IJESRT

INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY

ONE INCREMENTAL ADAPTATION STRATEGY OF THE DECISION-MAKING FORMAL NEURON

Archil Prangishvili, Natia Namicheishvili *
Faculty of Informatics and Control Systems, Georgian Technical University, Tbilisi, Georgia

ABSTRACT

The general objective of formal neuron adaptation is to give the reliable inputs more influence in determining the output than the unreliable inputs have. Adaptation is introduced into the threshold decision element by a circuit that performs two operations: estimates the error probability of each input and uses the estimate to change each vote-weight. A cyclic error-counting adaptation procedure is the one in which vote-weights are changed periodically (cyclically), based on data collected during computations in a period (cycle). If appropriate incremental adjustments are made to the estimate of inputs' each error after each computation, then we deal with continuous adaptation. There are two methods for detecting errors: comparison with the output decision (closed-loop adaptation) and comparison with an externally supplied correct answer (open-loop adaptation). In this paper, the input weighted sum is compared with the desired value of this sum, which is multiplied by the correct binary answer. Our results for incremental changes of the input weights of the adaptive formal neuron are based on the Widrow-Hoff algorithm and stochastic approximation method.

Keywords: vote-weight, Mahalanobis distance, entropy sensitivity criterion, Widrow-Hoff algorithm, Robbins-Monro algorithm.

INTRODUCTION

Let us suppose that the binary signal x coded, say, as $+1$ and -1 is supplied to n same-type data buses $B_1, B_2, \dots, B_n$. Because of probable errors of the buses, the value of the variable x is computed as $x_1, x_2, \dots, x_n$.

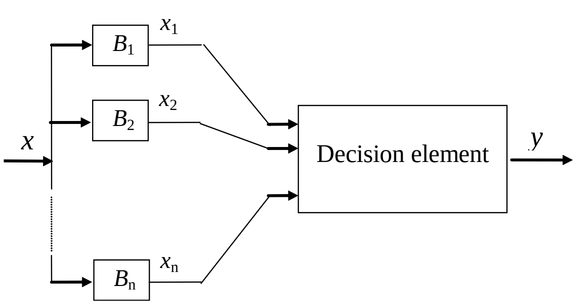

Fig. 1: ***Decision element (gate)***

As a result, n versions are obtained for the value of the variable x supplied for recognition. Clearly, each of the values $x_i (i = \overline{1, n})$ is a binary digit taking values $+1$ and -1. This redundant information (in the form of n versions for the value of the variable x) further arrives at the inputs of the so-called decision or restoration element whose model is shown in Fig. 1.

As is known, the decision element is a device which uses the known binary signals $x_1, x_2, \dots, x_n$ at n inputs to determine a binary output signal y called a decision. Saying it differently, the decision element is a switching circuit that realizes some binary function y of n binary arguments $x_1, x_2, \dots, x_n$:

$$y = f(x_1, x_2, \dots, x_n). \qquad (1.1)$$

The reliability of the decision element essentially depends on a type of function (1.1) realized by this element. It is obvious that in the ideal case the decision y produced by the decision element must coincide with the true value of the binary variable x.

The decision element that realizes the function

ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 2.114

$$y = \operatorname{sgn}\left(\sum_{i=1}^{n} x_i\right), \qquad (1.2)$$

where

$$\operatorname{sgn} z = \begin{cases} -1, \text{ if } z < 0 \\ \ \ 0, \text{ if } z = 0 \\ +1, \text{ if } z > 0 \end{cases}, \qquad (1.3)$$

is called a «majority-rule» decision element. Here zero stands for an unreliable (indeterminate) value of y. For the concrete input combination $x_1, x_2, \ldots, x_n$, the unreliability (indeterminacy) of the output variable y means that either this combination is never realized (n is an odd number in the sum $z = x_1 + x_2 + \ldots + x_n$) or for $z = 0$ the decision y is not produced at all. The graph of the function $y = \operatorname{sgn} z$ is shown in Fig. 2. Here the arrows indicate that the cusp does not belong to the graph, while the point at the origin indicates that the output variable y is unreliable.

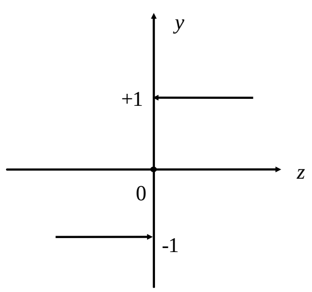

Fig. 2: ***Graph of the function y=sgnz***

The main components of the «majority–rule» decision element are a summation device which gives the output signal

$$z = \sum_{i=1}^{n} x_i ,$$

and a nonlinear two-terminal circuit with the characteristic $y = \operatorname{sgn} z$ shown in Fig. 3.

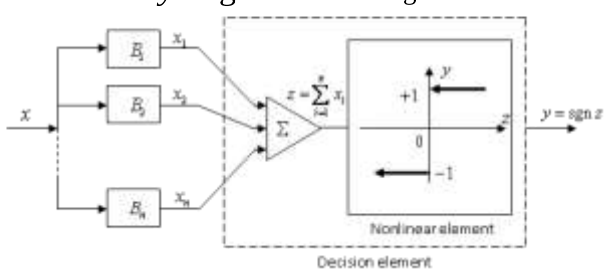

Fig. 3: ***Model of the «majority-rule» decision element***

It is quite obvious that this element gives the decision y as a result of voting by the principle of a simple majority of output signal values. For this reason it is often referred to as the voting element. The majority rule was first described by J. von Neumann in the work **[1]** and later extended by V.I. Varshavsky to analog systems with redundancy **[2]**. The majority rule was investigated in many other aspects, too, but we do touch upon them in the present paper.

The functioning of the restoration circuit with a «majority-rule» decision element cannot be considered satisfactory if the error probabilities $q_1, q_2, \ldots, q_n$ of the binary data buses $B_1, B_2, \ldots, B_n$ have different values so that to each x_i arriving from the output of the binary bus B_i to the i th input of the decision element we have to assign its weight $a_i \left(i = \overline{1,n}\right)$, where a_i is an arbitrary real number $\left(-\infty < a_i < +\infty\right)$. In that case, the output decision y must be given as a result of weighted voting according to the relation

$$y = \operatorname{sgn}\left(\sum_{i=1}^{n} a_i x_i - \Theta\right), \qquad (1.4)$$

where Θ is the so-called threshold or quorum of the element. Hence it is frequently called a threshold or quorum element though it might as well be called a weighted voting element. It is easy to see that the «majority-rule» decision element is a threshold element with weighted coefficients $a_i = 1\left(i = \overline{1,n}\right)$ and threshold $\Theta = 0$.

Let us formally assume that $\Theta = a_{n+1}$, and $x_{n+1} \equiv -1$. The latter implies that there exists some phantom data bus B_{n+1} which always produces the output signal $x_{n+1} \equiv -1$ whatever the input signal x is. Then relation (1.4) can be written in the form

$$y = \operatorname{sgn}\left(\sum_{i=1}^{n+1} a_i x_i\right). \qquad (1.5)$$

The model of a decision element shown in Fig. 4 fully agrees with this relation.

ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 2.114

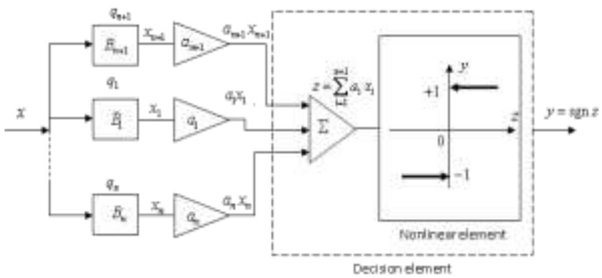

Fig. 4: ***Model of the decision element***

The need of such a model has arisen because if at the initial moment of time the error probabilities of data buses can be chosen practically equal to one another, in the course of time they nevertheless develop differences.

Threshold logic is quite well covered in the literature **[3]-[7]**. Basic results are related to the problem of synthesis of a threshold element (artificial neuron), i.e. it has been shown that the considered switching function of n binary variables can be realized by one threshold element (artificial neuron), and such weights and a threshold have been found that ensure the fulfillment of this function. Solutions have also been found for problems of the synthesis of networks of threshold elements, i.e. for problems of the synthesis of artificial neuron networks. However, on the whole, these studies are only superficially concerned with problems of threshold logic from the standpoint of making optimal decisions on the restoration of a true signal under the redundancy of data buses. Exactly this aspect of the threshold model of the functioning of a decision element in the correct signal restoration system interests us **[8]-[16]** and is the subject of investigation in the present paper.

Choice of vote-weights

Let us introduce the variable

$$Z = \sum_{i=1}^{n+1} a_i X_i \,, \qquad (2.1)$$

assuming that all x_i are random binary variables X_i because of the probability of error occurrence in the data buses B_i. Then it is obvious that Z, too, will be a random variable taking the value z on the real axis, and $Y = \operatorname{sgn} Z$ will be a random binary variable.

Besides, let us construct the random variable

$$\eta = X \cdot Z = \sum_{i=1}^{n+1} a_i \left(X \cdot X_i \right) = \sum_{i=1}^{n+1} \eta_i \,, \quad (2.2)$$

where X is the binary variable taking the values $+1$ and -1 supplied to the data buses $B_i \left(i = \overline{1, n+1} \right)$.

X is interpreted as a random variable with realizations x.

Realizations of the random variable η are denoted by v, and realizations of random variables η_i - by v_i. It is easy to see that the values of v are discrete and belong to the real axis, while the values of v_i are either $+a_i$ or $-a_i$ for all $i = \overline{1, n+1}$.

Let us turn our attention to random variables $X \cdot X_i \left(i = \overline{1, n+1} \right)$. It is easy to see that a discrete random value $X \cdot X_i$ takes the value -1 for $X_i = \neg X$ (where $\neg X$ denotes the inversion of the binary variable X) with probability q_i and the value $+1$ for $X_i = X$ with probability $1 - q_i$:

$$\left.\begin{array}{l} \text{Prob}\{X \cdot X_i = -1\} = \text{Prob}\{X_i \neq X\} = q_i \\ \text{Prob}\{X \cdot X_i = +1\} = \text{Prob}\{X_i = X\} = 1 - q_i \\ \quad i = \overline{1, n+1} \end{array}\right\}. \quad (2.4)$$

In particular,

$$\left.\begin{array}{l} q_{n+1} = \text{Prob}\{X \cdot X_{n+1} = -1\} = \text{Prob}\{X = +1\} \\ 1 - q_{n+1} = \text{Prob}\{X \cdot X_{n+1} = +1\} = \text{Prob}\{X = -1\} \end{array}\right\}, \quad (2.5)$$

since $X_{n+1} = -1$. Hence it follows that q_{n+1} is the a priori probability that the signal $X = +1$ will be supplied to the decision element for recognition, i.e. q_{n+1} is the a priori probability that at the threshold element output we shall have $+1$ as a correct signal.

Analogously, $1 - q_{n+1}$ is the a priori probability that the signal $X = -1$ will be supplied to the input of the decision element or, which is the same, the a priori probability that the output of the threshold element will be -1 as a correct decision.

The distributions $f\left(v_i\right)$ of probabilities of discrete random variables η_i are determined using the assumptions that $v_i = +a_i$ with probability $1 - q_i$ and $v_i = -a_i$ with probability q_i. Therefore, formally,

ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 2.114

$$\left.\begin{aligned} & f_i(v_i)=q_i^{(a_i-v_i)/2a_i}\cdot(1-q_i)^{(v_i+a_i)/2a_i} \\ & v_i=+a_i,-a_i \\ & i=\overline{1,n+1} \end{aligned}\right\}. \quad (2.6)$$

Criterion of Mahalanobis distance maximum

When solving the problem of weights optimization, we treat it as a classification problem, which means that the input signal x should be assigned to one of the two classes Ω_1 (when $x=+1$) and Ω_2 (when $x=-1$) on the basis of $n+1$ versions $x_1,x_2,\dots,x_n,x_{n+1}$ of this signal.

Under this approach, the restored signal should be considered as a random variable X with realization x. The characteristics of this variable are $n+1$ random variables $X_1,X_2,\dots,X_n,X_{n+1}$. It is helpful to consider X as a random vector, i.e. as an ordered set of $n+1$ numbers arranged as a column

$$\vec{X}=\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \\ X_{n+1} \end{bmatrix}=(X_1,X_2,\dots,X_n,X_{n+1})',$$

where the symbol «'» stands for the operation of transposition. Hence each element $X_i\,(i=\overline{1,n+1})$ serves as a component of the random vector X.

A realization of the random vector $\vec{X}$ is written as the observations vector

$$\vec{x}=(x_1,x_2,\dots,x_n,x_{n+1})',$$

where the components of the vector $\vec{x}$ are the realizations $x_1,x_2,\dots,x_n,x_{n+1}$ of the random variables $X_1,X_2,\dots,X_n,X_{n+1}$, respectively.

The vector $\vec{X}$ with components $X_1,X_2,\dots,X_n,X_{n+1}$ is described by the joint distribution function

$$f(\vec{x})=\text{Prob}\{X_1=x_1,X_2=x_2,\dots,X_n=x_n,X_{n+1}=x_{n+1}\},$$

where the probable values of $x_i\,(i=\overline{1,n+1})$ are equal to -1 and $+1$.

It is easy to see that in the class Ω_1 the vector $\vec{X}$ has the distribution

$$f_1(\vec{x})=\prod_{i=1}^{n+1}q_i^{\frac{1-x_i}{2}}\cdot(1-q_i)^{\frac{x_i+1}{2}}, \quad (2.7)$$

and in the class Ω_2 - the distribution

$$f_2(\vec{x})=\prod_{i=1}^{n+1}q_i^{\frac{x_i+1}{2}}\cdot(1-q_i)^{\frac{1-x_i}{2}}. \quad (2.8)$$

These relations hold under the assumption that the components $X_1,X_2,\dots,X_n,X_{n+1}$ of the vector $\vec{X}$ are independent of one another.

The mathematical expectation $\mu_i=\text{M}[X_i]$ of each of the components X_i is obtained from the partial distributions of the variables X_i. These $n+1$ mathematical expectations can be represented as the vector $\vec{\mu}$ of means

$$\vec{\mu}=\text{M}[\vec{X}]=(\mu_1,\mu_2,\dots,\mu_n,\mu_{n+1})'.$$

In particular, in the class Ω_1 the centroid of probabilities distribution of the vector $\vec{X}$ is given by the vector

$$\vec{\mu}_1=(\mu_{11},\mu_{12},\dots,\mu_{1n},\mu_{1(n+1)})', \quad (2.9)$$

and in the class Ω_2 - by the vector

$$\vec{\mu}_2=(\mu_{21},\mu_{22},\dots,\mu_{2n},\mu_{2(n+1)})', \quad (2.10)$$

where

$$\left.\begin{aligned} & \mu_{1i}=1-2q_i \\ & \mu_{2i}=2q_i-1 \\ & i=\overline{1,n+1} \end{aligned}\right\}. \qquad (2.11)$$

From the partial distributions for X_i we compute the dispersions σ_i^2 of random variables $X_i\,(i=\overline{1,n+1})$, and from the joint distribution of the components X_i and X_j we compute the covariance σ_{ij} of X_i and X_j:

$$\Sigma_{ij}=\text{M}\left[\left(X_i-\mu_i\right)\left(X_j-\mu_j\right)\right]=\begin{cases}\sigma_i^2, \text{ if } i=j \\ \sigma_{ij}, \text{ if } i\neq j\end{cases}.$$

(2.12)

Note that $\sigma_{ii}=\sigma_i^2$ and $\sigma_{ij}=\sigma_{ji}$. Dispersions and covariances generate, in total, the covariance

ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 2.114

matrix which is a generalization of the notion of dispersion of a one-dimensional random variable:

$$\Sigma_0 = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} & \cdots & \sigma_{1(n+1)} \\ \sigma_{21} & \sigma_2^2 & \sigma_{23} & \cdots & \sigma_{2(n+1)} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \sigma_{(n+1)1} & \sigma_{(n+1)2} & \sigma_{(n+1)3} & \cdots & \sigma_{(n+1)(n+1)}^2 \end{bmatrix}.$$

Since in the considered case the components $X_1, X_2, \ldots, X_n, X_{n+1}$ of the vector $\overrightarrow{X}$ are independent of one another, we have $\sigma_{ij} = 0$ for all $i \neq j$ and Σ_0 is the diagonal matrix

$$\Sigma_0 = \begin{bmatrix} \sigma_1^2 & 0 & 0 \cdots 0 \\ 0 & \sigma_2^2 & 0 \cdots 0 \\ \vdots & \vdots & \vdots \cdots \vdots \\ 0 & 0 & 0 \cdots \sigma_{n+1}^2 \end{bmatrix}.$$

It can be easily verified that the dispersions σ_i^2 are identical in both classes and defined by the relations

$$\left. \begin{array}{c} \sigma_{(1)i}^2 = \sigma_{(2)i}^2 \equiv \sigma_i^2 = 4q_i\left(1-q_i\right) \\ i = \overline{1, n+1} \end{array} \right\}. \quad (2.13)$$

Therefore the covariance matrices are also identical:

$$\Sigma_1 = \Sigma_2 \equiv \Sigma_0. \qquad (2.14)$$

Thus, when investigating the threshold element, the parameters $\vec{\mu}_1$, $\vec{\mu}_2$ and Σ_0 can be assumed to be known, while the principle of its functioning implies that the threshold element computes a linear combination of observations

$$z = a_1x_1 + a_2x_2 + \cdots + a_nx_n + a_{n+1}x_{n+1}$$

called a linear discriminant function. The observations vector $\vec{x}$ is assigned to the class Ω_1 if $z > \Theta$, and to the class Ω_2 if $z < \Theta$. When $z = 0$, no decision is produced.

Let us introduce into consideration a random variable Z by the relation

$$Z = \sum_{i=1}^{n+1} a_i X_i. \qquad (2.15)$$

Or

$$Z_i = \sum_{i=1}^{n+1} Z_i, \qquad (2.16)$$

where

$$Z_i = a_i X_i \,. \qquad (2.17)$$

If the observation $\vec{x}$ arrived from Ω_1, then sum (2.15) has the distribution

$$F_1(z) = \mathop{*}_{i=1}^{n+1} f_{1i}(z_i), \qquad (2.18)$$

where $*$ is the symbol of the operation of convolution, and

$$f_{1i}(z_i) = q_i^{\frac{a_i - z_i}{2a_i}} (1-q_i)^{\frac{z_i + a_i}{2a_i}}, \qquad (2.19)$$

where z_i are equal either to $+a_i$ or to $-a_i$.

The mathematical expectation of the random variable Z is

$$m_1 = \sum_{i=1}^{n+1} a_i \mu_{1i} = \sum_{i=1}^{n+1} a_i\left(1-2q_i\right). \quad (2.20)$$

Analogously, if the observation $\vec{x} = \left(x_1, x_2, \ldots, x_n, x_{n+1}\right)'$ is from the class Ω_2, then the variable Z has the following probabilities distribution

$$F_2\left(z\right) = \mathop{*}_{i=1}^{n+1} f_{2i}\left(z_i\right), \qquad (2.21)$$

where

$$f_{2i}(z_i) = q_i^{\frac{z_i + a_i}{2a_i}} \left(1-q_i\right)^{\frac{a_i - z_i}{2a_i}} \qquad (2.22)$$

and the probable values of z_i are $+a_i$ and $-a_i$.

If the observation $\vec{x}$ has arrived from the class Ω_2, the mathematical expectation of the random variable Z is determined by the formula

$$m_2 = \sum_{i=1}^{n+1} a_i \mu_{2i} = \sum_{i=1}^{n+1} a_i\left(2q_i - 1\right). \quad (2.23)$$

Comparing relations (2.20) and (2.23), we easily conclude that

$$m_1 = -m_2. \qquad (2.24)$$

By the analysis of (2.19) and (2.22) we obtain

$$\left. \begin{array}{c} f_{2i}\left(z_i\right) = f_{1i}\left(-z_i\right) \\ i = \overline{1, n+1} \end{array} \right\}. \qquad (2.25)$$

Therefore

$$F_2(z) = F_1(-z)\,. \qquad (2.26)$$

The dispersion σ_z^2 of the random variable Z is the same in both cases and defined by the formula

$$\sigma_z^2 = \sum_{i=1}^{n+1}\sum_{j=1}^{n+1} a_i \sigma_{ij} a_j, \qquad (2.27)$$

which, in view of relations (2.13), for $\sigma_{ij} = 0$ yields

ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 2.114

$$\sigma_z^2 = \sum_{i=1}^{n+1} 4a_i^2 q_i \left(1-q_i\right). \quad (2.28)$$

For heuristic reasons, the weights $a_1, a_2, \dots, a_n, a_{n+1}$ must be chosen so that the mathematical expectations m_1 and m_2 would stand apart from each other as far as possible and the dispersion σ_z^2 be minimal.

To this end, it suffices to introduce as a target function the generalized Mahalanobis distance **[4]**

$$\rho = \frac{\left(m_1 - m_2\right)^2}{\sigma_z^2} \quad (2.29)$$

which in the setting of our problem has the form

$$\rho = \frac{\left[\sum_{i=1}^{n+1} a_i \left(\mu_{1i} - \mu_{2i}\right)\right]^2}{\sum_{i=1}^{n+1}\sum_{j=1}^{n+1} a_i \sigma_{ij} a_j}. \quad (2.30)$$

The weights $a_i \left(i = \overline{1, n+1}\right)$, which give generate the maximum of this expression. satisfy the following system of equations

$$\left.\begin{aligned} \frac{\partial \rho}{\partial a_i} = 0 \\ i = \overline{1, n+1} \end{aligned}\right\}. \quad (2.31)$$

Using expression (2.30), we obtain

$$\frac{\mu_{1s} - \mu_{2s}}{\sum_{j=1}^{n+1} \sigma_{sj} a_j} = \frac{\sum_{i=1}^{n+1} a_i \left(\mu_{1i} - \mu_{2i}\right)}{\sum_{i=1}^{n+1} a_i \sum_{j=1}^{n+1} \sigma_{ij} a_j}. \quad (2.32)$$

Any vector $\left(a_1, a_2, \dots, a_n, a_{n+1}\right)'$ that satisfies the system of equations

$$\left.\begin{aligned} \sum_{j=1}^{n+1} \sigma_{sj} a_j = k\left(\mu_{1s} - \mu_{2s}\right) \\ s = \overline{1, n+1} \end{aligned}\right\}, \quad (2.33)$$

where k is an arbitrary constant, will be a solution of system (2.32) as well.

Since in system (2.29) $\sigma_{ij} = 0$ for $i \neq j$ and $\sigma_{ii} = \sigma_i^2$, we obtain

$$\left.\begin{aligned} a_s \sigma_s^2 = k\left(\mu_{1s} - \mu_{2s}\right) \\ s = \overline{1, n+1} \end{aligned}\right\},$$

whence

$$\left.\begin{aligned} a_s = k\frac{\mu_{1s} - \mu_{2s}}{\sigma_s^2} \\ s = \overline{1, n+1} \end{aligned}\right\}. \quad (2.34)$$

With expressions (2.11) and (2.13) taken into account, we eventually obtain

$$\left.\begin{aligned} a_i = k\frac{1 - 2q_i}{2q_i\left(1 - q_i\right)} \\ i = \overline{1, n+1} \end{aligned}\right\}. \quad (2.35)$$

If for $q_i < 1/2$ it is desirable to have a positive weight, and for $q_i > 1/2$ a negative one, then the constant k must satisfy the condition $0 < k < \infty$.

For these weights, the maximal value $\rho_{\max}$ of the generalized distance ρ is equal to the difference (more exactly, to the absolute value of the difference) of the mathematical expectations m_1 and m_2 of the sum Z for the classes Ω_1 and Ω_2:

$$\rho_{\max} = \left|m_1 - m_2\right| = \sum_{i=1}^{n+1} \frac{\left(1 - 2q_i\right)^2}{q_i\left(1 - q_i\right)}. \quad (2.36)$$

Consequently, when the weights a_i are chosen by relations (2.35), we have the equality

$$\frac{\sigma_z^2}{\left|m_1 - m_2\right|} = 1. \quad (2.37)$$

From formula (2.36) we see that an increase of the number of $n+1$ inputs of the threshold element, if only the error probabilities of these inputs are not equal to $1/2$, leads to the monotonic growth of the generalized distance ρ and therefore to a decrease of the probability of incorrect restoration of the signal by the decision element.

Criterion of entropy sensitivity

In the decision element, the system of $n+1$ data buses $B_1, B_2, \dots, B_n, B_{n+1}$ can be regarded as some source of binary information with the entropy E defined by the formula

$$E = \sum_{i=1}^{n+1} E_i, \quad (2.38)$$

ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 2.114

where E_i $\left(i=\overline{1,n+1}\right)$ is the entropy of the discrete random variable $X \cdot X_i$, whose distribution is given as a set of probabilities q_i and $\left(1-q_i\right)$ which meet its realizations $x \cdot x_i = -1$ and $x \cdot x_i = +1$, respectively.

Therefore

$$E_i = k \cdot \left[-\left(1-q_i\right) \cdot \ln\left(1-q_i\right) - q_i \cdot \ln q_i\right], \quad (2.39)$$

where k $(0<k<\infty)$ is an arbitrary positive constant, but the entropy may usually be defined also assuming that $k=1$.

Therefore

$$E = k \cdot \sum_{i=1}^{n+1}\left[-\left(1-q_i\right) \cdot \ln\left(1-q_i\right) - q_i \cdot \ln q_i\right]. \quad (2.40)$$

If in two data buses B_i and B_j the identical changes Δq_i and Δq_j $\left(\Delta q_i = \Delta q_j\right)$ of their error probabilities q_i and q_j bring about various changes $\left(\Delta E\right)_i$ and $\left(\Delta E\right)_j$ of the entropy E, then it is natural to assign a greater weight to that data bus which has caused a greater change of the entropy. In other words, the weight a_i must serve as a measure of the entropy change of the information source (the set of binary buses) depending on an increment of the probability q_i:

$$\left.\begin{aligned} a_i &= \frac{\partial E}{\partial q_i} \\ i &= 1, n+1 \end{aligned}\right\}. \quad (2.41)$$

Using expression (2.40) for the entropy E, from (2.41) we obtain

$$a_i = k \cdot \ln \frac{1-q_i}{q_i}. \quad (2.42)$$

2.3 Relationship between the weights computed by the above two criteria

We denote by a_{im} the weights defined by relations (2.35) and providing the maximum for the generalized Mahalanobis distance:

$$k \cdot \frac{1-2q_i}{2q_i\left(1-q_i\right)} = a_{im}. \quad (2.43)$$

The weights computed on the basis of the entropy approach are denoted by a_{ie}:

$$k \cdot \ln \frac{1-q_i}{q_i} = a_{ie}. \quad (2.44)$$

It is easy to verify that

$$a_{im} = \frac{1}{2} \cdot k \cdot \left(\frac{1-q_i}{q_i} - \frac{q_i}{1-q_i}\right). \quad (2.45)$$

On the other hand,

$$\left.\begin{aligned} \frac{1-q_i}{q_i} &= \exp\left(\ln\frac{1-q_i}{q_i}\right) = \exp\left(\frac{1}{k} \cdot a_{ie}\right) \\ \frac{q_i}{1-q_i} &= \exp\left(-\ln\frac{1-q_i}{q_i}\right) = \exp\left(-\frac{1}{k} \cdot a_{ie}\right) \end{aligned}\right\}. \quad (2.46)$$

Using (2.46) in (2.45), we obtain

$$a_{im} = k \cdot sh\left(\frac{1}{k} \cdot a_{ie}\right). \quad (2.47)$$

In particular, if $k=1$, then

$$a_{im} = sh\left(a_{ie}\right). \quad (2.48)$$

Thus if the normalizing factor k is chosen equal to 1, then weights (2.43), which provide the maximum for the Mahalanobis distance, are related to weights (2.44) found by using the entropy approach and the monotone transformation (2.48) and applying the law of hyperbolic sine.

Fig. 5 shows, for comparison, the graphs of relations (2.35) and (2.42) for the case $k=1$.

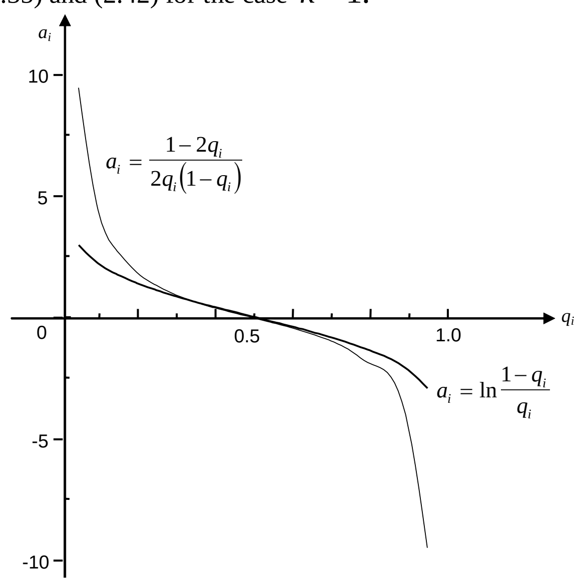

Fig. 5: ***Weight - probability relationship diagram***

ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 2.114

Adaptation of the restoring formal neuron

Adaptation is understood as a process of control of the input weights of a threshold element **[5]** with the purpose of making them match the current error probabilities of these inputs. The objective of the control is to impart more influence to the reliable inputs in determining the output than the less reliable inputs have. Therefore, at each moment of time t the weight a_i $\left(-\infty < a_i < +\infty\right)$ of the ith $\left(i = \overline{1, n+1}\right)$ input of the decision element must be determined by the error probability $q_i(t)$ of this input at the moment of time t **:**

$$a_i = f_a\left(q_i(t)\right).$$

The weights control process is complicated by the fact that we do not have at our disposal sensors of error probabilities $q_i(t)$, which could measure various physical quantities. Only the statistical estimates of these probabilities can be determined by the mismatch of the signal X_i produced by the data bus B_i either with the true variable X supplied by a binary variable for recognition or with the decision Y produced by the threshold element.

In this context, there may be two types of adaptation:: adaptation without feedback (open-loop adaptation) and adaptation with feedback (closed-loop adaptation).

Regardless of whether there is a feedback or not, adaptation is viewed as two procedures: cyclic error-counting adaptation and continuous error-counting adaptation.

The cyclic error-counting adaptation procedure is the one in which the vote-weights are changed periodically, using data collected by computations during a period (cycle) that includes the number of observations at cyclic of time.

In the case of the continuous error-counting adaptation, error probabilities can be estimated by using the devices which correct the vote-weights of all inputs after each observation at a cyclic moment of time.

Hence we distinguish adaptations with cyclic and continuous correction of weights.

Judging by the manner in which a cyclic moment of time is fixed, for performing the correction the third type of adaptation can also be indicated when changes in weights occur at random moments of time after the data buses reach certain states. In particular, the critical state of the data bus B_i can be determined also by the maximum value q_0 of the error probability q_i.

Methods of adaptation without feedback have a limited application because of the necessity to supply a correct answer externally. They have mainly two areas of application:

- The initial adaptation of a device;
- Periodic adjustments made to weights during the operation of the decision element by using checking programs with the known answers.

Continuous adaptation by the Widrow - Hoff least-mean-squares algorithm

Preliminaries

It is absolutely clear that it is in principle possible to organize the adaptation procedure of a threshold element in such a manner that in estimating its input error probabilities the weight these inputs are readjusted not at the end of a cycle by the results of M comparisons, but permanently, upon the completion of each comparison. An algorithm of this adaptation may, for example, be a procedure of penalty and bonus type. In that case, at the $(k+1)$th step the algorithm makes changes in the weight vector $\vec{a}(k)$ of the preceding step differently, depending on whether at the kth step of adaptation the pattern was classified by means of the vector $\vec{a}(k)$ correctly or incorrectly . In particular, if the pattern is classified correctly, then the bonus may consist in that no changes are made to the vector of weights, whereas if the pattern is classified incorrectly, then the threshold element is penalized either by an increase or a decrease of the vector of weights. A conclusion on the correct or incorrect recognition of the signal X by the data buses $B_1, B_2, \cdots, B_n, B_{n+1}$ can be made both by comparing the input signals $X_i\left(i = \overline{1, n+1}\right)$ with an externally supplied correct answer and also by comparing them with the decision Y. However, sometimes, instead of comparing X or Y with each individual signal X_i, comparison can be made with the result of the collective interaction of all $X_i\left(i = \overline{1, n+1}\right)$, i.e. with the output signal produced by the threshold element summer.

Continuous open-loop adaptation (without feedback)

Let us assume that the signal X is a random sequence of digits $+1$ and -1 corresponding to the

ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 2.114

a priori probabilities $q_{n+1} = P(\Omega_1)$ and $1 - q_{n+1} = P(\Omega_2)$. Also, in the threshold element inputs there appears the observations vector $\vec{x} = (x_1, x_2, \cdots, x_n, x_{n+1})'$ according to the probabilistic laws

$$P(\vec{x}/\Omega_1) \equiv f_1(\vec{x}) = \prod_{i=1}^{n+1} q_i^{\frac{1-x_i}{2}} \cdot (1-q_i)^{\frac{x_i+1}{2}},$$

(4.1)

$$P(\vec{x}/\Omega_2) \equiv f_2(\vec{x}) = \prod_{i=1}^{n+1} q_i^{\frac{x_i+1}{2}} \cdot (1-q_i)^{\frac{1-x_i}{2}},$$

(4.2)

where $x_{n+1} \equiv -1$.

For each observations vector $\vec{x}$, let us introduce a random variable of classification, or a mark ξ, such that

$\xi = +m_0$

for $\vec{x}$ corresponding to the signal $X = +1(\text{class } \Omega_1)$, and

$\xi = -m_0$

for $\vec{x}$ corresponding to the signal $X = -1(\text{class } \Omega_2)$.

For this, it suffices to construct ξ by the relation

$$\xi = m_0 \cdot X. \qquad (4.3)$$

Then, in the process of adaptation, data are represented as a sequence of pairs

$(\vec{x}(1), \xi_1), (\vec{x}(2), \xi_2), \cdots, (\vec{x}(k), \xi_k), \cdots.$

In view of formulas (4.1) and (4.2), the Bayes separating function has the form

$$\Upsilon_0(\vec{x}) = \ln \frac{P(\vec{x}/\Omega_1)}{P(\vec{x}/\Omega_2)} = \sum_{i=1}^{n+1} x_i \cdot \ln \frac{1-q_i}{q_i}. \qquad (4.4)$$

The purpose of adaptation consists exactly in, firstly, approximating $\Upsilon_0(\vec{x})$ with unknown parameters $q_i \ (i = \overline{1, n+1})$ by means of the finite series

$$\Upsilon_0(\vec{x}) = \sum_{i=1}^{n+1} x_i \cdot a_i \qquad (4.5)$$

and, secondly, determining the weight vector $\vec{a} = (\hat{a}_1, \hat{a}_2, \cdots, \hat{a}_n, \hat{a}_{n+1})'$ that minimizes the root-mean-square error of approximation

$$\varepsilon_0^2 = M\left[\left(\sum_{i=1}^{n+1} a_i x_i - \Upsilon_o(\vec{x})\right)^2\right]. \qquad (4.6)$$

For the minimization of ε_0^2 it suffices to know the Bayes separating function $\Upsilon_0(\vec{x})$ defined by formula (4.4) with exactly unknown probabilities $q_i \ (i = \overline{1, n+1})$.

To avoid this difficulty, we will consider ξ as a noise-corrupted value of the function $\Upsilon_0(\vec{x})$. Then the weights vector $\hat{\vec{a}} = (\hat{a}_1, \hat{a}_2, \cdots, \hat{a}_n, \hat{a}_{n+1})'$, which minimizes ε_0^2, will also minimize the criterion function

$$J(\vec{x}, \vec{a}) = M\left[\left(\sum_{i=1}^{n+1} a_i x_i - \xi\right)^2\right]. \qquad (4.7)$$

Taking the partial derivatives of the criterion function with respect to weights a_i, we obtain

$$\left. G_{0i}(\vec{x}, \vec{a}) = \frac{\partial J(\vec{x}, \vec{a})}{\partial a_i} = M\left[2\left(\sum_{j=1}^{n+1} a_j x_j - \xi\right) \cdot x_i\right] \right\} .$$
$$i = \overline{1, n=1}$$

Assuming that instead of the real values of $G_{0i}(\vec{x}, \vec{a})$, we observe their noise-corrupted values

$$\left. h_{0i}(\vec{x}, \vec{a}) = 2\left(\sum_{j=1}^{n+1} a_j x_j - \xi\right) \cdot x_i \right\}, \qquad (4.8)$$
$$i = \overline{1, n+1}$$

such that

$$\left. M[h_{0i}(\vec{x}, \vec{a})] = G_{0i}(\vec{x}, \vec{a}) \right\}$$
$$i = \overline{1, n+1}$$

and

$$\left. \sigma_i^2(\vec{x}, \vec{a}) = M\left[\left(G_{0i}(\vec{x}, \vec{a}) - h_{0i}(\vec{x}, \vec{a})\right)^2\right] < L \right\}$$
$$i = \overline{1, n+1}$$

for all weights $a_i \ (i = \overline{1, n+1})$, where $L < \infty$ is a positive constant, we may apply the Robbins - Monro algorithm **[6]** for the iterative definition of the zero $\hat{a}_i$ of the function $G_{0i}(\vec{x}, \vec{a}), i = \overline{1, n+1}$.

Denoting by $a_i(1)$ an arbitrary initial estimate of the root of the equation

$G_{0i}(\vec{x}, \vec{a}) = 0,$

ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 2.114

and by $a_i(k)$ the estimate of this root obtained at the kth iteration step, the procedure of correction by means of the Robbins - Monro algorithm can be written in the form

$$\left.\begin{aligned} a_i(k+1) = a_i(k) - \beta_k \cdot h_{0i}\left(\vec{x}(k), \vec{a}(k)\right) \\ i = \overline{1, n+1} \end{aligned}\right\}, \quad (4.9)$$

where β_k is an element of the sequence of positive numbers which satisfy the conditions

$$\left.\begin{aligned} \lim_{k\to\infty} \beta_k = 0 \\ \sum_{k=1}^{\infty} \beta_k = \infty \\ \sum_{k=1}^{\infty} \beta_k^2 < \infty \end{aligned}\right\}. \qquad (4.9')$$

Such a sequence can be exemplified by a harmonic sequence.

Therefore error corrections by the Robbins-Monro algorithm are proportional to the value of the variable $h_{0i}\left(\vec{x}(k), \vec{a}(k)\right)$ in the preceding observation.

The substitution of (4.8) into the general expression (4.9) gives

$$\left.\begin{aligned} a_i(k+1) = a_i(k) + \rho_k \cdot x_i(k) \cdot \left[\xi - \sum_{j=1}^{n+1} a_j(k) x_j(k)\right] \\ i = \overline{1, n+1} \end{aligned}\right\}, (4.10)$$

where $\rho_k = 2 \cdot \beta_k$.

Relation (4.10) is in essence the correction algorithm which Widrow and Hoff used in their work **[7].** In our case, for a stochastic increment of the ith weight

$$\Delta a_i(k) = a_i(k+1) - a_i(k)$$

that takes place at the $(k+1)$th iteration step, we write (4.10) in the following form

$$\left.\begin{aligned} \Delta a_i(k) = \rho_k \cdot X_i(k) \cdot \left[m_0 X - \sum_{j=1}^{n+1} a_j(k) X_j(k)\right] \\ i = \overline{1, n+1} \end{aligned}\right\}. (4.11)$$

Our further investigation pursues the following aims:

- finding a mathematical expectation for the increment value (4.11);
- obtaining expressions for weights which correspond to the steady state when the mathematical expectations of (4.11) are equal to zero.

Let us assume that the input errors of the threshold element are independent of one another and write (4.11) in the form

$$\left.\begin{aligned} \Delta a_i(k) = \rho_k m_0 X X_i(k) - \rho_k X_i(k) \cdot \sum_{j=1}^{n+1} a_j(k) X_j(k) \\ i = \overline{1, n+1} \end{aligned}\right\}.$$

Here

$$\sum_{j=1}^{n+1} a_j(k) X_j(k) = \sum_{\substack{j=1 \\ j\neq i}}^{n+1} a_j(k) X_j(k) + a_i(k) X_i(k),$$

where the condition $j \neq i$ in the subscript of the summation symbol means that the sum does not contain a summand with index i.

Thus

$$\left.\begin{aligned} \Delta a_i(k) = m_0 \rho_k X X_i(k) - \\ + \rho_k \cdot X_i(k) \cdot \sum_{\substack{j=1 \\ j\neq i}}^{n+1} a_j(k) X_j(k) - \\ + \rho_k a_i(k) X_i^2(k) \\ i = \overline{1, n+1} \end{aligned}\right\}. \qquad (4.12)$$

For the mathematical expectations of individual summands in the right-hand part of relation (4.12) we have

$$\begin{aligned} M\left[m_0 \rho_k X X_i(k)\right] = \\ = m_0 \rho_k \left[(1-q_i) \cdot 1 + q_i \cdot (-1)\right] = \qquad (4.13) \\ = m_0 \rho_k (1 - 2q_i), \end{aligned}$$

$$M\left[\rho_k a_i(k) \cdot X_i^2(k)\right] = \rho_k a_i(k). \quad (4.14)$$

As to the mathematical expectation of the summand

$$\rho_k \cdot X_i(k) \cdot \sum_{\substack{j=1 \\ j\neq i}}^{n+1} a_j(k) X_j(k),$$

it should be calculated using two conditions with $X = +1$ and $X = -1$.

We have

$$M\left[\rho_k X_i(k) \cdot \sum_{\substack{j=1 \\ j\neq i}}^{n+1} a_j(k) X_j(k) \Big/ X = +1\right] =$$

$$= \rho_k \cdot (1 - 2q_i) \cdot \sum_{\substack{j=1 \\ j\neq i}}^{n+1} a_j(k) (1 - 2q_j),$$

(4.15)

ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 2.114

$$M\left[\rho_k X_i(k)\cdot\sum_{\substack{j=1\\ j\neq i}}^{n+1} a_j(k) X_j(k) \Big/ X=-1\right]=$$
$$=\rho_k\cdot(2q_i-1)\cdot\sum_{\substack{j=1\\ j\neq i}}^{n+1} a_j(k)(2q_j-1)= \qquad (4.16)$$
$$=\rho_k\cdot(1-2q_i)\cdot\sum_{\substack{j=1\\ j\neq i}}^{n+1} a_j(k)(1-2q_j).$$

Therefore expressions (4.15) and (4.16) coincide and it can be claimed that independently of the class supplied for recognition,

$$M\left[\rho_k X_i(k)\cdot\sum_{\substack{j=1\\ j\neq i}}^{n+1} a_j(k) X_j(k)\right]= \qquad (4.17)$$
$$=\rho_k\cdot(1-2q_i)\cdot\sum_{\substack{j=1\\ j\neq i}}^{n+1} a_j(k)(1-2q_j).$$

Taking into account (4.13), (4.14) and (4.17) we have

$$\left.\begin{array}{l} M\left[\Delta a_i(k)\right]=\rho_k\left\{(1-2q_i)\cdot\Psi-a_i(k)\right\} \\ \Psi=\left[m_0-\sum_{\substack{j=1\\ j\neq i}}^{n+1} a_j(k)(1-2q_j)\right] \\ i=\overline{1,n+1} \end{array}\right\}. \quad (4.18)$$

Since here

$$\sum_{\substack{j=1\\ j\neq i}}^{n+1} a_j(k)(1-2q_j)=$$
$$=\sum_{j=1}^{n+1} a_j(k)(1-2q_j)-a_i(k)(1-2q_i),$$

using this fact in (4.18), we come to the expression

$$\left.\begin{array}{l} M\left[\Delta a_i(k)\right]= \\ =\rho_k\left\{(1-2q_i)\cdot\Upsilon+a_i(k)(1-2q_i)^2-a_i(k)\right\} \\ \Upsilon=\left[m_0-\sum_{j=1}^{n+1} a_j(k)(1-2q_j)\right] \\ i=\overline{1,n+1} \end{array}\right\}.$$

It is easy to see that

$$a_i(k)\cdot(1-2q_i)^2-a_i(k)=-4a_i(k)\cdot q_i\cdot(1-q_i).$$

Therefore

$$\left.\begin{array}{l} M\left[\Delta a_i(k)\right]=\rho_k\left\{(1-2q_i)\cdot\Upsilon+\lambda\right\} \\ \lambda=-4a_i(k)q_i(1-2q_i) \\ \Upsilon=\left[m_0-\sum_{j=1}^{n+1} a_j(k)(1-2q_j)\right] \\ i=\overline{1,n+1} \end{array}\right\}. \qquad (4.19)$$

Since the Robbins-Monro algorithm converges when conditions (4.9') are fulfilled, the weights will stop being changed (on the average) for some finite k. In this steady state the following equalities are fulfilled:

$$\left.\begin{array}{l} M\left[\Delta a_i(k)\right]=0 \\ i=\overline{1,n+1} \end{array}\right\}. \qquad (4.20)$$

We denote by $\hat{a}_i$ .the weights which have reached the steady state. Then for $\hat{a}_i$ $\left(i=\overline{1,n+1}\right)$ we have the equations

$$\left.\begin{array}{l} \hat{a}_i=\dfrac{1-2q_i}{2q_i(1-q_i)}\cdot\dfrac{m_0-\sum_{j=1}^{n+1}\hat{a}_j(1-2q_j)}{2} \\ i=\overline{1,n+1} \end{array}\right\}.$$

(4.21)

It is easy to see that the variables

$$\left.\begin{array}{l} a_{im}=\dfrac{1-2q_i}{2q_i(1-q_i)} \\ i=\overline{1,n+1} \end{array}\right\} \qquad (4.22)$$

are the weights which give the maximum to the Mahalanobis distance between the sets of values of the random sum

$$Z=\sum_{j=1}^{n+1} a_j X_j$$

in the classes Ω_1 and Ω_2.

To the weights

$$\left.\begin{array}{l} a_{ie}=\ln\dfrac{1-q_i}{q_i} \\ i=\overline{1,n+1} \end{array}\right\} \qquad (4.23)$$

computed in terms of entropic sensitivity and *being optimal in the Bayes sense* (i.e. in the sense of the a posteriori probability maximum), weights (4.22) are related in the following manner

$$\left.\begin{array}{l} a_{im}=sh(a_{ie}) \\ i=\overline{1,n+1} \end{array}\right\}. \qquad (4.24)$$

ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 2.114

Now, introducing the constant γ depending on the value of a steady-state weight, we obtain

$$\left.\begin{array}{c}\hat{a}_i = \gamma \cdot a_{im} = \gamma \cdot sh\left(a_{ie}\right) \\ i = \overline{1,\, n+1}\end{array}\right\}, \qquad (4.25)$$

where

$$\gamma = \frac{m_0 - \sum\limits_{j=1}^{n+1} \hat{a}_j \left(1-2q_j\right)}{2}. \qquad (4.26)$$

The expression for γ can be written in a more compact form if we take into account the fact that the maximal value $\rho_{\max}$ of the Mahalanobis distance ρ is defined by the relation

$$\rho_{\max} = \sum_{j=1}^{n+1} \frac{\left(1-2q_j\right)^2}{q_j\left(1-q_j\right)} = 2 \cdot \sum_{j=1}^{n+1} a_{jm}\left(1-2q_j\right). \qquad (4.27)$$

Indeed, in view of expressions (4.25) and (4.27) we can write formula (4.26) in the form

$$\gamma = \frac{m_0 - \gamma \cdot \sum\limits_{j=1}^{n+1} a_{jm}\left(1-2q_j\right)}{2} = \frac{m_0 - \frac{1}{2}\cdot\gamma\cdot\rho_{\max}}{2}.$$

Hence

$$\gamma = \frac{2m_0}{4+\rho_{\max}}. \qquad (4.28)$$

Substituting (4.28) into (4.25), we obtain

$$\left.\begin{array}{c}\hat{a}_i = \dfrac{2m_0}{4+\rho_{\max}} \cdot a_{im} = \dfrac{2m_0}{4+\rho_{\max}} sh\left(a_{ie}\right) \\ i = \overline{1,\, n+1}\end{array}\right\}. \qquad (4.29)$$

Thus, as a result of continuous adaptation without feedback the weights are established, which are proportional to the weights which give the maximum to the Mahalanobis distance.

Continuous closed-loop adaptation (with feedback)

In the case of continuous adaptation with feedback, when weight increments are determined by the Widrow-Hoff algorithm, the mark ξ is computed as $m_0 \cdot Y$, where Y is the output decision of the threshold element and, instead of the error probability q_i of the ith input, we have to operate with the probability d_i that this input will mismatch the variable Y.

Fig. 6 shows the flow-diagram of realization of the methods of continuous adaptation (both without and with feedback) by the Widrow-Hoff algorithm.

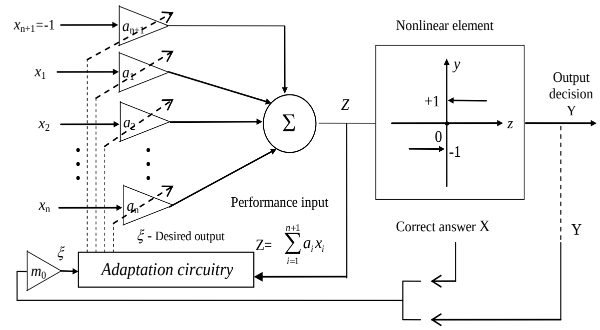

Fig. 6: ***Flow-diagram for closed- and open-loop adaptation by the Widrow-Hoff algorithm***

Results and discussion

a) The weights a_{ie} computed by the entropy sensitivity criterion are defined by formula (2.44):

$$a_{ie} = k \cdot \ln\frac{1-q_i}{q_i} \quad \left(i = \overline{1, n+1}\right),$$

while the weights a_{im} computed by the criterion of Mahalanobis distance maximum are defined by formula (2.43):

$$a_{im} = k \cdot \frac{1-2q_i}{2q_i\left(1-q_i\right)} \quad \left(i = \overline{1, n+1}\right).$$

b) The relationship between a_{ie} and a_{im} is established by means of the monotone transformation (2.47):

$$a_{im} = k \cdot sh\left(\frac{1}{k}\cdot a_{ie}\right) \quad \left(i = \overline{1, n+1}\right).$$

c) A maximal value of the Mahalanobis distance is defined by expression (4.27):

$$\rho_{\max} = \sum_{j=1}^{n+1} \frac{\left(1-2q_j\right)^2}{q_j\left(1-q_j\right)}.$$

d) As a result of continuous adaptation, weights (4.29) are established, which are proportional to the weights which give the maximum to the Mahanobis distance:

$$\left.\begin{array}{c}\hat{a}_i = \dfrac{2m_0}{4+\rho_{\max}} \cdot a_{im} = \dfrac{2m_0}{4+\rho_{\max}} sh\left(a_{ie}\right) \\ i = \overline{1,\, n+1}\end{array}\right\}.$$

We can state that the method of continuous adaptation is qualitatively very close to an optimal method defined by the weights a_{ie} computed by the criterion of entropy sensitivity since the weight a_{im} in the above formula computed by the criterion of

Mahalanobis distance maximum is proportional to $sh(a_{ie})$ and monotonically (but not in direct proportion) depends on a_{ie}.

CONCLUSION

It is shown that this adaptation procedure has an equilibrium setting of vote-weights proportional to the hyperbolic sine of the entropic vote-weights. This adaptation method may be said to be optimum in the sense of the P.C. Mahalanobis's criterion and qualitatively suboptimum from the standpoint of the entropic sensitivity definition since shX is monotonic related to X.

The method would be most valuable when the inputs either operate correctly or have error probability of one-half. As correctly noticed by W.H. Pierce **[5]**, this method may require less equipment to implement than other incremental procedures, because of the feature that the magnitudes of the vote-weights are changed by equal amounts.

ACKNOWLEDGEMENTS

This study was supported by the special Grant-in-Aid for Research of Georgian Technical University, Tbilisi, Georgia. The authors express their thanks to the management of Georgian Technical University, and to Prof. Zurab Tsveraidze, Dr.Sci. (Tech.), the Dean of the Faculty of Informatics and Control Systems.

The authors also express their gratitude to Dr. Tinatin Kaishauri and Dr. Levan Imnaishvili for the helpful advice and suggestions, and to Leila Kobelashvili for rendering assistance in preparing the English version of the paper.

REFERENCES

[1] J. von Neumann, Probabilistic logics and the synthesis of reliable organisms from unreliable components, In: C. Shannon and J. McCarthy (Editors), Automata Studies, Princeton University Press, 1956, 43-98.

[2] V.I. Varshavsky, Functionality and Synthesis of Threshold. Elements (in Russian), Proceedings of the USSR Academy of Sciences **139**(5) (1961) 1071-1074.

[3] M.L. Dertouzos, Threshold Logic: a Synthesis Approach, The MIT Press (M.I.T. Research Monograph Series), Cambridge, 1965, 256 p.

[4] P.C. Mahalanobis, On the Generalized Distance in Statistics, Proceedings of the National Institute of Sciences of India, 12 (1936), 49-55.

[5] W.H. Pirce, Failure-Tolerant Computer Design, Academic Press, New York and London, 1965, 242 pp.

[6] H. Robbins, and S. Monro, A Stochastic Approximation Method, The Annals of Mathematical Statistics, **22**(3) (1951), 400-407.

[7] B. Widrow, and M.E. Hoff, Jr., Adaptive Switching Circuits, In: IRE WESCON Convention Record, 1960, New York: IRE, Part 4, 96-104.

[8] J.G. Gogiashvili, O.M. Namicheishvili, and G.G. Chonia, Optimization of Weights for Threshold Redundancy of Binary Channels by the Method of (Mahalanobis's) Generalized Distance, In: Abstracts' Book of the Second International Conference on Mathematical Methods in Reliability MMR'2000: Methodology, Practice and Interference, Bordeaux, France, 2000, v. 1, 463-466

[9] J.G. Gogiashvili, O.M. Namicheishvili, and G.G. Shonia, Cyclic Adaptation of Threshold Elements, Bulletin of the Georgian Academy of Sciences **170**(3) (2004), 485-487.

[10] A.M. Elizbarashvili, and O.M. Namicheishvili, Mathematical Theory of Redundancy Based on Formal Neuron Model, In: Proceedings of the European Computing Conference ECC'09, Tbilisi, Georgia, 2009, 381-385.

[11] A.I. Prangishvili, O.M. Namicheishvili, Threshold Redundancy of Binary Sensors, In: Proceedings of the NATO Advanced Research Workshop on Philosophy and Synergy of Information: Sustainability and Security, Tbilisi, Georgia, 2011, IOS Press, Amsterdam, 2012, 88-99.

[12] A.I. Prangishvili, O.M. Namicheishvili, Formal Neuron and the Models of its Adaptation in the Redundancy Theory, In: Proceedings of the International Scientific Conference ICTM-2010, Devoted to the 80th Anniversary of I.V. Prangishvili, Series: Computer Science, Technology and Applications, Nova Science Publishers, New York, 2012, 1-9.

[13] A.I. Prangishvili, O.M. Namicheishvili, and M.A. Gogiashvili Two Methods of Obtaining a Minimal Upper Estimate for the Error Probability of the Restoring Formal Neuron, In: Proceedings of the International Conference on Mathematical Methods, Mathematical Models and Simulation in Science and Engineering MMSSE 2014, Interlaken,

ISSN: 2277-9655
Scientific Journal Impact Factor: 3.449
(ISRA), Impact Factor: 2.114

Switzerland, 2014, Mathematics and Computers in Science and Engineering Series 23, 37-41.

[14] A.I. Prangishvili, O.M. Namicheishvili, and M.A. Gogiashvili Two Methods of Obtaining a Minimal Upper Estimate for the Error Probability of the Restoring Formal Neuron, International Journal of Engineering and Technical Research (IJETR), **2**(2) (2014), 64-68:http://www.erpublication.org/IJETR/-vol_issue.php?abc1=20

[15] G. Levitin, Maximizing survivability of vulnerable weighted voting system, Reliability Engineering & System Safety, **83**(1) (2004) 17-26.

[16] G. Levitin, and A. Lisnianski, Reliability optimization for weighted voting system, Reliability Engineering & System Safety, **71**(2) (2001) 131-138.

AUTHOR BIBLOGRAPHY

	**Archil Prangishvili** Academician of the National Academy of Sciences of Georgia, Rector of Georgian Technical University, Professor, Dr.Sc.(Tech.). Research field: Control Systems and Information Technology.
	**Natia Namicheishvili** is with the Informatics and Control Systems Faculty, Department of computer engineering, Georgian Technical University, Tbilisi, Georgia 0175.