# АСИМПТОТИКА ПРОЦЕССОВ ПОСЛЕДОВАТЕЛЬНОГО ОБУЧЕНИЯ НЕЙРОННЫХ СЕТЕЙ В СТОХАСТИЧЕСКОЙ СРЕДЕ: СПЕЦИАЛЬНЫЙ СЛУЧАЙ

**Азарсков В.Н.,**
*д. т.н., зав. кафедрой СУЛА Национального авиационного университета (Киев)*
*azarskov@nau.edu.ua*
**Житецкий Л.С.**
*к. т.н., проф. кафедры СУЛА Национального авиационного университета (Киев)*
*leonid_zhiteckii@i.ua*
**Николаенко С.А.**
*аспирант, Международный научно-учебный центр информационных технологий и систем (Киев)*
*s_nikolaenko@ukr.net*

**Аннотация.** *Изучаются асимптотические свойства стандартных градиентных алгоритмов последовательного обучения нейросетевых моделей, функционирующих в стохастической среде, для одного специального случая, когда допустима точная идентификация неизвестной нелинейности. Установлены достаточные условия сходимости этих алгоритмов.*

**Ключевые слова:** *современная теория управления, идентификация, нелинейность, градиентный алгоритм, сходимость, нейронная сеть, последовательное обучение.*

# ASYMPTOTICS OF SEQUENTIAL LEARNING PROCESSES FOR NEURAL NETWORKS IN STOCHASTIC ENVIRONMENT: SPECIAL CASE

**V.N. Azarskov**
*National Aviation University (Kiev)*
**L.S. Zhiteckii**
*National Aviation University (Kiev)*
**S.A. Nikolaienko**
*Int. Research and Training Center for Inform. Technologies & Systems (Kiev)*

**Abstract.** *Asymptotical properties of the standard gradient algorithms for sequential learning in neural network models working in the stochastic environment for a special case, when the exact identification of an unknown nonlinearity is admissible, are studied. The sufficient convergence conditions of these algorithms are established. Simulation results are given.*

**Key words:** *modern control theory, identification, nonlinearity, gradient algorithm, convergence, neural network, sequential learning.*

Известно, что нейронные сети являются универсальными аппроксиматорами довольно широкого класса нелинейностей. Важнейшей проблемой, которая возникает при построении нейросетевых моделей, заключается в том, как обеспечить сходимость процессов обучения таких моделей [1].

В последнее время в западной литературе появился ряд публикаций, посвященных анализу сходимости рекуррентных алгоритмов последовательного обучения многослойных нейросетей, ориентированных на функционирование в стохастической среде в режиме “on-line”. Основные результаты в этом направлении исследований представлены в работах

[2, 3]. Трудности, появляющиеся при установлении условий сходимости алгоритмов последовательного обучения нейросетевых моделей, состоят в том, как теоретически гарантировать ограниченность векторов весовых коэффициентов нейронов сети при неограниченной продолжительности самого процесса обучения [3].

Применительно к одному специальному случаю, упомянутому в [1, p.304], когда нейронная сеть содержит один скрытый слой и допускает идеально точную аппроксимацию нелинейности, в работе [4] авторами впервые был установлен один результат, касающийся асимптотических свойств стандартного градиентного алгоритма последовательного обучения с постоянным шагом. Установленные позже достаточные условия сходимости этого класса алгоритмов обучения нейросетей были включены в доклад, сделанный на 16-м симпозиуме ИФАК по системам идентификации (Брюссель, Бельгия, 2012 г.) [5] и обобщен в недавней работе [6].

Эффективным инструментом для исследования асимптотики процессов обучения на основе классической градиентной процедуры для достаточного общего случая представляется метод функции Ляпунова и ее стохастический аналог, получивший существенное развитие в работе [7], в которой установлены различные достаточные условия сходимости так называемых псевдоградиентных алгоритмов обучения нелинейно параметризуемых моделей. Иной подход к решению данной задачи, основанный на изучении квазифейеровских случайных последовательностей, можно найти в монографии [8].

Настоящая статья продолжает исследования асимптотических свойств градиентных алгоритмов обучения нейросетей, проведенное авторами в [4-6] в направлении установления условий их глобальной сходимости с вероятностью 1. В рамках этих исследований используются некоторые базовые результаты, полученные в свое время в [7, 8].

Рассматривается некоторая неизвестная нелинейная функция, описываемая уравнением

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

В этом уравнении $y \in \mathrm{IR}$ и $x \in \mathrm{IR}^N$ – соответственно выходная скалярная переменная и вектор входных переменных, доступные для измерения в каждый $n$-й дискретный момент времени $(n = 1, 2, \ldots)$. Следуя [4, 6], будем считать, что

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

где $F:\ \mathrm{IR}^N \to \mathrm{IR}$ – неизвестное отображение.

Для аппроксимации нелинейности (1) применим двухслойную нейронную сеть, содержащую $M$ $(M \geq 1)$ нейронов в скрытом слое. Входами каждого $j$-го нейрона этого слоя в момент времени $n$ являются компоненты вектора $x(n-1)$. Выходной сигнал в $n$-й момент времени описывается выражением

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

Здесь $x_i(n-1)$ – $i$-й компонент вектора $x(n-1)$, а $w_{ij}^{(1)}$ и $b_j^{(1)}$ – соответственно весовые коэффициенты и смещения $j$-го нейрона. Функция $\sigma(\cdot)$ представляет так называемую активационную функцию [1]. В скрытом (втором) слое находится единственный нейрон, выходы которого служат входами для выходного слоя. Выходной сигнал второго слоя $y^{(2)}(n)$ в момент времени $n$ определяется как

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

где $w_1^{(2)}, \ldots, w_M^{(2)}$ – веса этого нейрона, а $b^{(2)}$ – смещение.

Поскольку функция $\sigma(\cdot)$ естественным образом предполагается нелинейной, то из (3), (4) с учетом (2) следует, что $y^{(2)}(n)$ – нелинейная функция, зависящая от $x(n-1)$ и от $(M(N+2)+1)$-мерного вектора параметров

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

где T обозначает знак транспонирования. Чтобы подчеркнуть этот факт, определим выходной сигнал нейронной сети в форме

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

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

Введем следующее основополагающее допущение: существует, по крайней мере, один вектор $w = w^* \in \mathrm{IR}^{M(N+2)+1}$ такой, что $F(x)$ может быть точно аппроксимирована функцией $\mathrm{y}_{\mathrm{NN}}(x, w^*)$ в смысле

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

для всех $x$ из некоторого компактного множества $X \subset \mathrm{IR}^N$. Именно это допущение как раз и упоминается в [1] как идеальный (специальный) случай.

Определим теперь бесконечную обучающую последовательность пар $\{(x(n-1),\, y(n))\}_{n=1}^{\infty}$, в которых векторы $x(n-1)$ берутся из множества $X$. Далее, сформулируем алгоритм обучения для коррекции оценок вектора параметров $w(n)$ как следующую стандартную градиентную процедуру:

$$\begin{aligned} w(n) &= w(n-1) + \\ &+ \eta\; e(n, w(n-1))\ \mathrm{grad}_w \mathrm{y}_{\mathrm{NN}}(x(n-1), w(n-1)). \end{aligned} \tag{7}$$

В этом алгоритме

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

– текущая ошибка оценивания, $\mathrm{grad}_w \mathrm{y}_{\mathrm{NN}}(x(n-1), w(n-1))$ – градиент функции $\mathrm{y}_{\mathrm{NN}}(x, w)$ в точке $w = w(n-1)$, $\eta \equiv \mathrm{const} > 0$ – шаг обучения.

Задача состоит в изучении асимптотических свойств последовательности $\{w(n)\} = w(1), w(2), \ldots$ весовых векторов, порожденной рекуррентным алгоритмом (7), (8) при данном начальном $w(0)$. Более определенно, требуется установить условие сходимости этого алгоритма в специальном случае, оговоренном предположением (6).

Следуя [7], для анализа асимптотических свойств алгоритма (7), (8), т.е. его поведения при $w(n) \to \infty$ введем скалярную неотрицптельную функцию $V(w)$ вида

$$V(w) = 0 \text{ при } w = w^*, V(w) > 0 \text{ при } w \neq w^*. \tag{9}$$

При наличии одноточечного множества $W^* = \{w^*\}$ функция $V(w)$ удовлетворяющая требованию (9), обычно определяется как квадрат расстояния вектора $w$ от $w^*$:

$$V(w) = \| w^* - w \|^2 . \tag{10}$$

Здесь и далее $\|\cdot\|$ обозначает евклидову норму вектора. Согласно [7] эта функция должна быть непрерывно дифференцируема по всем компонентам вектора *w*, а ее градиент удовлетворять условию

$$\begin{aligned} &\| \mathrm{grad}\, V(w') - \mathrm{grad}\, V(w'') \| < L \| w' - w'' \| \\ &\forall w', w'' \in \mathrm{IR}^{\mathrm{M(N+2)+1}}, \end{aligned} \tag{11}$$

в котором $L$ обозначает константу Липшица; см. [7, условие Б].

Оказывается, что если нейронная сеть содержит скрытый слой, то множество $W^*$ состоит из нескольких изолированных точек $w^*$ [5, 6]. В частности, в простейшем случае, когда имеется один нейрон в скрытом слое ($N = 1$, $M = 1$), а активационная функция $\sigma(\cdot)$ описывается выражением

$$\sigma(s) = \frac{1}{1 + \exp(-s)}, \tag{12}$$

то область $W^*$ содержит две точки:

$$w^{*(1)} = [w_1^*, w_2^*, w_3^*, w_4^*]^T \text{ и}$$

$$w^{*(2)} = [-w_1^*, -w_2^*, -w_3^*, w_3^* + w_4^*]^T.$$

В случае, когда множество $W^*$ неодноточечное, функция $V(w)$ может быть выбрана согласно [7] следующим образом:

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

Анализ выражения (13) показывает, что функция $V(w)$ вида (13) в отличие от (10), к сожалению, не обладает свойством непрерывной дифференцируемости во всех точках $w$ из $\mathrm{IR}^{M(N+2)+1}$. В частности, при $N = 1$, $M = 1$, когда эта функция определяется как

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

где

$$V^{(i)}(w) = \| w^{*(i)} - w \|^2,$$

условие (11) не выполняется на границе между двумя областями в пространстве $\mathrm{IR}^4$, определяемой условием $V^{(1)}(w) = V^{(2)}(w)$. Этот досадный факт демонстрирует рис. 1, на котором показаны также некоторые окрестности $W(w^{*(1)})$ и $W(w^{*(2)})$ точек $w^{*(1)}$ и $w^{*(2)}$.

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

Поскольку $V_n \ge 0$, то условие (14), при котором $V_n$ не возрастает, достаточно для существования предела

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

где $V_\infty$ – случайная величина, зависящая от $w(0)$ и $\{x(n)\}$. Тем не менее, условие (14) в принципе не является необходимым для существования этого предела. С другой стороны, такого предела, вообще говоря, может и не быть [4].

При моделировании алгоритма обучения (7), (8) в стохастической среде сразу же обнаружилось [4], что если нейросеть содержит скрытый слой, то требование (15) монотонного убывания $V_n$ не выполняется даже в относительно малой окрестности $\| w^* - w(0) \|$ начального приближения $w(0)$ к $w^*$ Ориентируясь на

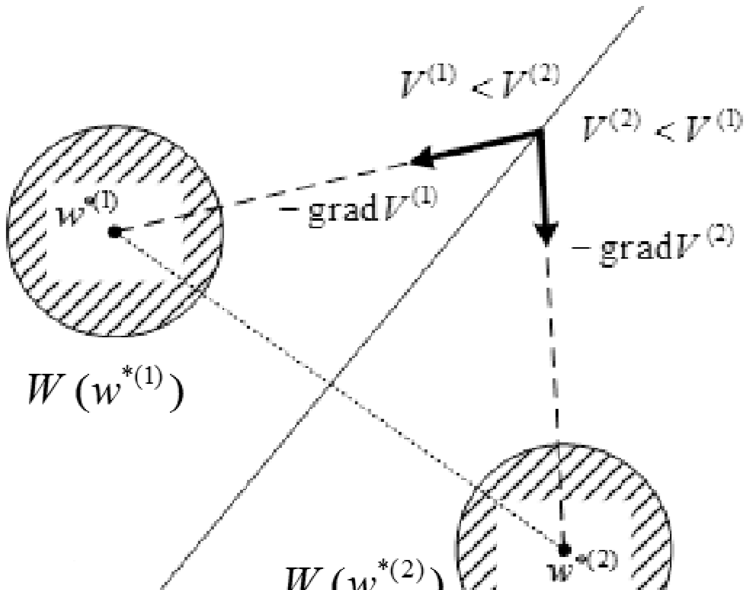


*Рис. 1. Иллюстрация свойств нейронной сети с одним скрытым слоем и N = 1.*

Переменная $V_n := V(w(n))$ немедленно становится так называемой функцией Ляпунова [5–7 ] для алгоритма (7), (8), если только

функционирование нейросети в этой среде, перейдем к стохастической версии требования (14) в форме

$$\mathrm{M}\{V_n \mid V_{n-1},\ldots,V_0\} \le V_{n-1}, \qquad (16)$$

где символ $\mathrm{M}\{V_n \mid \cdot\}$ обозначает условное математическое ожидание случайной $V_n$. Заметим, что случайная величина $V_n$, которая определяется соотношением (16), известна в теории вероятностей как супермартингал (в ранней литературе по теории вероятностей употреблялся термин «полумартингал»; см. [8, с.109]).

Пусть $W(w^{*(i)})$ обозначает некоторую окрестность вектора $w^{*(i)}$такую, что

$$V^{(i)}(w) < V^{(j)}(w)$$

для всех $w \in W(w^{*(i)}), j \ne i$ (см. рис. 1). Предположим далее, что начальная оценка $w(0)$ принадлежит $i$-ой окрестности $W(w^{*(0)})$: $w(0) \in W(w^{*(i)})$. Повторяя практически полностью доказательство теоремы 3 в работе [5], устанавливаем, что если вдоль всей траектории движения, порожденной процедурой обучения (7), (8), выполняется соотношение

$$\begin{aligned}&\mathrm{M}_x\{e(n, w(n-1))\,\mathrm{grad}_{\mathrm{w}}^{\mathrm{T}} y_{\mathrm{NN}}(w(n-1), x(n-1))(w^{*(i)} - w(n)\} \ge \\ &\ge \mathrm{M}_x\{e^2(n, w(n-1))\,\|\,\mathrm{grad}_{\mathrm{w}} y_{\mathrm{NN}}(w(n-1), x(n-1))\,\|^2\}, \end{aligned} \qquad (17)$$

в котором $\mathrm{M}_x\{\}$ – символ математического ожидания по $x$, а

$$0 < \eta < 1, \qquad (18)$$

то выполняется соотношение (16). Поскольку же $V_n \ge 0\ \forall n$, то это означает, что при выполнении (17) случайная величина $V_n$ – положительный супермартингал. По известной теореме Дуба о сходимости мартингалов [9, п.29.3] при $n \to \infty$ почти наверное (п.н.) существует некоторый конечный предел $V_\infty^{(i)}$ :

$$\lim_{n\to\infty} V_n^{(i)} = V_\infty^{(i)} < \infty \quad \text{п.н.} \qquad (19)$$

Если теперь ввести совершенно не ограничительное предположение, что $p(x) \ne 0$ на $X$ (за исключением быть может некоторых векторов $x \in X$, то по известной в теории вероятностей лемме Бореля-Кантелли [9, с.242] в соотношении (19) имеем $V_\infty^{(i)} = 0$ (п.н.), откуда

$$\lim_{n\to\infty} w(n) = w^{*(i)} \quad \text{п.н.} \qquad (20)$$

К сожалению, соотношение (17) представляется неконструктивным: его нельзя проверить до проведения самого обучения. Кроме того (и это не менеее существенно), фигурирующий в (17) вектор $w^{*(i)}$ априори неизвестен. Между тем такую проверку можно было бы заменить проверкой другого соотношения

$$\begin{aligned}&\int_X (y_{\mathrm{NN}}(w', x) - y_{\mathrm{NN}}(w'', x)) \quad \mathrm{grad}_{\mathrm{w}}^{\mathrm{T}} y_{\mathrm{NN}}(w'', x)(w' - w'')p(x)dx \ge \\ &\ge \int_X [y_{\mathrm{NN}}(w', x) - y_{\mathrm{NN}}(w'', x)]^2 \quad \|\,\mathrm{grad}_{\mathrm{w}} y_{\mathrm{NN}}(w'', x)\,\|^2\ p(x)dx, \\ &w',\ w'' \in \mathrm{IR}^{M(N+2)+1}, \end{aligned} \qquad (21)$$

приведенного в работе [6]. Но такая проверка требует, вообще говоря, перебора различных возможных векторов $w', w''$ на $\mathrm{IR}^{M(N+2)+1}$. К тому же само по себе соотношение определяет только условие, необходимое для выполнения (17).

Проведенные модельные эксперименты показали, что если начальная оценка $w(0)$ находится слишком далеко от множества $W^*$ то условие (17) не выполняется. Поэтому требование (17) вместе с требованием (18) правомерно интерпретировать как условие локальной сходимости алгоритма обучения (7), (8) с вероятностью 1.

Оказывается, что при больших начальных значениях $\| w^{*(i)} - w(0) \|$ выполняется не условие (17), приводящееся к (16), а другое более жесткое условие вида

$$\mathrm{M}\{V_n \mid V_{n-1},\ldots,V_0\} \le V_{n-1} + \chi_n, \qquad (22)$$

в котором функция $V_n$ определяется выражением (13), а переменная $\chi_n \ge 0$ обладает свойством

$$\sum_{n=1}^{\infty} \mathrm{P}\{\chi_n > 0\} \overset{\text{п.н.}}{<} \infty, \qquad (23)$$

где $\mathrm{P}\{\cdot\}$ обозначает вероятность случайного события, заключенного в фигурных скобках.

На основании упомянутой ранее леммы Бореля-Кантелли можно заключить, что при выполнении (23) с вероятностью 1 существует некоторый случайный момент $n = n^* < \infty$ такой, что

$$\chi_n = 0 \quad \forall n \geq n^*.$$

Поэтому

$$\sum_{n=1}^{\infty} \mathrm{M}\,\{\chi_n\} < \infty \quad \text{п.н.} \tag{24}$$

Согласно [8, гл. 3, теорема 3] в условиях (22), (24) справедливо такое асимптотическое свойство алгоритма обучения (7), (8):

$$V_n \xrightarrow[n\to\infty]{} V_\infty \quad \text{п.н.} \tag{25}$$

Эти условия в силу (25) как раз и являются достаточными условиями глобальной сходимости алгоритма обучения (7), (8) с вероятностью 1.

Для проверки условий сходимости данного алгоритма и демонстрации его поведения в асимптотике проводилось моделирование процесса обучения нейросети при малых и больших расстояниях начальных векторов $w(0)$ от множества $W^*$, определяемых начальными значениями функции Ляпунова

$$V_0 = V(w(0)) = \min\{\| w^{*(1)} - w(0) \|, \| w^{*(2)} - w(0) \|\}.$$

В качестве нелинейности (1) была взята функция

$$y = \frac{3{,}75 + 0{,}05\exp(-7{,}15x)}{1 + 0{,}19\exp(-7{,}15x)}.$$

Эта нелинейная функция может быть безошибочно аппроксимирована двухслойной нейронной сетью, описываемой уравнениями (3), (4), (12) с векторами весов $w^{*(1)} = [7{,}15,\ 1{,}65,\ 3{,}45,\ 0{,}3]^{\mathrm{T}}$ и $w^{*(1)} = [-7{,}15,\ -1{,}65,\ -3{,}45,\ 3{,}75]^{\mathrm{T}}$.

Во всех экспериментах шаг $\eta$ согласно (18) был взят равным $\eta = 0{,}01$. Последовательность $\{x(n)\}$ генерировалась как последовательность независимых равномерно распределенных псевдослучайных чисел в диапазоне $X = [-1{,}0,\ 1{,}0]$. Продолжительность процесса обучения не превышала 40 000 шагов.

Результаты одного из проведенных модельных экспериментов при малом $V_0$ (эксперимент №1) представлены на рис. 2. В этом эксперименте начальные оценки были выбраны такими: $w_{11}^{(1)}(0) = 1{,}4$, $b_1^{(1)}(0) = -0{,}1$, $w_1^{(2)}(0) = -0{,}56$, $b^{(2)}(0) = 0{,}46$.

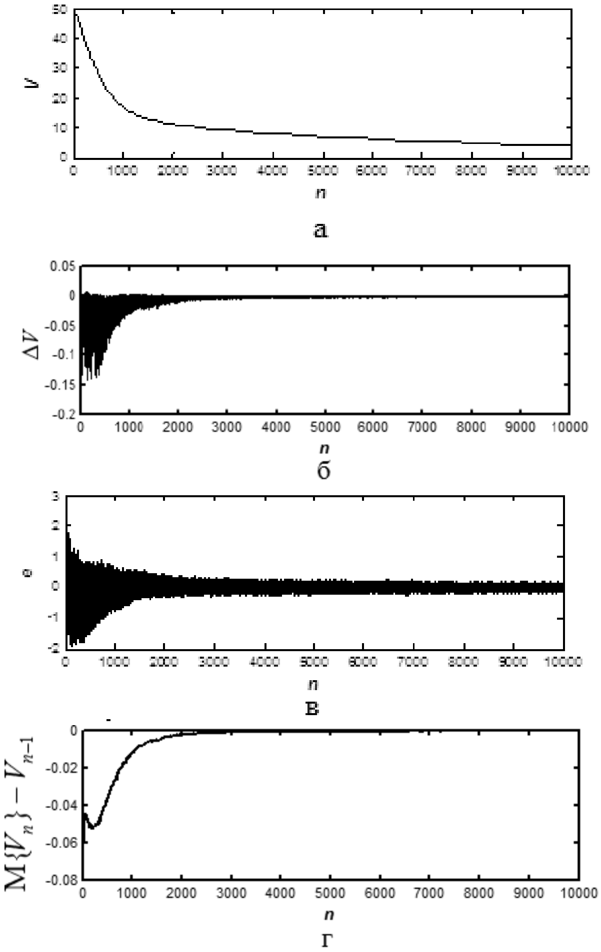


*Рис. 2. Процесс обучения нейросети в эксперименте №1: а – функция $V_n$; б – первая разность $\Delta V_n$; в – ошибка оценивания $e(n)$; г – разность между условным математическим ожиданием $\mathrm{M}\{V_n \mid \cdot\}$ и предыдущим значением $V_{n-1}$.*

Из рис. 2,а видно, что функция $V_n$ немонотонно уменьшается: ее первая разность $\Delta V_n = V_n - V_{n-1}$ изменяет свой знак (см. рис. 2,б). Тем не менее $e(n) \to 0$ с ростом $n$, как показывает рис. 2,в. Эта замечательная особенность процесса обучения следует из того факта, что удовлетворяется условие (16) локальной сходимости алгоритма. Рис. 2,г, на котором условное математическое ожидание $\mathrm{M}\{V_n \mid \cdot\}$ было оценено численно, демонстрирует указанную особенность.

Во второй серии модельных экспериментов начальные оценки $w(0)$ выбирались на сравнительно больших расстояниях от $W^*$ На рис. 3 представлены результаты моделирования алгоритма обучения, полученные при проведении одного эксперимента в этой серии (эксперимент №2). В этом эксперименте были приняты такие начальные значения компонентов вектора $w(0): w_{11}^{(1)}(0) = 0{,}53, b_1^{(1)}(0) = -0{,}5,$ $w_1^{(2)}(0) = -0{,}92, b^{(2)}(0) = 1{,}4.$

Как показывает рис. 3,а, в процессе обучения вектор $w(n)$, начиная свое движение в области $W(w^{*(1)})$, сначала переходит в область $W(w^{*(2)})$, а затем снова возвращается в область $W(w^{*(1)})$. При этом наблюдается заметное немонотонное изменение функции $V_n$ (см. рис. 3,б и в) и наглядно иллюстрируется выполнение условий (22), (23) (см. рис. 3,г и д). Из рис. 3,е видно, что в процессе обучения текущий средний квадрат ошибки

$$e_{\text{ср}}^2(n, w(n-1)) := \mathrm{M}_x\{e^2(n, w(n-1))\}$$

стремится к 0 по мере роста $n$, как и должно быть при $\gamma \in (0, 1)$.

Проведенные модельные эксперименты подтвердили, что при выполнении условий (22), (23) совместно с (18) обеспечивается глобальная сходимость алгоритма обучения (7), (8) нейросети на почти всех обучающих последовательностях $\{(x(n-1), y(n))\}_{n=1}^{\infty}$.

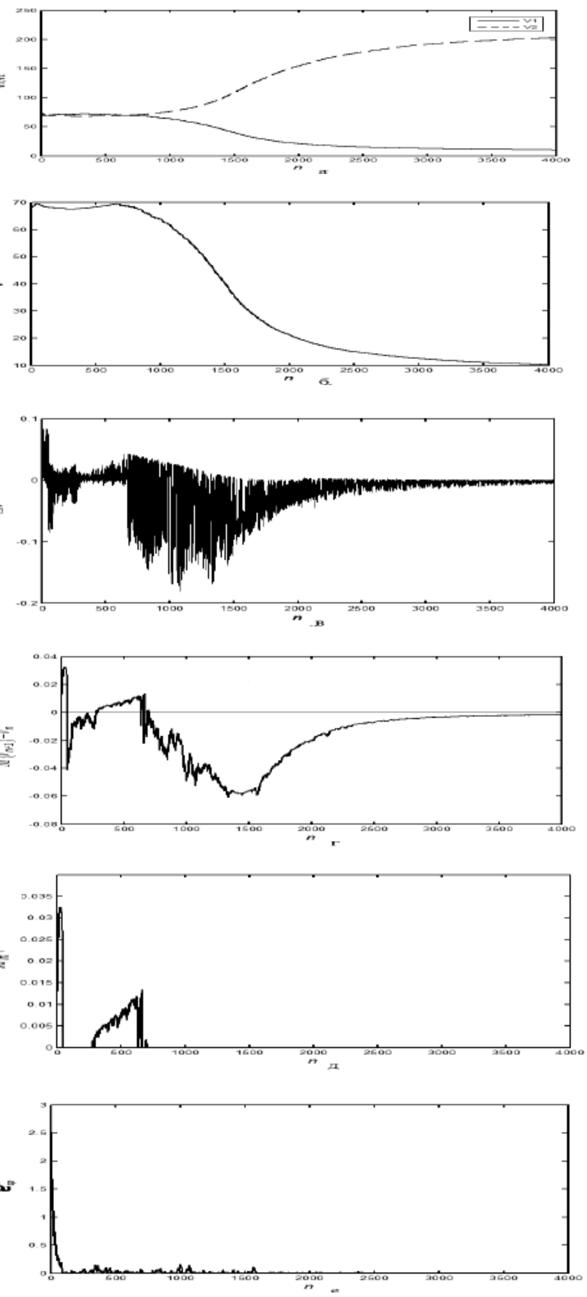

*Рис. 3. Процесс обучения нейросети в эксперименте №2:*

*а – функции $V_1, V_2$; б – функция $V_n$; в – первая разность $\Delta V_n$; г – разность между условным математическим ожиданием $\mathrm{M}\{V_n \mid \cdot\}$ и предыдущим значением $V_{n-1}$; д – переменная $\chi_n$; е – средний квадрат $e_{\text{ср}}^2(n, w(n-1))$ текущей ошибки оценивания.*

## Список литературы


1. 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. Journal “IEEE Trans. on Neural Networks”. 1999. No 2(10). P. 303–311.
2. Gaivoronski A. A. Convergence properties of backpropagation for neural nets via theory of stochastic gradient methods. Journal “Optim. Methods Software”. 1994. V. 4. P. 117–134.
3. Zhang H., Wu W., Liu F., Yao M. Boundedness and convergence of online gradient method with penalty for feedforward neural networks. Journal “IEEE Trans. Neural Networks”. 2009. No 6(20). P. 1050–1054.
4. Азарсков В.Н., Житецкий Л.С., Николаенко С.А. Достаточные условия сходимости градиентных алгоритмов последовательного обучения нейросетей: стохастический случай. Сборник научных трудов 14-й Всероссийской научно-технической конференции «Нейроинформатика-2012» (Москва, 2012). 2012. Ч. 1. с. 44 – 54.
5. 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). 2012. P. 1593–1598.
6. Azarskov V. N., Zhiteckii L. S., Nikolaienko S. A. Sequential learning processes in neural networks applied as models of nonlinear systems. Journal “Electronics and Control Systems”. 2013. No 3(37). P. 124 – 132.
7. Поляк Б.Т. Сходимость и скорость сходимости итеративных стохастических алгоритмов. I. Общий случай. Журнал «Автоматика и телемеханика». 1976. №12. с. 83–94.
8. Ермольев Ю.М. Методы стохастического программирования. М.: Наука. 1976. 239с.
9. Лоэв М. Теория вероятностей. М.: изд-во иностр. лит. 1962. 719с.