MATHEMATICAL MODELING OF PROCESSES AND SYSTEMS

UDC 681.5(045)

[1]V. N. Azarskov,
[2]L. S. Zhiteckii,
[3]S. A. Nikolaienko

SEQUENTIAL LEARNING PROCESSES IN NEURAL NETWORKS APPLIED AS MODELS OF NONLINEAR SYSTEMS

Institute of Air Navigation, National Aviation University, Kyiv, Ukraine
E-mails: [1]azarskov@nau.edu.ua, [2]leonid_zhiteckii@i.ua, [3]s_nikolaenko@ukr.net

Abstract. *Asymptotic properties of the online gradient algorithm with a constant step size employed for learning in neural network models of nonlinear systems having one hidden layer are examined. Some conditions guaranteeing the convergence of this algorithm are established.*

Keywords: nonlinear system; neural network model; gradient algorithm; learning; convergence.

Introduction

Neural networks containing at least one hidden layer play a role of universal models for any reasonable complex nonlinear systems, in particular, flight control systems. Namely, for the reliable operation of the modern aircraft control systems it is necessary to ensure the timely detection of failure situations, their localization and reconfiguration of the onboard control system. Currently, active studies in the application of the neural networks for designing fault-tolerant aircraft control systems are conducted. Again, the problem of synthesizing the algorithms for the detection and faults localization and the system reconfiguration using of neural networks is shown. Effectiveness of the proposed approach is confirmed by the results of simulation of the flight control system of the aircraft using a nonlinear mathematical model of F-16 fighter.

The fact above mentioned motivates the theoretical studies of learning algorithms for the neural network models. Significant breakthrough in this research area has been achieved in recent works [1]–[12]. Namely, the convergence results have been derived in [11] 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 algorithms with deterministic (non-stochastic) nature has been given in [12] by assuming that the learning set is finite. The difficulties in establishing the convergence results are that the neural networks contain the parameters which appear nonlinearly in their equations. To the best of author's knowledge, there are no results in literature concerning the convergence properties of training procedures with a fixed step size applicable to the case of infinite learning set.

This paper extends the recent work [13] of the authors. The main effort is focused on establishing conditions under which the online gradient algorithms applied for sequential learning neural network models with a constant step size will converge in the case of infinite learning set.

Problem statement

Let

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

be some nonlinear unknown function describing a complex system. In this equation, $y \in \mathrm{IR}$ and $x \in \mathrm{IR}^N$ are the output scalar and input vector variables, respectively, available for the measurement at each nth time instant $(n = 1, 2, \ldots)$. This implies that

$$y(n) = F(x(n-1)) \tag{2}$$

with an unknown mapping $F:\ \mathrm{IR}^N \to \mathrm{IR}$.

To approximate (1), 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 $x(n-1)$. Its output signal at the nth time instant is given by

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

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

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 as

$$y^{(2)}(n) = \sum_{j=1}^{M} w_j^{(2)} y_j^{(1)}(n) + b^{(2)}, \qquad (4)$$

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

Since $\sigma(\cdot)$ is assumed to be nonlinear, it follows from (3), (4) together with (2) that $y^{(2)}(n)$ is a nonlinear function depending on $x(n-1)$ and also on the $(M(N+2)+1)$-dimensional parameter vector

$$w = [w_{11}^{(1)},\ldots,w_{N1}^{(1)}, b_1^{(1)},\ldots,w_{1M}^{(1)},\ldots,w_{NM}^{(1)}, b_M^{(1)} \vdots \; w_1^{(2)},\ldots,w_M^{(2)}, b^{(2)}]^T.$$

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

$$y^{(2)}(n) = \mathrm{NN}(x(n-1), w) \qquad (5)$$

with $\mathrm{NN}: \mathrm{IR}^N \times \mathrm{IR}^{M(N+2)+1} \to \mathrm{IR}$.

The following basic assumption is made. There exists at least an unique $w = w^* \in \mathrm{IR}^{M(N+2)+1}$ such that $F(x)$ can explicitly be approximated by $\mathrm{NN}(x, w^*)$ in the sense of

$$F(x) \equiv \mathrm{NN}(x, w^*) \qquad (6)$$

for all x from a given compact set $X \subset \mathrm{IR}^N$. This assumption mentioned in [14] as the ideal case has only the mathematical meaning. Its introduction is motivated by the fact that the standard gradient type learning procedure with a constant step size cannot converge in the non-ideal cases if there is an infinite subsequence $\{x(n_k\} := x(n_1), x(n_2),\ldots$ satisfying

$$x(n_k) \in X \setminus \{x:\ F(x) - \mathrm{NN}(x, w) = 0\} \quad (k = 1,2,\ldots)$$

for any fixed w.

Define the infinite sequence $\{(x(n-1), y(n))\}_{n=1}^{\infty}$ of the measurable pairs in which $x(n-1)$s are taken from X. Then, the online learning algorithm for updating the parameter estimate $w(n)$ is formulated as the following standard recursive gradient procedure:

$$w(n) = w(n-1) + \eta\, e(n, w(n-1))\, \mathrm{grad}_w \mathrm{NN}(x(n-1), w(n-1)). \qquad (7)$$

In this algorithm,

$$e(n, w(n-1)) = y(n) - \mathrm{NN}(x(n-1), w(n-1)) \qquad (8)$$

is the current estimation error and the variable $\mathrm{grad}_w \mathrm{NN}(x(n-1), w(n-1))$ denotes the gradient of $\mathrm{NN}(x, w)$ at the point $w = w(n-1)$, and $\eta \equiv \mathrm{const} > 0$ is its step size (the learning rate).

Equations (2) and (7) together with (5), (6) and (8) describe the closed-loop system for sequential learning of the neural network exploiting as a model of (1). For better understanding its performance, the structure of this system is depicted in Fig. 1.

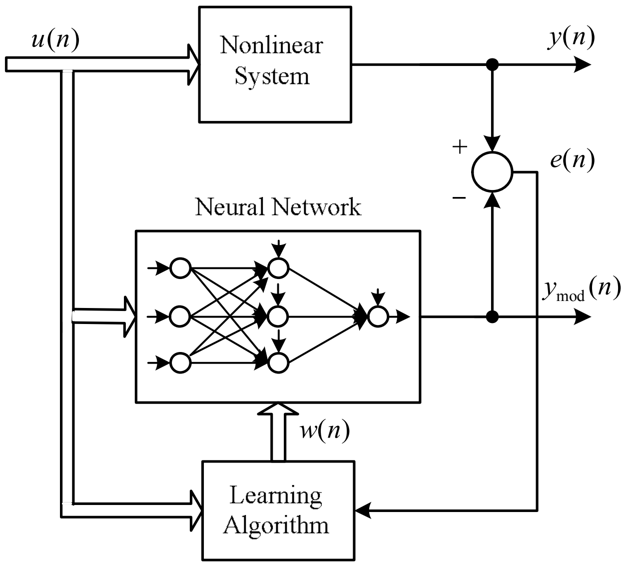


Fig. 1. Configuration of learning system

The problem is to study the properties of sequence $\{w(n)\}$ caused by (7), (8) as n tends to ∞ (in some sense specified below).

Basic definitions and investigation tool

Before going to study the asymptotical properties of $\{w(n)\}$, we need some preliminaries including several definitions.

Definition 1 [13]. Introduce the notation $x(n, \omega)$ of a scalar variable which depends on an ω as on peculiar event parameter for every fixed n. Let $\{\omega\}$ be a set of ω*s*. Then, for any integer r and for arbitrarily chosen numbers $x^{(1)},\ldots,x^{(r)}$, the sequence $\{x(n, \omega)\}$ is said to be non-stochastic (irregular) on IR in wide sense if these is some $\omega \in \{\omega\}$ such that

$$x(1, \omega) = x^{(1)},\ldots, x(r, \omega) = x^{(r)}. \qquad (9)$$

In contrast to the purely stochastic representation, the sequence given in (9) is quite not predictable: for example, there is no its expectation, in general. Note that in the particular case where $\{\omega\}$ contains a single ω, the sequence satisfying (9) becomes simply deterministic.

To analyze the asymptotic behavior of (7), (8), the scalar non-negative function $V(w)$ given by

$$V(w)=0 \text{ for } w=w^*, \quad V(w)>0 \text{ for } w\neq w^* \quad (10)$$

is exploited.

In the presence of the one-point set $W^*=\{w^*\}$, the function $V(w)$ satisfying (9) is usually chosen as

$$V(w)=\left\|w^*-w\right\|^2, \quad (11)$$

where $\|\cdot\|$ denotes the usual Euclidean vector norm. It turned out that if the neural network contains the hidden layer, then W^* consists of several isolated w^*s. In particular, in the simplest case, when there is one neuron in the hidden layer ($N=1$, $M=1$) and the activation function, $\sigma(\cdot)$ is described by

$$\sigma(s)=\left[1+\operatorname{exs}(-s)\right]^{-1}, \quad (12)$$

W^* contains two points: $w^{*(1)}=[w_1^*, w_2^*, w_3^*, w_4^*]^T$ and $w^{*(2)}=[-w_1^*, -w_2^*, -w_3^*, w_3^*+w_4^*]^T$.

In the case when W^* is not one-point, instead of (11), $V(w)$ may be chosen as follows:

$$V(w)=\inf_{\mathbf{w}^*\in W^*}\left\|w^*-w\right\|^2. \quad (13)$$

It can be observed that $V(w)$ specified by (13) is not continuous (in contrast to (11)).

The variable $V_n:=V(w(n))$ becomes immediately the Lyapunov function of the algorithm (7), (8) if only

$$V_n\leq V_{n-1} \quad \forall n. \quad (14)$$

Since $V_n\geq 0$, the condition (14) under which V_n does not increase is sufficient for existing a limit

$$\lim_{n\to\infty} V_n=V_\infty, \quad (15)$$

where $V_\infty=V_\infty(\omega)$ is a random value (in general) depending on $w(0)$ and $\{x(n)\}$. Nevertheless, the property (14) is not necessary for existing its limit. On the other hand, this limit may not exist.

Definition 2. The algorithm (7), (8) will be called weakly convergent if the limit (15) with $V_\infty\neq 0$ exists.

Definition 3. The algorithm (7), (8) will be called strongly convergent if $\lim_{n\to\infty} V(n)=0$ meaning that $w(n)\xrightarrow[n\to\infty]{} w^*$.

Any non-negative V_n which is the Lyapunov function of the learning algorithm (7), (8) may be employed as a tool to derive its convergence properties if $\{x(n)\}$ is a non-stochastic sequence. It turns out that in stochastic case, its counterpart satisfying

$$\mathsf{E}\{V_n\,|\,V_{n-1},\ldots,V_0\}\leq V_{n-1} \quad (16)$$

plays the role of the similar Lyapunov function (instead of (14)). In this expression, the symbol $\mathsf{E}\{V_n\,|\,\cdot\}$ denotes the conditional expectation of random V_n. Note that $\{V_n\}$ specified by (16) is known in the probability theory as the supermartingale [15].

An asymptotic property of learning algorithm

We first observed in simulation examples that the sequence $\{w(n)\}$ may not converge in the presence of non-stochastic $\{x(n)\}$ from X. Such an ultimate feature of (7), (8) implies that

$$\lim_{n\to\infty} w(n)=w_\infty \quad (17)$$

may not exist. Nevertheless, if (17) is achieved, then the following result can be established:

a) $\{w(n)\}$ converges to some $w_\infty\in\liminf W_n$, in sense of (17) where

$$\liminf W_n:=\bigcup_{n=1}^{\infty}\bigcap_{k=n}^{\infty}W_k \quad (18)$$

denotes the limit set introduced in [15, sect. 1.3] in which $W_n:=\{w:\ y(n)-\mathrm{NN}(x(n-1), w)=0\}$;

b) the estimation error given by (8) vanishes as n goes to infinity:

$$\lim_{n\to\infty} e(n,w(n-1))=0. \quad (19)$$

Note that the limit set $\liminf W_n$ represents a nonlinear manifold on $\mathrm{IR}^{M(N+2)+1}$ whose dimension satisfies $0\leq\dim\liminf W_n\leq M(N+2)$.

It can be understood that the algorithm (7), (8) "attempts" to solve the infinite set of the equations

$$y(n)-\mathrm{NN}(x(n-1), w)=0, \quad n=1,2,\ldots \quad (20)$$

with respect to unknown $w\in\mathrm{IR}^{M(N+L+2)+1}$. In fact, this algorithm may give the solution $w=w_\infty$ of the remainder of (20), which is determined as the limit set (18) but not a $w\in W^*$.

It is important to observe that if (19) is satisfied, then the algorithm (7), (8) ensures the so-called functional identification of (1) [13]. In this case, the discrepancy between the outputs of the system (1) to be identified and its neural network's model

$$y_{\text{mod}}(n) = \text{NN}(x(n-1), w)$$

depicted in Fig. 1, for $w = w(n-1)$ goes to zero as $n \to \infty$.

Observations

To study the asymptotic properties of (7), (8) in the presence of non-stochastic $\{x(n)\}$, four simulation experiments with the scalar nonlinear system

$$y = \frac{3.75 + 0.05\exp(-7.15x)}{1 + 0.19\exp(-7.15x)}$$

were conducted. This nonlinear function can explicitly be approximated by the two-layer neural network described by (3), (4), (12) with the components of two $w = w^{*(1)}$, $w = w^{*(2)}$ summarized in the table.

Parameters of neural network model

Exp. No	Parameter	$w_{11}^{(1)}$	$b_1^{(1)}$	$w_1^{(2)}$	$b^{(2)}$
1–4	Components of $w^{*(1)}$	7.15	1.65	3.45	0.3
	Components of $w^{*(2)}$	-7.15	-1.65	-3.45	3.75
1	Initial estimate	1.40	-0.10	-0.56	0.46
	Final estimate	≈ 7.1	≈ 1.6	≈ 3.4	≈ 0.3
2	Initial estimate	0.53	-0.50	-0.92	1.04
	Final estimate	5.41	1.32	3.82	-0.05
3	Initial estimate	0.38	-0.57	-0.98	1.14
	Final estimate	-5.13	-1.52	-4.20	3.78
4	Initial estimate	1.40	-0.10	-0.56	0.46
	Final estimate	do not exist			

In all of the experiments, η was taken as $\eta = 0.01$ and the duration of the learning processes was always equal to 40 000 steps.

In the first experiment, $\{x(n)\}$ was chosen to achieve the monotonic convergence of $\{V_n\}$ to a V_∞. Such a choice given that the first difference of V_n defined as $\Delta V_n := V_n - V_{n-1}$ was negative. It turned out that $V_\infty \approx 0$ meaning the strong convergence of (7), (8) (in the sense of Definition 3). Fig. 2 demonstrates this property. In this experiment, $\{w(n)\}$ with $w(0) = [1.40,\ -0.10,\ -0.56,\ 0.46]^{\text{T}}$ tends approximately to $w^{*(1)}$ (see the table). Again, V_n given by (13) as

$$V_n = \min\{V_n^{(1)}, V_n^{(2)}\},$$

where

$$V_n^{(i)} := \left\| w^{*(i)} - w(n) \right\|^2, \quad i = 1, 2$$

satisfies the identity $V_n \equiv V_n^{(1)}$.

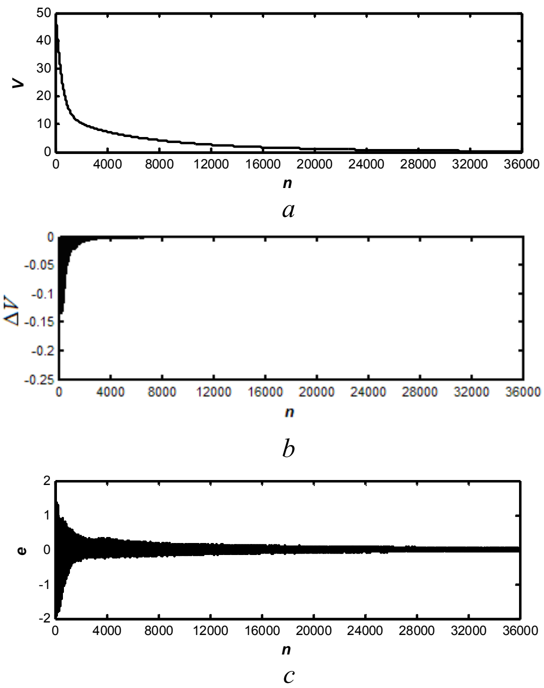
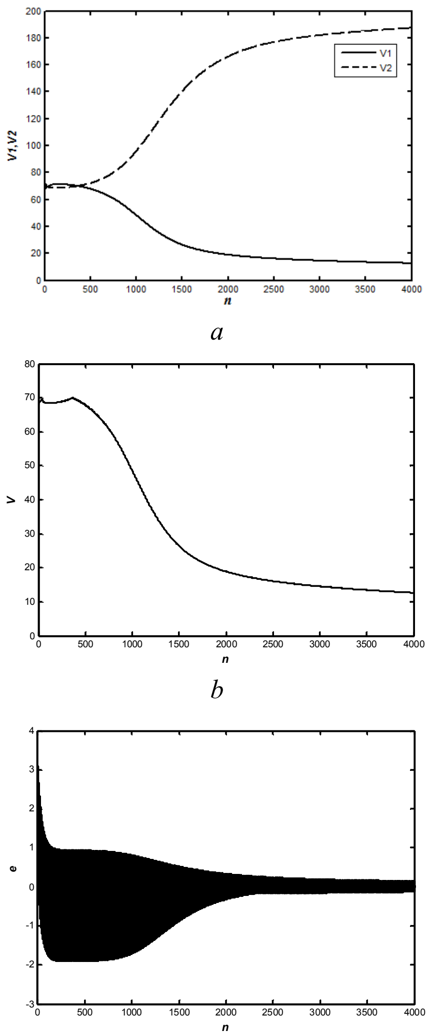


Fig. 2. Learning processes in simulation experiment 1: *a* is the function V_n; *b* is the first difference ΔV_n; *c* is the current model error $e(n)$

Fig. 3 illustrates the learning processes arising in the second experiment. In this experiment, the sequence $\{x(n)\}$ was taken cyclically from the finite set $X = \{\text{-}0.4442,\ 0.5158,\ 0.8761\}$. The initial estimate $w(0)$ whose components are given in the table was chosen in order to satisfy $V_n^{(1)}(0) < V_n^{(2)}(0)$. It was observed that at an initial stage of the learning process, the sequence $\{V_n^{(1)}\}$ was increasing so that $V_n^{(1)} > V_n^{(2)}$, as shown in Fig. 3a. Further, $\{V_n^{(1)}\}$ became decreasing. Such a behavior of this sequence leaded to appearing the feature that $V_n^{(1)} < V_n^{(2)}$ for all sufficiently large n.

Thus, we see that $\{V_n\}$ is convergent, however its convergence is not monotonic as in the first experiment. Namely, $\{w(n)\}$ converges to $w_\infty = [5.4120, 1.3172, 3.8233, -0.0475]^T$ which lies on the nonlinear manifold $\liminf W_n$ but not to one of two points $w_1^* = [7.15, 1.65, 3.45, 0.3]^T$ or to $w_1^* = [-7.15, -1.65, -3.45, 3.75]^T$.



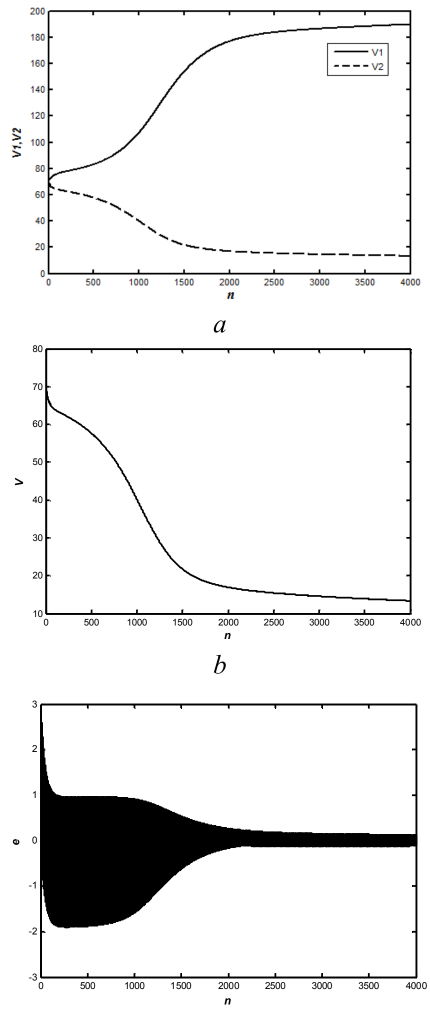
a *b* *c*

Fig. 3. Learning processes in simulation experiment 2: (*a*) are the functions $V_n^{(1)}$ and $V_n^{(2)}$; (*b*) is the function V_n; (*c*) is the current model error $e(n)$

In the third experiment, the initial vector $w(0)$ was chosen to be close to that in the second experiment (see Table). As in this previous experiment, $\{x(n)\}$ was taken from the same finite set X containing the tree points.

Results of the learning process is shown in Fig. 4. We observe that in this case, $V_n \equiv V_n^{(2)}$ (see Fig. 4a and b). However, V_∞ is nonzero because $\{w(n)\}$ goes to $w = [-5.13, -1.52, -4.20, 3.78]^{\mathrm{T}}$ belonging to the manifold $\liminf W_n$ determined by (18) but not to the nearest $w^{*(2)}$ from W^*.

Thus, as in the second experiment we deal with the weak convergence of $\{w(n)\}$ (in the sense of Definition 2) guaranteeing only that $e(n) \xrightarrow[n\to\infty]{} 0$.



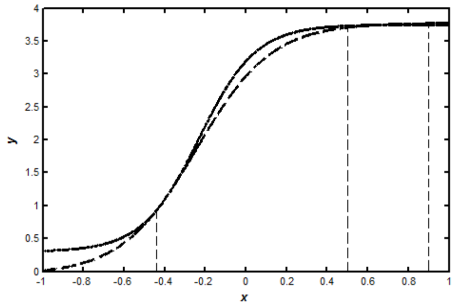
a *b* *c*

Fig. 4. Learning processes in simulation experiment 3: (*a*) are the functions $V_n^{(1)}$ and $V_n^{(2)}$; (*b*) is the function V_n; (*c*) is the current model error $e(n)$

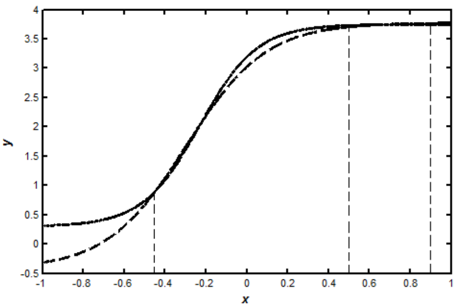
Results of the second and third experiments are presented in Fig. 5 showing the nonlinearity $y = F(x)$ given by (1) and its neural network model approximations $y_{\text{mod}} = F(x, w_\infty)$ obtained after 40 000 learning steps. It can be see that these nonlinearities are “somewhat” different whereas they are coinciding at three xs which belong to X.

Fig. 6 demonstrates the learning processes taking place in fourth experiment when $\{V_n\}$ has no limit when $\{x(n)\}$ is a non-stochastic sequence.

We see that the model error $e(n, w(n-1))$ does not go to zero, as n tends to ∞.

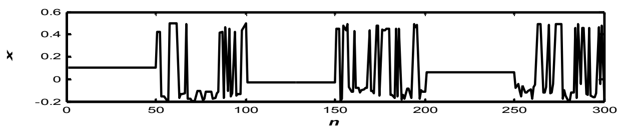


a

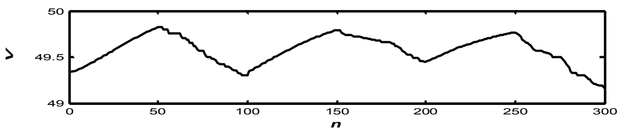


b

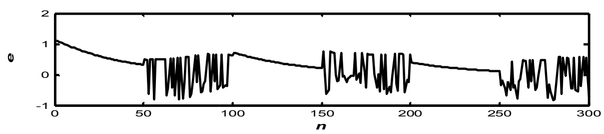
Fig. 5. The nonlinearities $y = F(x)$ (solid line) and $y_{\mathrm{mod}} = F(x, w_\infty)$ (dashed line): (*a*) in the experiment 2; (*b*) in the experiment 3



a



b



c

Fig. 6. Learning processes in simulation experiment 4: (*a*) is the input signal; (*b*) is the function V_n given by (14); (*c*) is the current model error $e(n)$

Stochastic case

Main theoretical result concerning the asymptotical behavior of (7), (8) in the stochastic case is based on following additional assumptions: $\{x_i(n)\}$ are the stochastic sequences of independent random variables having the probability density function

$$p(x(n) \mid x(n-1), \ldots, x(0)) \equiv p(x(n)) := p(x) \quad (21)$$

with the properties that

$$P\{x(n) \in X'\} = \int_{x \in X'} p(x)\, dx > 0, \quad (22)$$

for any subset $X' \subset X$, and

$$P\{x(n) \in X''\} = 0 \quad (23)$$

if $\dim X'' = 0$, where $P\{\cdot\}$ denotes the probability of corresponding event.

Let $W(w^*)$ denote a neighborhood of some $w^* \in W^*$ which does not contain another points of W^*. With this $W(w^*)$, we have established that if the assumptions (6), (21) – (23) are satisfied and the conditions

$$0 < \eta < 2,$$

$$\int_{x \in X} [\mathrm{NN}(x, w^*) - \mathrm{NN}(x, w)]\, \mathrm{grad}_w^T \mathrm{NN}(x, w) \cdot (w^* - w) p(x)\, dx$$
$$\geq \int_{x \in X} [\mathrm{NN}(x, w^*) - \mathrm{NN}(x, w)]^2 \cdot \|\mathrm{grad}_w \mathrm{NN}(x, w)\|^2 p(x)\, dx$$

hold for any $x \in X$ and for arbitrary w from $W(w^*)$, then the limit (15) is valid with probability 1. Again,

$$\lim_{n \to \infty} w(n) = w^*$$

almost sure (a.s.).

The proof of this result essentially utilizes the Borel – Cantelli lemma and Doob's martingale convergence theorem (see [13]).

To evaluate the asymptotical properties of the algorithm (7), (8) in the stochastic case, a simulation with $\{x(n)\}$ generated as the sequence of independent identically distributed (i.i.d.) pseudorandom numbers on $X = [-1.0,\ 1.0]$. The initial estimates were chosen as follows: $w_{11}^{(1)}(0) = 1.4$, $b_1^{(1)}(0) = -0.1$, $w_1^{(2)}(0) = -0.56$, $b^{(2)}(0) = 0.46$.

The asymptotical properties of the learning algorithm are illustrated in Fig. 7. It can be observed that V_n depicted in Fig. 7a is not monotonically decreasing. Nevertheless, $e(n) \to 0$ as n increases as shown in Fig. 7c. This important feature follows from the fact that (16) is satisfied. Fig. 7d in which the conditional expectation $\mathsf{E}\{V_n \mid \cdot\}$ was estimated numerically, demonstrates this feature.

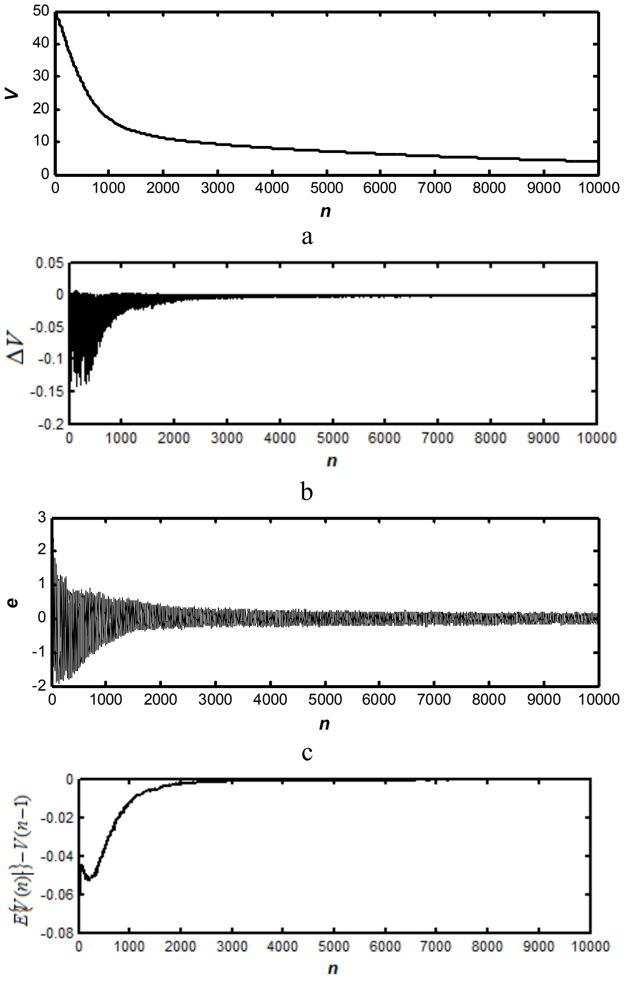


Fig. 7. The learning processes in a stochastic case: (a) the function V_n; (b) the first difference ΔV_n; (c) the estimation error $e(n)$; (d) the difference between the conditional expectation of V_n and its past value

Conclusion

In general case, the standard online gradient algorithms applied to sequential learning in neural networks with hidden layer may not converge. To guarantee their convergence, certain conditions need to be satisfied.

Simulation examples show the successful ultimate performance of the gradient learning algorithm used for identifying of nonlinear systems by means of the two-layer neural network.

References

[1]Behera, L.; Kumar, S.; Patnaik, A. "On adaptive learning rate that guarantees convergence in feedforward networks." *IEEE Trans. on Neural Networks*. no. 5(17). 2006. pp. 1116–1125.

[2]Shao, H.; Wu, W.; Liu, L. "Convergence and monotonicity of an online gradient method with penalty for neural networks." *WSEAS Trans. Math.* 2007. vol. 6. pp. 469–476.

[3]Tadic, V.; Stankovic, S. "Learning in neural networks by normalized stochastic gradient algorithm: Local convergence." *Proc. 5th Seminar Neural Netw. Appl. Electr. Eng*, Yugoslavia, Sept. 2000. pp. 11–17.

[4]Wu, W.; Feng, G.; Li, X. "Training multilayer perceptrons via minimization of ridge functions." *Advances in Comput. Mathematics*. 2002. vol. 17. pp. 331–347.

[5] Wu, W.; Feng, G.; Li, X.; Xu, Y. "Deterministic convergence of an online gradient method for BP neural networks." *IEEE Trans. on Neural Networks*. 2005. no. 3(16). pp. 1–9.

[6] Zhang, N.; Wu, W.; Zheng, G. "Convergence of gradient method with momentum for two-layer feedforward neural networks." IEEE *Trans. on Neural Networks*. 2006. no. 2(17). pp. 522–525.

[7] Finnoff, W. "Diffusion approximations for the constant learning rate backpropagation algorithm and resistance to local minima." *Neural Comput.* 1994. no. 2 (6). pp. 285– 295.

[8] 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.

[9] Mangasarian, O. L.; Solodov, M. V. "Serial and parallel backpropagation convergence via nonmonotone perturbed minimization." *Optim. Methods Software*. 1994. vol. 4. pp. 103–106.

[10] Fine, T. L.; Mukherjee, S. "Parameter convergence and learning curves for neural networks." *Neural Comput.* 1999. no. 3(11). pp. 749–769.

[11] Zhang, H.; Wu, W.; Liu, F.; Yao, M. "Boundedness and convergence of online gradient method with penalty for feedforward neural networks." *IEEE Trans. Neural Networks*. 2009. no. 6(20). pp. 1050–1054.

[12] Xu, Z.; Zhang, R.; Jing, W. "When does online BP training converge?" *IEEE Trans. Neural Networks*. 2009. no. 9(20). pp. 1529–1539.

[13] Zhiteckii, L. S., Azarskov, V. N.; Nikolaienko, S. A. "Convergence of learning algorithms in neural networks for adaptive identification of nonlinearly parameterized systems." *Proc. 16th IFAC Symposium on System Identification*, Brussels, Belgium, July 10 – 13, 2012. pp. 1593–1598.

[14] Tsypkin, Ya. Z.; Mason, J. D.; Avedyan, E. D.; Warwick, K.; Levin, I. K. "Neural networks for identification of nonlinear systems under random piecewise polynomial disturbances." *IEEE Trans. on Neural Networks*. 1999. no. 2(10). pp. 303–311.

[15] Loeve, M. "Probability Theory." *Springer-Verlag*, New York. 1963. 719 p.

Received 30 October 2013

Valerii Azarskov. Doctor of Engineering. Professor.
Aerospace Control Systems Department, National Aviation University, Kyiv, Ukraine.
Education: Kyiv Civil Aviation Engineers Institute, Kyiv, Ukraine (1968).
Research area: control theory and its application.
Publication: 250.
E-mail: azarskov@nau.edu.ua

Leonid Zhiteckii. Candidate of Engineering. Professor.
Aerospace Control Systems Department, National Aviation University, Kyiv, Ukraine.
Education: Odessa Polytechnic Institute, Odessa, Ukraine (1962).
Research area: control theory and its application.
Publication: 4 monographs.
E-mail: leonid_zhiteckii@i.ua

Sergii Nikolaienko. Postgraduate student.
Aerospace Control Systems Department, National Aviation University, Kyiv, Ukraine.
Education: Kiev National of Construction and Architecture University, Kiev, Ukraine (1998).
Research area: modeling, identification and their application.
Publication: 5.
E-mail: s_nikolaenko@ukr.net

В. М. Азарсков, Л. С. Житецький, С. А. Ніколаєнко. Процеси послідовного навчання в нейронних мережах, що застосовуються як моделі нелінійних систем

Вивчено асимптотичні властивості градієнтного алгоритму зі сталим кроковим коефіцієнтом, що використовується для навчання в реальному часі нейромережних моделей нелінійних систем з одним прихованим шаром. Встановлено деякі умови, що гарантують збіжність цього алгоритму.

Ключові слова: нелінійна система, нейромережна модель, градієнтний алгоритм, навчання, збіжність.

Валерій Миколайович Азарсков. Доктор технічних наук. Професор.
Кафедра систем управління літальних апаратів, Національний авіаційний університет, Київ, Україна.
Освіта: Київський інститут інженерів цивільної авіації, Київ, Україна (1968).
Напрям наукової діяльності: теорія управління та її прикладення.
Кількість публікацій: 250.
E-mail: azarskov@nau.edu.ua

Леонід Сергійович Житецький. Кандидат технічних наук. Професор.
Кафедра систем управління літальних апаратів, Національний авіаційний університет, Київ, Україна.
Освіта: Одеський політехнічний інститут, Одеса, Україна (1962).
Напрям наукової діяльності: теорія управління та її прикладення.
Кількість публікацій: 4 монографії.
E-mail: leonid_zhiteckii@i.ua

Сергій Анатолійович Ніколаєнко. Аспірант.
Кафедра систем управління літальних апаратів, Національний авіаційний університет, Київ, Україна.
Освіта: Київський національний університет будівництва та архітектури, Київ, Україна (1998).
Напрям наукової діяльності: моделювання, ідентифікація та їх прикладення.
Кількість публікацій: 5.
E-mail: s_nikolaenko@ukr.net

В. Н. Азарсков, Л. С. Житецкий, С. А. Николаенко. Процессы последовательного обучения в нейронных сетях, применяемых в качестве моделей нелинейных систем

Изучены асимптотические свойства градиентного алгоритма с постоянным шаговым коэффициентом, используемого для обучения в реальном времени нейросетевых моделей нелинейных систем с одним скрытым слоем. Установлены некоторые условия, гарантирующие сходимость этого алгоритма.

Ключевые слова: нелинейная система, нейросетевая модель, градиентный алгоритм, обучение, сходимость.

Азарсков Валерий Николаевич. Доктор технических наук. Профессор.
Кафедра систем управления летательных аппаратов, Национальный авиационный университет, Киев, Украина.
Образование: Киевский институт инженеров гражданской авиации, Киев, Украина.(1968).
Направление научной деятельности: системы управления.
Количество публикаций: 250.
E-mail: azarskov@nau.edu.ua

Житецкий Леонид Сергеевич. Кандидат технических наук. Профессор.
Кафедра систем управления летательных аппаратов, Национальный авиационный университет, Киев, Украина.
Образование: Одесский политехнический институт, Одесса, Украина (1962).
Направление научной деятельности: системы управления.
Количество публикаций: 4 монографии.
E-mail: leonid_zhiteckii@i.ua

Сергей Анатольевич Николаенко. Аспирант.
Кафедра систем управления летательных аппаратов, Национальный авиационный университет, Киев, Украина.
Образование: Киевский национальный университет строительства и архитектуры, Киев, Украина (1998).
Направление научной деятельности: моделирование, идентификация и их приложения.
Количество публикаций: 5.
E-mail: s_nikolaenko@ukr.net