УДК 681.5

# ADAPTATION AND LEARNING IN SOME CLASSES OF IDENTIFICATION AND CONTROL SYSTEMS

**L.S. Zhiteckii, S.A. Nikolaienko, K.Yu. Solovchuk**

***International Research and Training Center for Information Technologies and Systems of the National Academy of Science of Ukraine and Ministry of Education and Sciences of Ukraine, Kiev, Ukraine***

Рассмотрены задача обучения нейросетевых моделей для идентификации неопределенных нелинейных систем в стохастической среде и задача адаптивного управления линейным многомерным объектом без памяти с произвольным матричным коэффициентом усиления при наличии нерегулярных ограниченных возмущений, уровни которых априори неизвестны. Установлены достаточные условия сходимости алгоритма обучения нейронных сетей и асимптотические свойства алгоритма адаптивного управления многомерным объектом. Полученные результаты являются фундаментальными.

***Ключевые слова:*** нейронная сеть, градиентный алгоритм обучения, сходимость, многомерный объект без памяти, алгоритм адаптивного управления, ограниченность сигналов.

Розглянуто задачу навчання нейромережних моделей для ідентифікації невизначених нелінійних систем у стохастичному середовищі та задачу адаптивного керування лінійним багатовимірним об'єктом без пам'яті з довільним матричним коефіцієнтом підсилення за наявності нерегулярних обмежених збурень, рівні яких апріорі невідомі. Встановлено достатні умови збіжності алгоритму навчання нейронних мереж, а також асимптотичні властивості алгоритму адаптивного керування багатовимірним об'єктом. Одержані результати є фундаментальними.

***Ключові слова:*** нейронна мережа, градієнтний алгоритм навчання, збіжність, багатовимірній об'єкт без пам'яті, алгоритм адаптивного керування, обмеженість сигналів.

## INTRODUCTION

Since the appearance of the fundamental works [1, 2], substantial progress has been achieved during past decades in the area of identification and learning automatic systems. This research direction remains actual up to now because of its importance from both theoretical and practical points of view. In last time, new results for designing adaptive identification and control systems have been derived by the Ukrainian researchers including one of the authors and summarized in the books [3, 4].

Over the past years, interest has been increasing toward the use of multilayer neural networks as adjustable models for the adaptive identification of nonlinearly parameterized dynamic systems [5–8]. Several learning methods for updating the weights of neural networks have been advanced in literature. Most of these methods rely on the gradient concept [8]. Although this concept has been successfully used in many empirical studies, there are very few fundamental results dealing with the convergence of gradient algorithms for learning neural networks. One of these results is based on utilizing the Lyapunov stability theory [6, 9].

The asymptotic behaviour of online adaptive gradient algorithms for the network learning has been studied by many authors. In particular, White [10] investigated the convergence of the learning process for the so-called feedforward network models with single hidden layer by using the stochastic approximation theory. The convergence results have been derived in [11–17] among many others provided that input signals have a probabilistic nature. In their stochastic approach, the learning rate goes to zero as the learning process tends to infinity. Unfortunately, this gives that the learning goes faster in the beginning and slows down in the late stage.

The convergence analysis of learning algorithm with deterministic (non-stochastic) nature has been given in [18–23]. In contrast to the stochastic approach, several of these results allow employing a constant learning rate [20, 24]. However, they assume that learning set must be finite whereas in online identification schemes, this set is theoretically infinite. To the best of author's knowledge, there are no general results in literature concerning the global convergence properties of training procedures with a fixed learning rate applicable to the case of infinite learning set.

The distinguishing feature of multi-layer neural networks is that they describe some nonlinearly parameterized models needed to be identified. This leads to difficulties in deriving their convergence properties for a general case. To avoid these difficulties in non-stochastic case, the assumption that similar nonlinear functions need to be convex (concave) is introduced in [25]. However, such an assumption is not appropriate for neural network's description of nonlinearity.

A popular approach to analyze the asymptotic behavior of online gradient algorithms in stochastic case is based on the Martingale convergence theory [26]. This approach has been exploited in [27] to derive some local convergence in stochastic framework for standard online gradient algorithms with the constant learning rate used for updating the parameters (weights) of neural networks models.

It is well known that the design of an adaptive or of other advanced control systems requires a model that gives an accurate description of the plant to be controlled. Within the so-called generalized inverse model-based approach, a new perfect controller has recently been devised in [28] to stabilize an arbitrary multi-input multi-output memoryless plant whose are assumed to be known. To cope with an uncertain multivariable plant having the nonsingular matrix gain, the standard adaptive control algorithms can directly be exploited in adjusting its inverse model [29, item 4.2.3°]. See also [4, subsect. 4.2]. However, they are quite not admissible if this matrix gain becomes singular as it is noted in the textbook [30, item 5.2.3].

Recently, a new adaptive control method to dealing with the possibly singular matrix gain of the linear multivariable plant in the presence of non-stochastic upper bounded disturbance has been advanced in [31].

This paper is an extension of recent results of [27, 31] related to the adaptation and learning algorithms in certain classes of the identification and control systems using the gradient concept. Specifically, the purpose of the paper is to establish the global convergence conditions of the standard gradient online learning algorithm in

the two-layer neural network model by utilizing the probabilistic asymptotic analysis. Also, it is required to derive the convergent adaptive control algorithm guaranteeing the boundedness of the signals in the closed-loop system which contains the multivariable memoryless plant with an arbitrary matrix gain in the presence of unmeasurable disturbances whose bounds are unknown.

## STATEMENT OF THE PROBLEMS

Two different but close problems related to the adaptation and learning algorithms above mentioned are stated and solved.

**The problem of analyzing the asymptotic properties of the online gradient learning algorithm in the neural network model**. Let

$$y(n) = F(x(n)) \tag{1}$$

be the nonlinear equation describing a complex system to be identified. In this equation, $y(n) \in \mathrm{IR}$ and $x(n) \in \mathrm{IR}^N$ are the scalar output and the so-called state vector, respectively, available for the measurement at each $n$th time instant $(n = 1, 2, \ldots)$, and $F: \mathrm{IR}^N \to \mathrm{IR}$ represents some unknown nonlinear mapping. (Note that $x(n)$ may include the current inputs of this system and possibly its past inputs and also outputs; see [1, subsect. 5.15].) Without loss of generality, one supposes that the nonlinearity

$$y = F(x) \tag{2}$$

is the continuous and smooth function on a bounded but infinite set $X \subset \mathrm{IR}^N$ $(\mathrm{diam}\, X < \infty)$.

To approximate (2) by a suitable nonlinearly parameterized function, the two-layer neural network model containing $M$ $(M \geq 1)$ neurons in its hidden layer is employed. The inputs to the each $j$th neuron of this layer at the time instant $n$ are the components of $x(n)$. Its output signal at the $n$th time instant is specified as

$$y_j^{(1)}(n) = \sigma\left( b_j^{(1)} + \sum_{i=1}^{N} w_{ij}^{(1)} x_i(n) \right), \quad j = 1, \ldots, M, \tag{3}$$

where $x_i(n)$ denotes the $i$th component of $x(n)$, and $w_{ij}^{(1)}$ and $b_j^{(1)}$ are the weight coefficients and the bias of this $j$th neuron, respectively. $\sigma(\cdot)$ denotes the so-called activation function defined usually as the sigmoid functions

$$\sigma(s) = \frac{1}{1 + \exp(-s)} \tag{4}$$

or

$$\sigma(s) = \tanh(s). \tag{5}$$

There is only one neuron in the output (second) layer, whose inputs are the outputs of the hidden layer's neurons. The output signal of second layer, $y^{(2)}(n)$, at the time instant $n$ is determined by

$$y^{(2)}(n) = \sum_{j=1}^{M} w_j^{(2)} y_j^{(1)}(n) + b^{(2)}, \tag{6}$$

where $w_1^{(2)}, \ldots, w_M^{(2)}$ are the weights of this neuron and $b^{(2)}$ is its bias.

Since $\sigma(\cdot)$s defined by (4) and (5) are nonlinear, it follows from (3), (6) that $y^{(2)}(n)$ is the nonlinear function depending on $x(n-1)$ and also on the $(M(N+2)+1)$-dimensional parameter vector

$$\begin{aligned} w = [&w_{11}^{(1)}, \ldots, w_{N1}^{(1)}, b_1^{(1)}, \ldots, w_{1M}^{(1)}, \ldots, w_{NM}^{(1)}, b_M^{(1)} : \\ &w_1^{(2)}, \ldots, w_M^{(2)}, b^{(2)}]^T. \end{aligned} \tag{7}$$

To emphasize this fact, define the output signal of the neural network in the form

$$y^{(2)}(n) = \mathrm{NN}(x(n), w) \tag{8}$$

using the notation $\mathrm{NN}: \mathrm{IR}^N \times \mathrm{IR}^{M(N+2)+1} \to \mathrm{IR}$. Taking into account that the neural network plays the role of a model of (1), rewrite (8) as follows:

$$y_{\mathrm{mod}}(n) = \mathrm{NN}(x(n), w). \tag{9}$$

Now, define the variable

$$e = F(x) - \mathrm{NN}(x, w) \tag{10}$$

representing the discrepancy between the nonlinearity (2) and its neural network's model for a fixed $w$. Due to (1), it yields the current model error

$$e(n) = y(n) - \mathrm{NN}(x(n), w) \tag{11}$$

which can be measured at the $n$th time instant. Further, introduce the usual quadratic loss function

$$Q(x, w) = [F(x) - \mathrm{NN}(x, w)]^2. \tag{12}$$

To do an adaptation of the neural network model to the uncertain system (1), the standard online gradient learning algorithm

$$w(n) = w(n-1) - \eta(n) \nabla_w Q(x(n), w(n-1)) \tag{13}$$

taken, for example, from [1, 8] is utilized. In this algorithm, $\nabla_w Q(x(n), w(n-1))$ denotes the gradient of $Q(x, w)$ with respect to $w$ at $w = w(n-1)$ for given $x = x(n)$, and $\eta(n)$ is the learning rate (step size) of (13). Thus, (3), (6), (8) and (13) together with (9) and (12) describe the learning system necessary for the adaptive identification of (1). Suppose $\{x(n)\}$ is a sequence of vectors appearing randomly in accordance with some probability density function $p(x)$ such that

$$\int_X p(x)\,dx = 1.$$

Furthermore, $p(x)$ has the following properties:

$$P\{x(n) \in X'\} := \int_{X'} p(x)\,dx > 0$$

for any subset $X' \subset X$ whose dimension is $N$, and

$$P\{x(n) \in X''\} := \int_{X''} p(x)\,dx = 0$$

if $\dim X'' < N$, where $P\{\cdot\}$ denotes the probability of corresponding random event.

Additionally, it is assumed that $p(x)$ represents a continuous function which may become zero only at some isolated points on $X$.

Now, introduce the performance index

$$J(w) = E\{Q(x, w)\} \tag{14}$$

which evaluates the quality of learning process with $Q(x, w)$ given in (12). In this expression,

$$E\{Q(x, w)\} := \int_X [F(x) - \mathrm{NN}(x, w)]^2\, p(x)\,dx$$

denotes the expectation of $Q(x, w)$ with respect to the random $x$s.

The following problem is here stated. It is required to derive the conditions under which $\{w(n)\}$ caused by the learning algorithm (13) will converge in the sense that

$$J(w(n)) \to \inf_w J(w) \quad \text{as } n \to \infty \tag{15}$$

almost sure (a.s.) for any initial $w(0)$, where $J(w(n))$ is determined by (14) for $w = w(n)$.

**The problem of designing the adaptive control system which contains the linear memoryless plant with arbitrary matrix gain**. Now, consider the linear multivariable memoryless plant described by

$$y_n = Bu_n + v_n, \tag{16}$$

where $y_n = [y_n^{(1)}, \ldots, y_n^{(N)}]^T$ is the $N$-dimensional output vector to be measured at $n$th time instant, $u_n = [u_n^{(1)}, \ldots, u_n^{(N)}]^T$ is the $N$-dimensional vector of unmeasurable disturbances and

$$B = \begin{pmatrix} b_{11} & \ldots & b_{1N} \\ \ldots & \ldots & \ldots \\ b_{N1} & \ldots & b_{NN} \end{pmatrix} \tag{17}$$

is an arbitrary $N \times N$ matrix gain.

It is assumed that the elements of the matrix $B$ in (17) are all unknown. However, there are some interval estimates

$$\underline{b}_{ik} \le b_{ik} \le \bar{b}_{ik}, \quad i, k = 1, \ldots, N \tag{18}$$

with the known upper and lower bounds. This implies that $B$ in (16) may be ill-conditioned or even singular, in general. Hence its rank satisfies

$$\text{rank } B \le N.$$

Suppose $\{v_n^{(i)}\} \in \ell_\infty$, where $\ell_\infty$ denotes the space of all bounded scalar sequences $\{x_n\}$ having the norm $\|x\|_\infty = \sup\limits_{0 \le n < \infty} |x_n| < \infty$. Thus,

$$|v_n^{(i)}| \le \varepsilon_i < \infty \quad \forall\, i = 1, \ldots, N, \tag{19}$$

where $\varepsilon_i$ s are constant. We assume that they are unknown, and it is essential.

Let $y^0 = [y^{0(1)}, \ldots, y^{0(N)}]^T$ denote the desired output vector whose components satisfy

$$|y^{0(1)}| + \ldots + |y^{0(N)}| \ne 0.$$

The problem is to design an adaptive controller of the form

$$u_{n+1} = U_n(u_n, y_n, y^0), \tag{20}$$

to be able to guarantee the boundedness of all signals in the closed-loop system (16), (20), i.e.,

$$\limsup_{n \to \infty} (\|u_n\| + \|y_n\|) < \infty \tag{21}$$

provided that the assumptions (18) and (19) hold. In the expression (20), $U_n : \mathrm{IR}^N \times \mathrm{IR}^N \times \mathrm{IR}^N \to \mathrm{IR}^N$ represents a time-varying linear operator defined later.

## MAIN RESULTS

**The convergence conditions for the learning algorithm in neural network model**. The global stochastic convergence analysis of the gradient learning algorithm (13) (for an arbitrary $w(0)$) is based on employing the fundamental convergence conditions established in the following Key Technical Lemma which is the slightly reformulated Theorem 3 of [32].

**Key Technical Lemma**. Let $V(w)$ be the so-called Lyapunov function to be satisfy the following properties:

$$V(w) = 0 \text{ if } w \in W^* \text{ and } V(w) > 0 \text{ if } w \notin W^*, \tag{22}$$

and

$$\|\nabla V(w') - \nabla(w')\| \le L \|w' - w''\| \tag{23}$$

with the Lipschitz constant $L > 0$. Define the scalar variable

$$H(w) = \nabla_w V(w)^T \nabla_w E\{Q(x, w)\} \tag{24}$$

and denote

$$H_n(w) := \nabla_w V(w(n))^T \nabla_w \, E\{Q(x, \; w(n))\}.$$

Suppose:

(i) $H_n(w) \geq \theta_n V(w(n-1)), \; \theta_n > 0,$

(ii) $E\{\| \nabla_w Q(x, \; w(n)) \|^2\} \leq \tau_n V(w(n)),$

$\tau_n \geq 0.$

Introduce the additional variable

$$\nu_n = \eta(n)(\theta_n - L\eta(n)\tau_n/2). \tag{25}$$

Then the algorithm (13) yields $\lim_{n\to\infty} V_n = 0$ a.s. provided that $E\{w(0)\} < \infty$ and

$$0 \leq \nu_n \leq 1, \tag{26}$$

$$\sum_{n=0}^{\infty} \nu_n = \infty, \tag{27}$$

i.e., the limit

$$\lim_{n\to\infty} V_n = 0 \tag{28}$$

will be achieved with probability 1. ■

Related result followed from the Theorem 3′ of [32] is.

**Corollary**. Under the conditions of the Key Technical Lemma, if $\theta_n \equiv \theta = \text{const}$ and $\tau_n \equiv \tau = \text{const}$, and $\eta(n) \equiv \eta = \text{const}$, then $V_n \xrightarrow[n\to\infty]{} 0$ a.s. provided that

$$0 < \eta \leq 2(\theta - \varepsilon)/L\tau \quad (0 < \varepsilon < \theta) \tag{29}$$

is satisfied. ■

Consider, first, the case when $F(x)$ can exactly be approximated by a neural network representation for all $x \in X$ implying

$$F(x) \equiv \text{NN}(x, w^*). \tag{30}$$

In this case called in [8, p. 304] as the ideal case, we have $J(w^*) = 0$ (by virtue of (12), (14)).

Now, we are able to present the first convergence result summarized in the theorem below.

**Theorem 1**. Suppose the assumption (30) holds. Then the gradient algorithm (13) with a constant learning rate, $\eta(n) \equiv \eta$, will converge with probability 1 (in the sense that $V_n \xrightarrow[n\to\infty]{} 0$ a.s.) and

$$\lim_{n\to\infty} e(n) = 0 \quad \text{a.s.} \tag{31}$$

for any initial $w(0)$ chosen randomly so that $E\{Q(x, w(0))\} < \infty$ if the conditions (29) with $\theta$ and $\tau$ specified by

$$\theta := \inf_{w \notin W^*} \frac{\| \nabla_w E\{Q(x,w)\} \|^2}{E\{Q(x,w)\}}, \tag{32}$$

$$\tau := \sup_{w \notin W^*} \frac{E\{\| \nabla_w Q(x,w) \|^2\}}{E\{Q(x,w)\}} \tag{33}$$

are satisfied.

**Proof**. Set

$$V(w) = E\{Q(x,w)\}. \tag{34}$$

Then condition (22) and (23) can be shown to be valid. This indicates that $V(w)$ of the form (34) may be taken as the Lyapunov function. By virtue of (24) such a choice of $V(w)$ gives $H(w) = \| \nabla_w E\{Q(x,w)\} \|^2$. Putting $\theta_n \equiv \theta$ and $\tau_n \equiv \tau$ with $\theta$ and $\tau$ determined by (32) and (33), respectively, one can conclude that the conditions (i), (ii) of the Key Technical Lemma are satisfied. Applying its Corollary it proves that $\lim_{n\to\infty} V_n = 0$ with probability 1.

Due to the definition (34) of $V(w)$ together with the assumption (30), result (31) follows. ■

Now, consider general case, where $F(x)$ cannot exactly be approximated by $\mathrm{NN}(x, w)$ (as in (30)). Obviously, in this case, $\inf_w Q(x, w^*) \not\equiv 0$, and the choice of a constant learning rate, $\eta(n) = \eta$, is not appropriate [8].

The convergence results are established in the follow theorem.

**Theorem 2**. Subject to the conditions

$$\text{(a)} \quad \sum_{n=0}^{\infty} \eta(n) = \infty, \qquad \text{(b)} \quad \sum_{n=0}^{\infty} \eta^2(n) < \infty, \tag{35}$$

the gradient algorithm (13) yields

$$\lim_{n\to\infty} E\{Q(x, w(n))\} = \inf_{w} E\{Q(x, w))\} \quad \text{a.s.}$$

provided that $\theta > 0$ with $\theta$ determined by (32).

**Proof**. Setting

$$V_n = E\{Q(x, w(n))\} - \inf_{w} E\{Q(x, w)\}$$

it can show that the requirements (22) and (23) will be satisfied: $V(w^*) = 0,$ and $V(w) > 0$ for $w \neq w^*$. Since $E\{\| \nabla_w Q(x, w) \|\}^2 > 0$ for $w = w^*,$ it follows that condition (ii) of the Key Technical Lemma assumes $\tau_n \to \infty$ as $w(n) \to w^*$.

Suppose (ii) is not satisfied. Then, there is a finite $\bar{\tau}$ such that

$$E\{\| \nabla_w Q(x, w) \|\}^2 \le \tau_n V(w(n)) \text{ with } \tau_n \le \bar{\tau} < \infty. \tag{36}$$

Since $\tau_n$ is assumed to be finite, there exists a finite $n_0$ such that requirement (27) will be satisfied for all sufficiently large $n \ge n_0$ provided that (i) takes place with $\theta_n \ge \theta > 0$ and $E\{w(n_0)\} < \infty$ and the condition (b) of (35) is satisfied (due to the fact that (b) means $\eta(n) \to 0$ as $n \to \infty$).

Further, if the assumption $\tau_n \le \bar{\tau} < \infty$ holds then the series

$$\sum_{n=n_0}^{\infty} \eta_n \theta_n \text{ with } \theta_n \ge \theta > 0$$

diverges whereas the series

$$-\sum_{n=n_0}^{\infty} L\eta(n)\tau_n / 2$$

converges (because of the validity of (a)). This gives that (27) takes also place.

Since $\theta > 0,$ all the conditions of Key Technical Lemma are satisfied for $n \ge n_0$. By this Lemma, $\lim_{n\to\infty} V_n = 0$ a.s. Therefore, $\tau_n \xrightarrow[n\to\infty]{} \infty$ with probability 1. But this contradicts the assumption that $\tau_n \le \bar{\tau} < \infty$ (see (36)). Hence, this assumption is false. This fact proves the validity of result given in theorem. ■

**Remark 1**. Setting

$$\theta_n := \| \nabla_w E\{Q(x, w(n))\} \|^2 \big/ E\{Q(x, w(n))\},$$
$$\tau_n := E\{\| \nabla_w Q(x, w(n)) \|^2\} \big/ E\{Q(x, w(n))\}$$

it can be concluded that, under the condition of the Theorem 2, the following features are observed: $\theta_n > \theta > 0$, $\tau_n < \tau < \infty$ for all $n$.

**Remark 2**. The conditions established in the theorem 1 and 2 are sufficient to guarantee the global convergence of (13) (for any $w(0)$) with probability 1 both in ideal and non-ideal cases. Under these conditions, the requirement (15) in which

$$J(w(n)) \equiv E\{Q(x, w(n))\}$$

will obviously be satisfied (final result). Again, the essential feature of this result is that these convergence properties can be achieved without adding penalty term to $Q(x, w(n))$, as in [17].

Of course, the calculation of $\theta$ and $\tau$ for choosing the suitable constant learning rate, $\eta$, according to (32), (33) seems to be hard. Meanwhile, $\eta$ may be replaced by the time-varying $\eta(n)$ satisfying the requirements (29) if necessary. Note that they are usual in the stochastic learning theory [1].

**Adaptive control of the plant (16) with an arbitrary** $B$. Basic idea is the transaction from the adaptive identification of the true plant having the singular transfer matrix $B$ to the adaptive identification of a fictitious plant with the nonsingular transfer matrix of the form

$$\widetilde{B} = B + \delta_0 I, \tag{37}$$

where $I$ denotes the identity matrix and $\delta_0$ is a fixed quantity.

Although $\widetilde{B}$ as well as $B$ remain unknown, the requirement

$$\det \widetilde{B} \neq 0 \tag{38}$$

can always be satisfied by the suitable choice of $\delta_0$ in (37). In fact, each $i$th eigenvalue $\lambda_i(B)$ of $B$ lies in one of the $N$ closed regions of the complex $z$-plane consisting of all the Gerŝgorin discs [33, p. 146]

$$|z - b_{ii}| \leq \sum_{\substack{k=1 \\ k \neq i}}^{N} |b_{ik}|, \quad i = 1, \ldots, N. \tag{39}$$

Since, at least, one of the eigenvalues $\lambda_i(B)$ is equal to zero (due to the singularity of $B$), by virtue of (8) there are the numbers

$$\underline{\beta}^{(i)} := b_{ii} - \sum_{\substack{k=1 \\ k \neq i}}^{N} |b_{ik}|, \quad \overline{\beta}^{(i)} := b_{ii} + \sum_{\substack{k=1 \\ k \neq i}}^{N} |b_{ik}|, \tag{40}$$

such that if

$$|b_{i1}|+\ldots+|b_{iN}|\neq 0 \tag{41}$$

then either $\underline{\beta}^{(i)}\le 0$ but $\overline{\beta}^{(i)}>0$ or $\underline{\beta}^{(i)}<0$ but $\overline{\beta}^{(i)}\ge 0$. These numbers define the intersection points of the *i*th Gerŝgorin disc with the real axis of the complex *z*-plane as show in Figs 1 and 2, left. In both cases, $\underline{\beta}^{(i)}\,\overline{\beta}^{(i)}\le 0$ if (11) is satisfied because $\underline{\beta}^{(i)}$ and $\overline{\beta}^{(i)}$ cannot have the same sign.

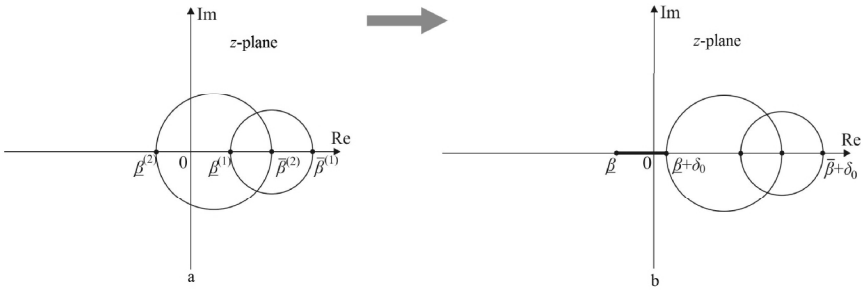


Fig. 1. The Gerŝgorin discs for N=2 in the case $|\underline{\beta}^{(2)}|<|\overline{\beta}^{(1)}|$

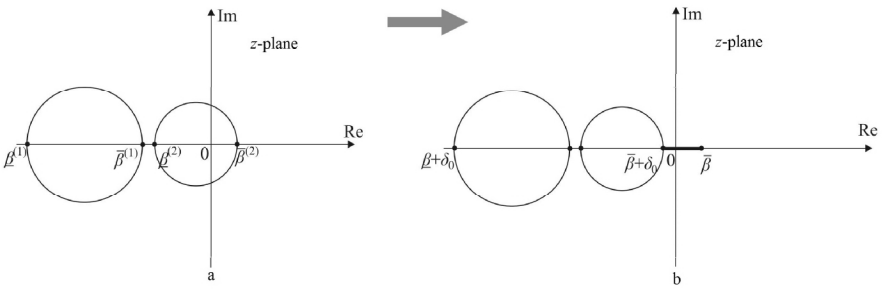


Fig. 2. The Gerŝgorin discs for N=2 in the case $|\overline{\beta}^{(2)}|<|\underline{\beta}^{(1)}|$

Denote

$$\underline{\beta}:=\min\{\underline{\beta}^{(1)},\ldots,\underline{\beta}^{(N)}\},\quad \overline{\beta}:=\max\{\overline{\beta}^{(1)},\ldots,\overline{\beta}^{(N)}\} \tag{42}$$

and consider the following cases: (a) $|\underline{\beta}|<|\overline{\beta}|$; (b) $|\underline{\beta}|>|\overline{\beta}|$ (The case when $|\underline{\beta}|=|\overline{\beta}|$ can be combined with any two cases.) In order to go to the transfer matrix $\widetilde{B}$ of the fictitious plant having the form (37) in the case (a), it is sufficient to shift the Gerŝgorin disc (39) right taking

$$\delta_0>|\underline{\beta}|, \tag{43}$$

as shown in Fig. 1, right. In the case (b), the discs (39) need to be shifted left according to

$$\delta_0 < -|\bar{\beta}|. \tag{44}$$

See Fig. 1, right. In both cases, the nonsingularity of $\widetilde{B}$ is guaranteed. Nevertheless, the conditions (43) and (44) cannot be satisfied, as yet. In fact, the numbers $\underline{\beta}$ and $\bar{\beta}$ given by the expressions (42) depend of $\underline{\beta}^{(i)}$ and $\bar{\beta}^{(i)}$s defined by (40). But they are unknown because $b_{ik}$s are all unknown.

To choose a number $\delta_0$ satisfying (38), we propose the following actions.

Define

$$\underline{\beta}^{(i)}_{\min} := \underline{b}_{ii} - \sum_{\substack{k=1 \\ k\neq i}}^{N} \max\{|\underline{b}_{ik}|, |\bar{b}_{ik}|\}, \tag{45}$$

$$\bar{\beta}^{(i)}_{\max} := \bar{b}_{ii} + \sum_{\substack{k=1 \\ k\neq i}}^{N} \max\{|\underline{b}_{ik}|, |\bar{b}_{ik}|\}, \tag{46}$$

minimizing and maximizing the right side of (40) for $\underline{\beta}^{(i)}$ and $\bar{\beta}^{(i)}$, respectively in $b_{ik} \in [\underline{b}_{ik}, \bar{b}_{ik}]$.

Now, introduce such quantities:

$$\begin{aligned} \underline{\beta}_{\min} &:= \min\{\underline{\beta}^{(1)}_{\min}, \ldots, \underline{\beta}^{(N)}_{\min}\}, \\ \bar{\beta}_{\max} &:= \max\{\bar{\beta}^{(1)}_{\max}, \ldots, \bar{\beta}^{(N)}_{\max}\}. \end{aligned} \tag{47}$$

Then $\delta_0$ has to satisfy the conditions

$$\begin{aligned} &\delta_0 > -\underline{\beta}_{\min} \quad \text{for} \quad |\underline{\beta}_{\min}| < |\bar{\beta}_{\max}|, \\ &\delta_0 < -\bar{\beta}_{\max} \quad \text{for} \quad |\underline{\beta}_{\min}| > |\bar{\beta}_{\max}|. \end{aligned} \tag{48}$$

It can be clarified that if (48) together with (45)–(47) will be satisfied then the condition (38) will without fail be ensured.

After determining the number $(\delta_0)$ we able to proceed to the consideration of the fictitious plant. Since the input variables $u_n^{(1)}, \ldots, u_n^{(N)}$ and the disturbances $v_n^{(1)}, \ldots, v_n^{(N)}$ of both true plant and fictitious plant are the same, this feature allows to describe our fictitious plant by the equation

$$\widetilde{y}_n = \widetilde{B}u_n + v_n, \tag{49}$$

similar to (1). In this equation, $\widetilde{y}_n = [\widetilde{y}_n^{(1)}, \ldots, \widetilde{y}_n^{(N)}]^T$ denotes the output vector of the fictitious plant.

It is interesting that the components of $\widetilde{y}_n$ can be measured while the components of $v_n$ in (49) remain unmeasurable. In fact, substituting (37) into (49) due to (16) we produce

$$\widetilde{y}_n = y_n + \delta_0 u_n. \tag{50}$$

It is seen from (50) that $\widetilde{y}_n$ can always be found indirectly having $u_n$ and $y_n$ to be measured.

Now, our problem reduces to the problem of adaptive control applicable to the fictitious plant (49) with the unknown transfer matrix $\widetilde{B}$ in the presence of arbitrary bounded disturbances $v_n^{(1)},\ldots,v_n^{(1)}$ whose bounds, $\varepsilon_i$ s, are also unknown. As in [4, item 4.2.3], the adaptive control law is designed in the from

$$u_{n+1} = u_n + \widetilde{B}_n^{-1}\widetilde{e}_n, \tag{51}$$

where instead of the current estimate $B_n$ of $\widetilde{B}$ is exploited where as the error vector

$$e_n = y^0 - y_n$$

is replaced by

$$\widetilde{e}_n = y^0 - \widetilde{y}_n. \tag{52}$$

with $\widetilde{y}_n$ given by the expression (50).

The adaptive identification algorithm used to determine the estimates $\widetilde{B}_n$ may be taken as

$$\widetilde{b}_n^{(i)} = \widetilde{b}_{n-1}^{(i)} - \gamma_n^{(i)} \frac{f(\widetilde{e}_n^{*(i)}, \varepsilon_{n-1}^{(i)})}{1+\|\nabla u_n\|^2} \nabla u_n \operatorname{sign} \widetilde{e}_n^{*(i)}, \quad i=1,\ldots,N, \tag{53}$$

which is similar to that in [4, item 4.2.3]. In this algorithm, the following notations are introduced $\widetilde{b}_n^{(i)T} = [\widetilde{b}_{i1}(n),\ldots,\widetilde{b}_{iN}(n)]$, $\nabla u_n := u_n - u_{n-1}$.

$$f(e,\bar{\varepsilon}) = \begin{cases} 0 & \text{if } |e| \le \bar{\varepsilon}, \\ |e| - \bar{\varepsilon} & \text{otherwise} \end{cases} \tag{54}$$

represents the dead-zone function depending on

$$\widetilde{e}_n^{*(i)} = \nabla\widetilde{y}_n^{(i)} - \widetilde{b}_{n-1}^{(i)\mathrm{T}} \nabla u_n, \tag{55}$$

which is the $i$th component of $\widetilde{e}_n^* = [\widetilde{e}_n^{*(1)},\ldots,\widetilde{e}_n^{*(N)}]^T$, and on the past estimate $\varepsilon_{n-1}^{(i)}$ of the unknown $\varepsilon^{(i)}$ found at the previous ($n$-1)th step. $\gamma_n^{(i)}$ is the coefficient chosen as

$$0 < \gamma' \le \gamma_n \le \gamma'' < 2 \tag{56}$$

to ensure $\det \widetilde{B}_n \ne 0$.

The algorithm for estimating $\varepsilon_n^{(i)}$s is specified by

$$\varepsilon_n^{(i)} = \varepsilon_{n-1}^{(i)} + \gamma_n^{(i)} \frac{f(\tilde{e}_n^{*(i)}, \varepsilon_{n-1}^{(i)})}{1 + \|\nabla u_n\|^2}, \quad i = 1, \ldots, N. \tag{57}$$

**Remark 3**. The dead-zone function $f(e, \bar{\varepsilon})$ depicted in Fig. 3 differs from the dead-zone function utilized in the standard adaptation algorithms similar to (53) in that it is nonnegative and also its size is time-varying variable determined by the past estimate of the unknown bound on the disturbance.

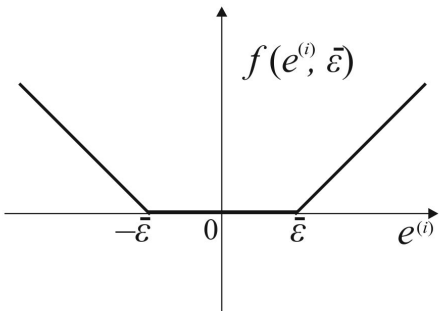


Fig. 3. The dead-zone function of the form (54)

The asymptotic properties of the adaptive control algorithm designed above are established in the following theorem.

**Theorem 3**. Let the assumptions (18), (19) and (41) be valid. Consider the adaptive feedback control system containing the plant (16), the controller (51), (52) together with the adaptation algorithm (53) to (57). Put $\varepsilon_0^{(i)} = 0$ for all $i = 1, \ldots, N$ and choose any initial estimate $\widetilde{B}_0 = B_0 + \delta_0 I$ from the conditions $\underline{b}_{ik} \le b_{ik}(0) \le \bar{b}_{ik}$.

Then:

(i) the sequence $\{\widetilde{B}_n\} := \widetilde{B}_1, \widetilde{B}_2, \ldots,$ caused by (53)–(56) converges, i.e., $\lim\limits_{n\to\infty} \widetilde{B}_n = \widetilde{B}_\infty$;

(ii) the sequence $\{\varepsilon_n^{(i)}\} = \varepsilon_0^{(i)}, \varepsilon_1^{(i)}, \ldots$ caused by (57) is nondecreasing and is convergent, i.e.,

$\lim\limits_{n\to\infty} \{\varepsilon_n^{(i)}\} = \varepsilon_\infty^{(i)}, \quad i = 1, \ldots, N;$

(iii) the requirement (21) is satisfied.

**Proof.** The validity of (i) and (ii) follows immediately from the results which can be found in [4, subsect. 4.2]. It is based on exploiting the fact that

$$V^{(i)}(n) = V_{\tilde{b}}^{(i)}(n) + V_{\varepsilon}^{(i)}(n),$$

with

$$V_{\tilde{b}}^{(i)}(n) := \| \tilde{b}^{(i)} - \tilde{b}_n^{(i)} \|^2 \text{ and } V_{\varepsilon}^{(i)}(n) := \| 2\varepsilon - \varepsilon_n^{(i)} \|^2,$$

where $\tilde{b}^{(i)} = [\tilde{b}_{i1}, \ldots, \tilde{b}_{iN}]^T$, is the Lyapunov function of the algorithm (53) to (57).

The proof of (iii) is based on the boundedness property for $\{\tilde{y}_n^{(i)}\}$ established in [4, subsect. 4.2]. Due to this property it can be written

$$\limsup_{n\to\infty} \| \tilde{y}_n \| < \infty. \tag{58}$$

Substituting (49) into (58) gives

$$\limsup_{n\to\infty} (\tilde{B}u_n + v_n) < \infty. \tag{59}$$

Since $\{v_n\} \in \underbrace{\ell_\infty \times \cdots \times \ell_\infty}_{N}$, and $\tilde{B}$ is nonsingular from (59) it can be concluded that $\{u_n\} \in \underbrace{\ell_\infty \times \cdots \times \ell_\infty}_{N}$. By virtue of the boundedness of $\{v_n\}$ it yields the boundedness of $\{y_n\}$. Finally, this fact proves the validity of (iii). ■

## CONCLUSIONS

The Lyapunov function approach is the suitable tool for analyzing the asymptotic behavior both of the gradient learning algorithm in the neural network identification systems and of the adaptive gradient algorithm in the certain closed-loop control systems.

Using the approach above mentioned, the two groups of global sufficient conditions which guarantee that the online gradient learning algorithm in neural network model for the identification of uncertain nonlinear systems acting in the stochastic environment will converge with probability 1. The first group of these conditions define the requirements under which this algorithm will be convergent a.s. with a constant learning rate. Such an asymptotic property holds in the ideal case where the nonlinearity to be identified can exactly be described by a neural network model. The second group of convergence conditions shows that this property can also be achieved in non-ideal case. Note that adding a penalty term to the current error function is indeed not necessary to guarantee this property.

It is established that in a worst case where the matrix gain of multivariable plant to be controlled is unknown and may be singular, and the bounds on the arbitrary unmeasurable disturbances remain unknown, the convergence of the gradient adaptation algorithm and the boundedness of all signals in the adaptive closed-loop system can be guaranteed.


1. Tsypkin Ya.Z. Adaptation and Learning in Automatic Systems. N.Y.: Academic Press, 1971.
2. Tsypkin Ya.Z. Foundation of the Theory of Learning Systems. N.Y.: Academic Press, 1973.
3. Kuntsevich V.M. Control under Uncertainty Conditions: Guaranteed Results in Control and Identification Problems. Kiev: Nauk. dumka, 2006. (in Russian).
4. Zhiteckii L.S. and Skurikhin V.I. Adaptive Control Systems with Parametric and Nonparametric Uncertainties. Kiev: Nauk. dumka, 2010. (in Russian).
5. Suykens J. and Moor B.D. Nonlinear system identification using multilayer neural networks: some ideas for initial weights, number of hidden neurons and error criteria. In Proc. 12nd IFAC World Congress, 1993, vol. 3, pp. 49–52.
6. Kosmatopoulos E.S., Polycarpou M.M., Christodoulou M.A. and Ioannou P.A. High-order neural network structures for identification of dynamical systems. IEEE Trans. on Neural Networks, 1995, vol. 6, pp. 422–431.
7. Levin A.U. and Narendra K.S. Recursive identification using feedforward neural networks. Int. J. Control, 1995, vol. 61, pp. 533–547.

8. Tsypkin Ya.Z., Mason J.D., Avedyan E.D., Warwick K. and Levin I.K. Neural networks for identification of nonlinear systems under random piecewise polynomial disturbances. IEEE Trans. on Neural Networks, 1999, vol. 10, pp. 303–311.
9. Behera L., Kumar S., and Patnaik A. On adaptive learning rate that guarantees convergence in feedforward networks. IEEE Trans. on Neural Networks, 2006, vol. 17, pp. 1116–1125.
10. White H. Some asymptotic results for learning in single hidden-layer neural network models. J. Amer. Statist. Assoc., 1987, vol. 84, pp. 117–134.
11. Kuan C M. and Hornik K. Convergence of learning algorithms with constant learning rates. IEEE Trans. on Neural Networks, 1991, vol. 2, pp. 484 – 489.
12. Luo Z. On the convergence of the LMS algorithm with adaptive learning rate for linear feedforward networks. Neural Comput., 1991, vol. 3, pp. 226–245.
13. Finnoff W. Diffusion approximations for the constant learning rate backpropagation algorithm and resistance to local minima. Neural Comput., 1994, vol. 6, pp. 285– 295.
14. Gaivoronski A.A. Convergence properties of backpropagation for neural nets via theory of stochastic gradient methods. Optim. Methods Software, 1994, vol. 4, pp. 117–134.
15. Fine T.L. and Mukherjee S. Parameter convergence and learning curves for neural networks. Neural Comput., 1999, vol. 11, pp. 749–769.
16. Tadic V. and Stankovic S. Learning in neural networks by normalized stochastic gradient algorithm: Local convergence. In Proc. 5th Seminar Neural Netw. Appl. Electr. Eng., 2000, pp. 11–17.
17. Zhang H., Wu W., Liu F. and Yao M. Boundedness and convergence of online gradient method with penalty for feedforward neural networks. IEEE Trans. on Neural Networks, 2009, vol. 20, pp. 1050–1054.
18. Mangasarian O.L. and Solodov M.V. Serial and parallel backpropagation convergence via nonmonotone perturbed minimization. Optim. Methods Software, 1994, pp. 103–106.
19. Wu W., Feng G. and Li X. Training multilayer perceptrons via minimization of ridge functions. Advances in Comput. Mathematics, vol. 17, pp. 331–347, 2002.
20. Zhang N., Wu W. and Zheng G. Convergence of gradient method with momentum for two-layer feedforward neural networks. IEEE Trans. on Neural Networks, 2006, vol. 17, pp. 522–525.
21. Wu W., Feng G., Li X and Xu Y. Deterministic convergence of an online gradient method for BP neural networks. IEEE Trans. on Neural Networks, 2005, vol. 16, pp. 1–9.
22. Xu Z.B., Zhang R. and Jing W.F. When does online BP training converge? IEEE Trans. on Neural Networks, 2009, vol. 20, pp. 1529–1539.
23. Shao H., Wu W. and Liu L. Convergence and monotonicity of an online gradient method with penalty for neural networks. WSEAS Trans. Math., 2007, vol. 6, pp. 469–476.
24. Ellacott S.W. The numerical analysis approach. In Mathematical Approaches to Neural Networks (Taylor J.G. ed; B.V.: Elsevier Science Publisher), 1993, pp. 103–137.
25. Skantze F.P., Kojic A., Loh A.P. and Annaswamy A.M. Adaptive estimation of discrete time systems with nonlinear parameterization. Automatica, 2000, vol. 36, pp. 1879–1887.
36. Loeve M. Probability Theory. N.Y.: Springer-Verlag, 1963.
27. Zhiteckii L.S., Azarskov V.N. and Nikolaienko S.A. Convergence of learning algorithms in neural networks for adaptive identification of nonlinearly parameterized systems. In Proc. 16th IFAC Symposium on System Identification, 2012, pp. 1593–1598.
28. Skurikhin V.I., Gritsenko V.I., Zhiteckii L.S. and Solovchuk K.Yu. Generalized inverse operator method in the problem of optimal controlling linear interconnected static plants. Dopovidi Natsionalnoi Akademii Nauk Ukrainy, 2014, no. 8, pp. 57–66. (in Russian).
29. Fomin V.N., Fradkov A.L. and Yakubovich V.A. Adaptive Control of Dynamic Systems. Moscow: Nauka, 1981. (in Russian).
30. Goodwin G.C. and Sin K.S. Adaptive Filtering, Prediction and Control. Engewood Cliffs. NJ.: Prentice-Hall, 1984.
31. Azarskov V.N., Zhiteckii L.S. and Solovchuk K.Yu. Adaptive robust control of multivariable static plants with possibly singular transfer matrix. Electronics and Control Systems, 2013, no. 4, pp. 47–53.

32. Polyak B.T.Convergence and convergence rate of iterative stochastic algorithms, I: General case. Autom. Remote Control, 1976, vol. 12, pp. 1858–1868.
33. Marcus M. and Minc H. A Survey of Matrix Theory and Matrix Inequalities. Boston: Allyn & Bacon Inc. 1964.

UDC 681.5

# ADAPTATION AND LEARNING IN SOME CLASSES OF IDENTIFICATION AND CONTROL SYSTEMS

**L.S. Zhiteckii, S.A. Nikolaienko, K.Yu. Solovchuk**

***International Research and Training Center for Information Technologies and Systems of the National Academy of Science of Ukraine and Ministry of Education and Sciences of Ukraine, Kiev, Ukraine***

**Introduction.** The paper deals with studying the asymptotical properties of the standard discrete-time gradient online learning algorithm in the two-layer neural network model of the uncertain nonlinear system to be identified. Also, the design of the discrete-time adaptive closed-loop system containing the linear multivariable memoryless plant with possibly singular but unknown matrix gain in the presence of unmeasurable bounded disturbances having the unknown bounds are addressed in this paper. It is assumed that the learning process in the neural network model is implemented in the stochastic environment whereas the adaptation of the plant model in the control system is based on the non-stochastic description of the external environment.

**The purpose** of the paper is to establish the global convergence conditions of the gradient online learning algorithm in the neural network model by utilizing the probabilistic asymptotic analysis and to derive the convergent adaptive control algorithm guaranteeing the boundedness of the signals in the closed-loop system which contains the multivariable memoryless plant with an arbitrary matrix gain in the presence of unmeasurable disturbances whose bounds are unknown.

**Results.** The Lyapunov function approach as the suitable tool for analyzing the asymptotic behavior both of the gradient learning algorithm in the neural network identification systems and of the adaptive gradient algorithm in the certain closed-loop control systems is utilized. Within this approach, the two groups of global sufficient conditions guaranteeing the convergence of the online gradient learning algorithm in neural network model with probability 1 are obtained. The first group of these conditions defines the requirements under which this algorithm will be convergent almost sure with a constant learning rate. Such an asymptotic property holds in the ideal case where the nonlinearity to be identified can exactly be described by a neural network model. The second group of convergence conditions shows that this property can also be achieved in non-ideal case. It turns out that adding a penalty term to the current error function is indeed not necessary to guarantee this property. It is established that in a worst case where the matrix gain of multivariable plant is unknown and may be singular, and the bounds on the arbitrary unmeasurable disturbances remain unknown, the convergence of the gradient adaptation algorithm and the boundedness of all signals in the adaptive closed-loop system can be ensured.

**Conclusions.** In order to guarantee the global convergence of the online learning algorithm in the neural network identification system with probability 1,

the certain conditions should be satisfied. Also the boundedness of all signals in the closed-loop adaptive control system containing the multivariable memoryless plant whose matrix gain is unknown and possibly singular can be achieved even if the bounds on the unmeasurable disturbances are unknown.

***Keywords:*** neural network, gradient learning algorithm, convergence, multivariable memoryless plant, adaptive control algorithm, boundedness of the signals.

1. Tsypkin Ya.Z. *Adaptation and Learning in Automatic Systems*. N.Y.: Academic Press, 1971.
2. Tsypkin Ya.Z. *Foundation of the Theory of Learning Systems*. N.Y.: Academic Press, 1973.
3. Kuntsevich V.M. *Control under Uncertainty Conditions: Guaranteed Results in Control and Identification Problems*. Kiev: Nauk. dumka, 2006. (in Russian).
4. Zhiteckii L.S. and Skurikhin V.I. *Adaptive Control Systems with Parametric and Nonparametric Uncertainties*. Kiev: Nauk. dumka, 2010. (in Russian).
5. Suykens J. and Moor B.D. Nonlinear system identification using multilayer neural networks: some ideas for initial weights, number of hidden neurons and error criteria. In *Proc. 12nd IFAC World Congress*, 1993, vol. 3, pp. 49–52.
6. Kosmatopoulos E.S., Polycarpou M.M., Christodoulou M.A. and Ioannou P.A. High-order neural network structures for identification of dynamical systems. *IEEE Trans. on Neural Networks*, 1995, vol. 6, pp. 422–431.
7. Levin A.U. and Narendra K.S. Recursive identification using feedforward neural networks. *Int. J. Control*, 1995, vol. 61, pp. 533–547.
8. Tsypkin Ya.Z., Mason J.D., Avedyan E.D., Warwick K. and Levin I.K. Neural networks for identification of nonlinear systems under random piecewise polynomial disturbances. *IEEE Trans. on Neural Networks*, 1999, vol. 10, pp. 303–311.
9. Behera L., Kumar S., and Patnaik A. On adaptive learning rate that guarantees convergence in feedforward networks. *IEEE Trans. on Neural Networks*, 2006, vol. 17, pp. 1116–1125.
10. White H. Some asymptotic results for learning in single hidden-layer neural network models. *J. Amer. Statist. Assoc.*, 1987, vol. 84, pp. 117–134.
11. Kuan C M. and Hornik K. Convergence of learning algorithms with constant learning rates. *IEEE Trans. on Neural Networks*, 1991, vol. 2, pp. 484 – 489.
12. Luo Z. On the convergence of the LMS algorithm with adaptive learning rate for linear feedforward networks. *Neural Comput.*, 1991, vol. 3, pp. 226–245.
13. Finnoff W. Diffusion approximations for the constant learning rate backpropagation algorithm and resistance to local minima. *Neural Comput.*, 1994, vol. 6, pp. 285–295.
14. Gaivoronski A.A. Convergence properties of backpropagation for neural nets via theory of stochastic gradient methods. *Optim. Methods Software*, 1994, vol. 4, pp. 117–134.
15. Fine T.L. and Mukherjee S. Parameter convergence and learning curves for neural networks. *Neural Comput.*, 1999, vol. 11, pp. 749–769.
16. Tadic V. and Stankovic S. Learning in neural networks by normalized stochastic gradient algorithm: Local convergence. In *Proc. 5th Seminar Neural Netw. Appl. Electr. Eng.*, 2000, pp. 11–17.

17. Zhang H., Wu W., Liu F. and Yao M. Boundedness and convergence of online gradient method with penalty for feedforward neural networks. *IEEE Trans. on Neural Networks*, 2009, vol. 20, pp. 1050–1054.
18. Mangasarian O.L. and Solodov M.V. Serial and parallel backpropagation convergence via nonmonotone perturbed minimization. *Optim. Methods Software*, 1994, pp. 103–106.
19. Wu W., Feng G. and Li X. Training multilayer perceptrons via minimization of ridge functions. *Advances in Comput. Mathematics*, vol. 17, pp. 331–347, 2002.
20. Zhang N., Wu W. and Zheng G. Convergence of gradient method with momentum for two-layer feedforward neural networks. *IEEE Trans. on Neural Networks*, 2006, vol. 17, pp. 522–525.
21. Wu W., Feng G., Li X and Xu Y. Deterministic convergence of an online gradient method for BP neural networks. *IEEE Trans. on Neural Networks*, 2005, vol. 16, pp. 1–9.
22. Xu Z.B., Zhang R. and Jing W.F. When does online BP training converge? *IEEE Trans. on Neural Networks*, 2009, vol. 20, pp. 1529–1539.
23. Shao H., Wu W. and Liu L. Convergence and monotonicity of an online gradient method with penalty for neural networks. *WSEAS Trans. Math*., 2007, vol. 6, pp. 469–476.
24. Ellacott S.W. The numerical analysis approach. In *Mathematical Approaches to Neural Networks* (Taylor J.G. ed; B.V.: Elsevier Science Publisher), 1993, pp. 103–137.
25. Skantze F.P., Kojic A., Loh A.P. and Annaswamy A.M. Adaptive estimation of discrete time systems with nonlinear parameterization. *Automatica*, 2000, vol. 36, pp. 1879–1887.
26. Loeve M. *Probability Theory*. N.Y.: Springer-Verlag, 1963.
27. Zhiteckii L.S., Azarskov V.N. and Nikolaienko S.A. Convergence of learning algorithms in neural networks for adaptive identification of nonlinearly parameterized systems. In *Proc. 16th IFAC Symposium on System Identification*, 2012, pp. 1593–1598.
28. Skurikhin V.I., Gritsenko V.I., Zhiteckii L.S. and Solovchuk K.Yu. Generalized inverse operator method in the problem of optimal controlling linear interconnected static plants. *Dopovidi Natsionalnoi Akademii Nauk Ukrainy*, 2014, no. 8, pp. 57–66. (in Russian).
29. Fomin V.N., Fradkov A.L. and Yakubovich V.A. *Adaptive Control of Dynamic Systems*. Moscow: Nauka, 1981. (in Russian).
30. Goodwin G.C. and Sin K.S. *Adaptive Filtering, Prediction and Control*. Engewood Cliffs. NJ.: Prentice-Hall, 1984.
31. Azarskov V.N., Zhiteckii L.S. and Solovchuk K.Yu. Adaptive robust control of multivariable static plants with possibly singular transfer matrix. *Electronics and Control Systems*, 2013, no. 4, pp. 47–53.
32. Polyak B.T.Convergence and convergence rate of iterative stochastic algorithms, I: General case. *Autom. Remote Control*, 1976, vol. 12, pp. 1858–1868.
33. Marcus M. and Minc H. *A Survey of Matrix Theory and Matrix Inequalities*. Boston: Allyn & Bacon Inc. 1964.

Получено 06.07.2015