A}_3$

In the region $\mathscr{A}_3(\frac{1}{2}\pi < \omega_k \leqq \pi;\ \omega_k + \eta_k < \pi)$ is $\cos \omega_k$ negative and denominator of Λ_k is negative too, because

(5.14) $$|\cos \omega_k| < \cos \eta_k .$$

Numerator of Λ_k will be positive if

(5.15) $$\frac{\cos \eta_k}{|\cos \omega_k|} > |\cos \xi_k| .$$

By (5.14) the fraction $\cos \eta_k / |\cos \omega_k| > 1$, so that inequality (5.15) is valid and

$\Lambda_k > 0$ in the region $\mathscr{A}_3$. The training methods will converge not only if

$$c_k < \frac{2|\hat{\mathbf{W}}_k|}{|\hat{\mathbf{Y}}_k|} \Lambda_k \,. \tag{5.16}$$

The convergence condition in $\mathscr{A}_3$ means that $|\cos \omega_{k+1}| < |\cos \omega_k|$, i.e.

$$\omega_{k+1} > \pi - \omega_k \tag{5.17}$$

and when the change of weight vector is so great (when is $c_k > 2|\hat{\mathbf{W}}_k| \, |\hat{\mathbf{Y}}_k|^{-1} \Lambda_k$) that inequality (5.17) is valid conversely, then respective convergence condition (5.11) must be converted too. For two-dimensional case is this situation illustrated on Fig. 5.3.

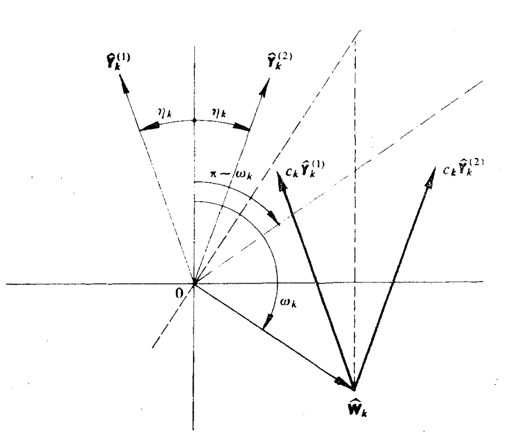

Fig. 5.3.

The training methods converge in region $\mathscr{A}_3$ also for any positive value of correction increment c_k.

The region $\mathscr{A}_4$

In the region $\mathscr{A}_4$ is $\cos \omega_k < 0$ and denominator of Λ_k is positive. Numerator of Λ_k can be positive, zero or negative if $\cos \eta_k / |\cos \omega_k| < |\cos \xi_k|$, $\cos \eta_k / |\cos \omega_k| = |\cos \xi_k|$ or $\cos \eta_k / |\cos \omega_k| > |\cos \xi_k|$. The possible values of angle ξ are defined by following relations

$$\begin{aligned} &\xi_k > \tfrac{1}{2}\pi \,, \quad \xi_k > \omega_k - \eta_k \,, \\ &\xi_k < \omega_k + \eta_k \quad \text{for} \quad \omega_k + \eta_k < \pi \,, \\ &\xi_k < 2\pi - \omega_k - \eta_k \quad \text{for} \quad \omega_k + \eta_k > \pi \,. \end{aligned} \tag{5.18}$$

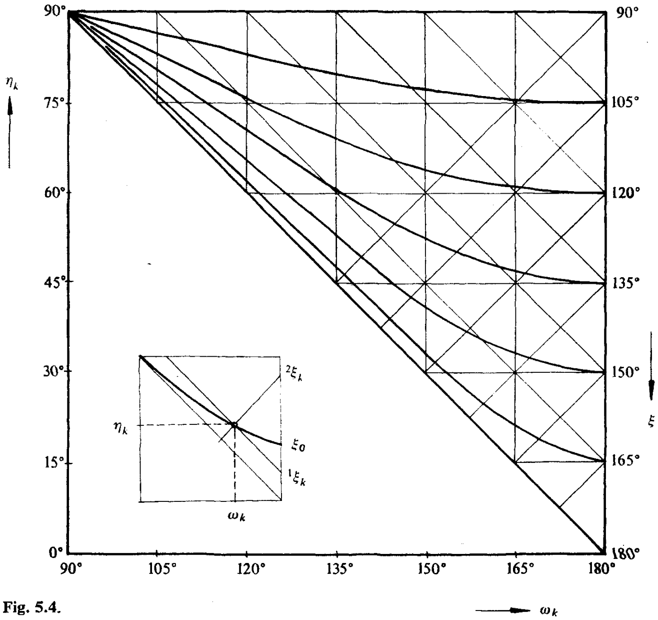

Fig. 5.4.

This situation is illustrated on Fig. 5.4. If the angles between vectors $\hat{\mathbf{W}}_k$, $\mathbf{Y}_k$ and $\mathbf{W}$ are ω_k and η_k, then the values of angle ξ_k between vectors $\hat{\mathbf{W}}_k$ and $\hat{\mathbf{Y}}_k$ can be ${}^1\xi_k \leqq \xi_k \leqq {}^2\xi_k$. Using the system of curves $\cos \eta_k / |\cos \omega_k| = \text{const}$, can be found the value ξ_0, for which $\cos \eta_k / |\cos \omega_k| = \cos \xi_0$ for each point (ω_k, η_k). For $\xi_k > \xi_0$ is $\cos \eta_k / |\cos \omega_k| < |\cos \xi_k|$ and for $\xi_k < \xi_0$ is $\cos \eta_k / |\cos \omega_k| > |\cos \xi_k|$.

The convergence condition in $\mathscr{A}_4$ is $c_k > 2|\hat{\mathbf{W}}_k| \, |\hat{\mathbf{Y}}_k|^{-1} \Lambda_k$. For determination of value of Λ_k is decisive the upper limit of ξ_k, i.e. $\xi_k = {}^2\xi_k$, because for this value is $\Lambda_k > 0$. The function $\Lambda_k = \Lambda(\omega_k, \eta_k, {}^2\xi_k)$ is limited and the highest value is $\Lambda_k = 1$ for $\Lambda(\omega_k, \eta_k, {}^2\xi_k) = \Lambda(\pi/2, \pi/2, \pi/2)$. The function Λ in $\mathscr{A}_4$ is shown on Fig. 5.5. In point $\omega = \frac{1}{2}\pi$, $\eta = \frac{1}{2}\pi$ the function is undefined and its value is $0{\cdot}5 \leqq \Lambda_k \leqq 1$, similarly in point $\omega_k = \pi$, $\eta_k = 0$ where $0 \leqq \Lambda_k \leqq 0{,}5$.

The training methods in region $\mathscr{A}_4$ converge always if the value of c_k is

$$c_k > 0 \quad \text{for} \quad \xi_k < \xi_0 \quad (\text{i.e. } \Lambda_k < 0)\,, \tag{5.19}$$

$$c_k > \frac{2|\hat{\mathbf{W}}_k|}{|\hat{\mathbf{Y}}_k|} > \frac{2|\hat{\mathbf{W}}_k|}{|\hat{\mathbf{Y}}_k|} \Lambda_k \quad \text{for} \quad \xi_k > \xi_0 \quad (\Lambda_k > 0)\,.$$

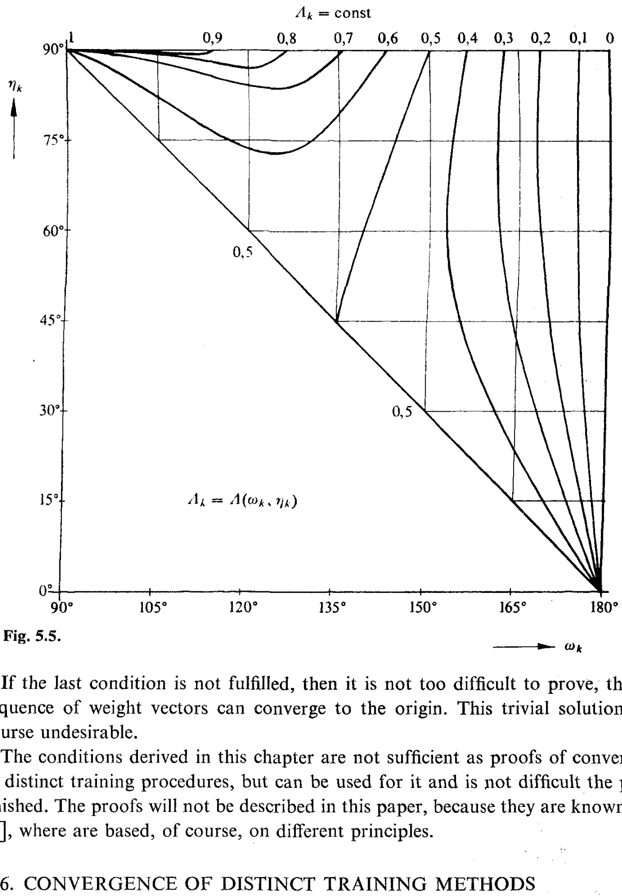

Fig. 5.5.

If the last condition is not fulfilled, then it is not too difficult to prove, that the sequence of weight vectors can converge to the origin. This trivial solution is of course undesirable.

The conditions derived in this chapter are not sufficient as proofs of convergence of distinct training procedures, but can be used for it and is not difficult the proofs finished. The proofs will not be described in this paper, because they are known from [3], where are based, of course, on different principles.

6. CONVERGENCE OF DISTINCT TRAINING METHODS

The conditions for the convergence of the distinct training procedures will be obtained from the relation (5.11) if a particular expression for the correction increment c_k will be put to it. Let us derive as an example such specific conditions for the training procedures (3.1), (3.3) and (3.4). It will be derived convergence conditions only for region $\mathscr{A}_1$, because in other regions the conditions are clear and very simple.

By the relation (5.11) the training procedure converges in $\mathscr{A}_1$ if

$$\lambda < \frac{2|\hat{\mathbf{W}}_k|}{|\hat{\mathbf{Y}}|_{\max}} \Lambda_k \tag{6.1}$$

where $|\hat{\mathbf{Y}}|_{\max} = \max_{k=1,2\ldots} |\hat{\mathbf{Y}}_k|$. For $\lambda = $ const. is

$$\lim_{k\to\infty} |\hat{\mathbf{W}}_k| = \lim_{k\to\infty} \frac{\lambda|\hat{\mathbf{Y}}|_{\max}}{2\Lambda_k} = \infty \tag{6.2}$$

because by (5.13) is $\lim_{k\to\infty} \Lambda_k = 0$.

The convergence condition for fixed increment training procedure will be fulfilled for linear case, when absolute value of weight vector will grow over all limits.

The fractional training procedure

The convergence condition in $\mathscr{A}_1$ for this method is by (5.11) and (3.3)

$$\lambda \frac{|\hat{\mathbf{W}}_k \cdot \hat{\mathbf{Y}}_k|}{|\hat{\mathbf{Y}}_k|^2} < \frac{2|\hat{\mathbf{W}}_k|}{|\hat{\mathbf{Y}}_k|} \Lambda_k\,, \tag{6.3}$$

and after the simple arrangement will be obtained

$$\lambda|\cos \xi_k| \leqq 2\Lambda_k \tag{6.4}$$

and eventually

$$\lambda \leqq 2 \frac{\cos \omega_k(\cos \eta_k + |\cos \xi_k| \cos \omega_k)}{|\cos \xi_k|\,(\cos^2 \omega_k - \cos^2 \eta_k)}\,. \tag{6.5}$$

The expression $\Lambda_k/|\cos \xi_k|$ can be written

$$\frac{\Lambda_k}{|\cos \xi_k|} = \frac{\cos \omega_k \cos \eta_k}{|\cos \xi_k|(\cos^2 \omega_k - \cos^2 \eta_k)} + \frac{\cos^2 \omega_k}{\cos^2 \omega_k - \cos^2 \eta_k} \tag{6.6}$$

and the following relations are evidently valid. By (5.11)

$$\frac{\cos \omega_k \cos \eta_k}{|\cos \xi_k|\,(\cos^2 \omega_k - \cos^2 \eta_k)} > 0 \quad \text{for all} \quad k\,, \tag{6.7}$$

$$\frac{\cos^2 \omega_k}{\cos^2 \omega_k - \cos^2 \eta_k} > 1 \quad \text{for all} \quad k\,, \tag{6.8}$$

and

$$\lim_{k\to\infty} \frac{\cos^2 \omega_k}{\cos^2 \omega_k - \cos^2 \eta_k} = 1 \tag{6.9}$$

hold, so that

$$\lim_{k\to\infty} \frac{\Lambda_k}{|\cos \xi_k|} = 1_{(+)} . \tag{6.10}$$

The convergence condition for the fractional training procedure valid in region $\mathscr{A}_1$ is

$$\lambda \leqq 2 , \tag{6.11}$$

which is the well known result, derived in another way in [2] or [3].

The fractional training procedure with the dead-zone and the influence of the solution region

For this case it must be by (5.11) and (3.4)

$$\frac{\lambda}{|\hat{\mathbf{Y}}_k|} \left(d + \frac{|\hat{\mathbf{W}}_k \cdot \hat{\mathbf{Y}}_k|}{|\hat{\mathbf{Y}}_k|} \right) < \frac{2|\hat{\mathbf{W}}_k|}{|\hat{\mathbf{Y}}_k|} \Lambda_k . \tag{6.12}$$

After the arrangement we shall obtain

$$\lambda < 2|\hat{\mathbf{W}}_k| \frac{\Lambda_k}{d + \hat{\mathbf{W}}_k| |\cos \xi_k|} . \tag{6.13}$$

As it is known that $\lim_{k\to\infty} \Lambda_k = 0$ and also $\lim_{k\to\infty} |\cos \xi_k| = 0$ it is evident that

$$\lim_{k\to\infty} \frac{\Lambda_k}{d + |\hat{\mathbf{W}}_k| |\cos \xi_k|} = 0 . \tag{6.14}$$

Since we suppose, that $\lambda = \text{const.}$, it must be

$$\lim_{k\to\infty} |\hat{\mathbf{W}}_k| = \infty . \tag{6.15}$$

The fractional training procedure for the case, when the dead-zone of the threshold element cannot be omitted, will converge, if the absolute value of weight vector can grow to infinity.

The specific convergence conditions for all variants of the mentioned typical training procedures can be derived in the same or in a similar way.

The circumstances which influence the convergence of the training procedure, are not so unfavourable in practical case as it was supposed in the previous chapters. Especialy the solution region – it will be denoted by $\mathscr{W}_s$ – has a positive influence contrary ro single solution weight vector which was assumed till now.

Let us suppose to simplify the matter that the form of the solution region $\mathscr{W}_s$ is a cone with vertex in the origin and solution vector $\boldsymbol{W}$ lies in the axis of this cone. The angle between vector $\boldsymbol{W}$ and any vector on the boundary of $\mathscr{W}_s$ is ω_s. The next relations are valid in region $\mathscr{A}_1$:

$$\omega_s < \omega_k < \tfrac{1}{2}\pi \quad \text{so that} \quad 0 < \cos \omega_k < \cos \omega_s\,, \tag{7.1}$$

$$0 < \eta_k < \tfrac{1}{2}\pi - \omega_s \quad \text{so that} \quad \sin \omega_s < \cos \eta_k\,, \tag{7.2}$$

$$\tfrac{1}{2}\pi < \xi_k < \tfrac{1}{2}\pi + \omega_k - \omega_s \quad \text{so that} \quad 0 < |\cos \xi_k| < 1 \tag{7.3}$$

and as before (5.12)

$$\lim_{k\to\infty} \eta_k = 0\,, \tag{7.4}$$

$$\lim_{k\to\infty} |\cos \xi_k| = 0\,. \tag{7.5}$$

Using (7.2) and (7.4) we can write after a simple arrangement

$$\Lambda_k > \frac{\cos \omega_k \sin \omega_s}{\cos^2 \omega_k} + |\cos \xi_k\,. \tag{7.6}$$

Since the angle ω_k decrease to the ω_s (for $\omega_k < \omega_s$ the training process terminate), the equation (7.6) can be re-written:

$$\Lambda_k > \operatorname{tg} \omega_s + |\cos \xi_k|\,. \tag{7.7}$$

If this relation will be used in inequality (5.11) for $\mathscr{A}_1$, we shall obtain

$$c_k \leqq \frac{2|\hat{\boldsymbol{W}}_k|}{|\hat{\boldsymbol{Y}}_k|}(\operatorname{tg} \omega_s + |\cos \xi_k|)\,. \tag{7.8}$$

The specific convergence conditions for the distinct training methods can be derived again.

From (7.8) and (3.1) we shall obtain considering (7.5) yet

$$(7.9) \qquad \frac{\lambda}{2 \operatorname{tg} \omega_s} |\hat{\mathbf{Y}}|_{\max} \leqq |\hat{\mathbf{W}}_k| \, ; \quad |\hat{\mathbf{Y}}|_{\max} = \max_{k=1,2,\dots} |\hat{\mathbf{Y}}_k| \, .$$

For some $|\hat{\mathbf{W}}|_{\max}$ a convenient value of λ can be calculated.

The fractional training procedure

From (7.8) and (3.3) we shall obtain easily

$$(7.10) \qquad \lambda \leqq \frac{\operatorname{tg} \omega_s}{|\cos \xi_k|} + 1 \, .$$

The value of fraction $\operatorname{tg} \omega_s / |\cos \xi_k| \to \infty$ for $k \to \infty$. The possible increasing of the value of λ must be considered, of course, in connection with the decreasing of the distance δ (3.2) to zero.

The fractional training procedure with dead-zone

From (7.8), (3.4) and (7.5) we shall obtain

$$(7.11) \qquad \lambda \leqq 2 |\hat{\mathbf{W}}_k| \frac{\operatorname{tg} \omega_s}{d}$$

or

$$(7.12) \qquad \frac{\lambda d}{2 \operatorname{tg} \omega_s} \leqq |\hat{\mathbf{W}}_k| \, .$$

It means that even for this case the convenient value of λ can be determined.

8. CONCLUSION

The contents of this paper can be summed up into several theorems. Their derivations are in previous chapters.

1. Only such training procedures will be considered, which can be described by equation (2.1).

2. The training procedures can always converge to the solution, if the ratio of absolute values of weight vector change and the weight vector can be arbitrarily small at the end of training process.

3. The influence of the solution region softens the convergence condition in all cases.

4. The fixed increment training procedure needs increasing of the weight vector's absolute value compared with initial vector. On the contrary the absolute value of the weight vectors are decreasing compared with previous weight vector if relaxation training procedure for $\lambda < 2$ is used.

5. The fixed increment training procedure converges always to the nonzero solution.

6. The training procedures of the relaxation type converge always to the point in the solution region or on its boundary. It means, that such point can be the origin; it is an undesirable trivial solution. The training procedure of the relaxation type has always a non-zero solution, if initial weight vector lies in the positive half-space of the solution vector.

7. If the dead-zone of the threshold element cannot be omitted, then relaxation methods need increasing of the weight vectors absolute value over all limits, too. This condition is softened also by the existence of the solution region.

8. The training procedure converge to the solution, if the angle between the solution vector and the weight vector decreases in every step. The convergence is independent of the absolute values of vectors.

Many practical cases are known which converge to the solution though does not fulfil the demand of unlimited values of weights. It would be necessary to complete the conditions of convergence also for the case when the limitations of weight must be considered. But the derivation of such conditions is not simple. The position of weight vector compared with solution vector, type of correction rule and "width" of solution region have influence on it.

(Received April 26th, 1968.)

REFERENCES

[1] S. Agmon: The Relaxation Method for Linear Inequalities. Canadian Journal of Mathematics *VI* (1954), 382—392.

[2] T. S. Motzkin, I. J. Schoenberg: The Relaxation Method for Linear Inequalities. Canadian Journal of Mathematics *VI* (1954), 393—404.

[3] N. J. Nilsson: Learning Machines. McGraw-Hill Book Comp., N. York 1965.

[4] C. H. Mays: Effect of Adaptation Parameters on Convergence Time and Tolerance for Adaptive Threshold Elements. IEEE Transaction on Electronics Computers (August 1964), 465—468.

Konvergence jedné skupiny korekčních trenovacích metod

SVATOPLUK BLÁHA

Většina trenovacích metod pro učící se stroje má společnou vlastnost: změna váhového vektoru je přímo úměrná prvkovému vektoru. Podmínky, které musí být splněny, aby trenovací metody konvergovaly k řešení lze formulovat jednotně pro všechny varianty uvedeného principu. Podmínky jsou odvozeny na základě skutečnosti, že metody konvergují tehdy a jen tehdy, zmenšuje-li se v každém kroku úhel mezi váhovým vektorem a vektorem řešení. Odvozené podmínky dávají dobrý obraz o průběhu trenovacího procesu.

Aplikací obecného vztahu na jednotlivé trenovací metody snadno obdržíme speciální podmínky konvergence, platné pro zvolenou metodu. Popsaný princip je možno použít i v tom případě, že nelze zanedbat pásmo necitlivosti prahového členu.

Ing. Svatopluk Bláha, CSc., Ústav teorie informace a automatizace ČSAV, Vyšehradská 49, Praha 2.