

DOI: <https://doi.org/10.5281/zenodo.16957906>

АНАЛИЗ СХОДИМОСТИ ОНЛАЙН ГРАДИЕНТНОГО МЕТОДА ДЛЯ НЕЙРОННЫХ СЕТЕЙ С МНОЖЕСТВЕННЫМИ ВЕСОВЫМИ СВЯЗАМИ

Худайбергенов Кабул Кадирбергенович^{1,2}, Марахимов Авазжон Рахимович³

¹PhD, доцент, Международный университет Кимё в Ташкенте, Ташкент, Узбекистан, kabul85@mail.ru;

²PhD, доцент, Научно-исследовательский институт развития цифровых технологий и искусственного интеллекта, kabul85@mail.ru;

³Доктор технических наук, профессор, Ташкентский государственный технический университет, Ташкент, Узбекистан, avaz.marakhimov@yandex.ru.

Аннотация. В данной статье исследуется свойство сходимости онлайн-градиентного обучения для алгоритма обратного распространения ошибки для нейронных сетей прямого распространения с двумя скрытыми слоями с множественными весовыми связями. Предполагается, что в каждом цикле обучения каждый пример обучения в наборе обучающих данных передается в стохастической форме в многослойную нейронную сеть прямого распространения с множественными весовыми связями ровно один раз. Приводиться свойство слабой и сильной сходимости для подходов к обучению, указывая, что градиент функции ошибок стремится к нулю, а весовые коэффициенты - к значению фиксированной точки соответственно. Сначала приводиться результат сходимости для подхода полностью стохастического порядка, а затем - для подхода специального стохастического порядка. Условия на функцию активации нейронной сети и скорость обучения, гарантирующие сходимость, смягчены по сравнению с существующими результатами. Свойства сходимости в данной статье изучаются для сигмоидального типа функции активации, однако эти результаты справедливы и для других типов функций активации.

Ключевые слова: нейронные сети, метод обратного распространения, онлайн градиентный метод, сходимость.

1. ВВЕДЕНИЕ

Нейронные сети (НС) в последние годы набирает еще больше актуальности в области когнитивных наук, вычислительного интеллекта и интеллектуальной обработки данных. А метод обратного распространения ошибок [1] является наиболее широко используемым методом обучения для НС прямого распространения. Существует два способа реализации алгоритма обратного распространения ошибок: подход с пакетным обучением и подход с онлайн обучением. В стандартном градиентном методе подход пакетного обучения накапливает весовую коррекцию по всем обучающим выборкам перед выполнением обновления параметров. С другой стороны, подход онлайн обучения обновляет веса сети сразу после подачи каждого примера из обучающей выборки. Некоторые авторы сравнивают эти два варианта обучения НС прямого распространения [2-4] и, а также получены некоторые

асимптотические свойства. В [5] объясняют, почему пакетное обучение почти всегда медленнее онлайн-обучения, особенно на больших обучающих наборах. В работе [5] теоретически анализирует свойства сходимости двух схем, применяемых к квадратичным функциям потерь, и показывает, в какой степени размер обучающего набора, дисперсия градиента на случай и скорость обучения влияют друг на друга.

Существует три подхода к онлайн-обучению НС в соответствии с различными методами выборки. Первый подход - ОГМ-ПСП (полностью стохастический порядок): на каждом шаге обучения из обучающего множества случайным образом выбирается одна из выборок и предъявляется сети [6]. Второй подход - ОГМ-ССП (специальный стохастический порядок): В каждом цикле обучения каждый образец из обучающего набора подается в сеть в стохастическом порядке ровно один раз [7-8]. Третий подход - ОГМ-ФП (фиксированный порядок): В каждом цикле обучения каждый образец из обучающего набора подается в сеть в фиксированном порядке ровно один раз [9-15].

Целью данной работы является всестороннее исследование слабой и сильной сходимости для ОГМ-ФП и ОГМ-ССП, которая заключается в том, что градиент функции ошибки обращается в ноль, а последовательность весов - в фиксированную точку, соответственно.

В [16] было предложено новая модель НС, которое получено на основе информационной связи между двумя нейронами. То есть два нейрона передают сообщения, высвобождая несколько типов нейротрансмиттеров, а не только один вид. А в работе [17] предложено новый тип НС с множественными весами связей (MCNN) путем расширения размеров связей между две единицы от одной к нескольким. В MCNN разные размеры весов связей соответствуют разным типам нейротрансмиттеров. В частности, вес связи между двумя модулями для существующих моделей нейронных сетей является одномерным, а вес связи между двумя модулями для новой модели - многомерным. Чтобы заставить веса связей между двумя модулями адаптивно сотрудничать и конкурировать, данной работе эвристически определяется входные и выходные данные каждого скрытого модуля в MCNN.

2. СТРУКТУРА МОДЕЛИ MCNN

Здесь описывается структура предлагаемой MCNN для задачи регрессии. Структура MCNN и ее нейроны показана на рис. 1, где в качестве примера берётся MCNN с двумерным вектором весовых соединений. В общем, модель MCNN может состоять из одного входного слоя, нескольких скрытых слоев и выходного слоя, где разрешается соединения внутри слоя и от более высокого к нижнему уровням запрещается. В MCNN также связь между двумя единицами характеризуется действительным весом. Сенсорный сигнал поступает в модель нейронной сети через входной слой и распространяется от входного слоя к выходному слою для получения прогноза. Модель трех слоистой MCNN сети с двумерными весами связей определяется следующим образом.

$$f(x_1, \dots, x_p) = \sum_{i=1}^n w_i \sigma \left(\sum_{j=1}^p (w_{i,j}^{(1)} x_j + b_j^{(1)}), \dots, \sum_{j=1}^p (w_{i,j}^{(H)} x_j + b_j^{(H)}) \right)$$

где $h=1, \dots, H$, - размерность весовых соединений в MCNN, p - размерность входного слоя, n - количество нейронов скрытого слоя. $\sigma(z_1, \dots, z_H)$ - сигмоидальная функция активация, которая определяется ниже

$$\sigma(z_1, \dots, z_H) = 1 / (1 + \sum_{h=1}^H e^{-\eta_h z_h}), \quad (9)$$

где $z_h \in \mathbf{R}$, $\eta_h \in \mathbf{R}$.

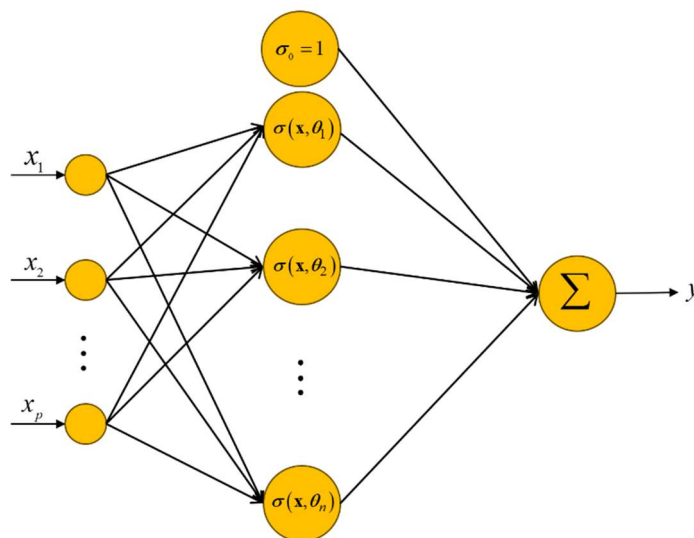


Рис. 1. Структура модели MCNN с двумерными весовыми связями ($\theta = \{w, b\}$).

3. МЕТОДЫ ОГМ-ФП И ОГМ-ССП

Начнем с введения MCNN прямого распространения с тремя слоями. Количество нейронов для входного, скрытого и выходного слоев равно p , n и 1, соответственно. Предположим, что набор обучающей выборки имеет вид $\{\mathbf{x}^j, O^j\}_{j=1}^J \subset R^p \times R$, где $\mathbf{x}^j$ и O^j являются входом и соответствующим желаемым выходом j -примера, соответственно. Пусть $\mathbf{V}^H = (v_{i,j}^{(h)})_{n \times p}$ - матрица весов ($\mathbf{V} = \{\mathbf{V}^1, \mathbf{V}^2, \dots, \mathbf{V}^H\}$) для $h=1, 2, \dots, H$, соединяющая входной и скрытый слой, и запишем как $\mathbf{v}_i = (v_{i1}^{(1)}, v_{i2}^{(2)}, \dots, v_{ip}^{(H)})^T$ для $i=1, 2, \dots, n$, H - размерность весовых соединений в MCNN. Вектор весов, соединяющий скрытый и выходной слой, обозначается как $\mathbf{u}_i = (u_1, u_2, \dots, u_n)^T \subset R^n$. Для упрощения изложения объединим матрицу весов $\mathbf{V}^H$ с вектором весов $\mathbf{u}$, и запишем как $\mathbf{w} = (\mathbf{u}^T, \mathbf{v}_1^{1,T}, \mathbf{v}_1^{2,T}, \dots, \mathbf{v}_1^{H,T}, \dots, \mathbf{v}_n^{1,T}, \mathbf{v}_n^{2,T}, \dots, \mathbf{v}_n^{H,T}) \in R^n$. Пусть $\sigma, f: R \rightarrow R$ заданы функции активации для скрытого и выходного слоев соответственно. Для удобства введем следующую векторную функцию

$$G(\mathbf{z}) = (\sigma(z_1^{(1)}, z_1^{(2)}, \dots, z_1^{(H)}), \sigma(z_2^{(1)}, z_2^{(2)}, \dots, z_2^{(H)}), \dots, \sigma(z_n^{(1)}, z_n^{(2)}, \dots, z_n^{(H)}))^T, \quad \forall \mathbf{z}_i^{(h)} \in R^n, \quad (10)$$

где $\mathbf{z} = (z_1^{(1)}, z_1^{(2)}, \dots, z_1^{(H)})$, $\sigma(\cdot)$ - сигмоидальная функция активация (9).

Для любого данного ввода $\mathbf{x} \in R^p$, выход скрытых нейронов $G(\mathbf{V}^1 \mathbf{x}, \mathbf{V}^2 \mathbf{x}, \dots, \mathbf{V}^H \mathbf{x})$, и окончательный фактический результат запишем следующим образом

$$y = f(\mathbf{u} \cdot G(\mathbf{V}^1 \mathbf{x}, \mathbf{V}^2 \mathbf{x}, \dots, \mathbf{V}^H \mathbf{x})). \quad (11)$$

Для любых фиксированных весов $\mathbf{w}$, ошибка НС определяется

$$\begin{aligned} E(\mathbf{w}) &= \frac{1}{2} \sum_{j=1}^J (O^j - f(\mathbf{u} \cdot G(\mathbf{V}^1 \mathbf{x}^j, \mathbf{V}^2 \mathbf{x}^j, \dots, \mathbf{V}^H \mathbf{x}^j)))^2 \\ &= \sum_{j=1}^J (f_j(\mathbf{u} \cdot G(\mathbf{V}^1 \mathbf{x}^j, \mathbf{V}^2 \mathbf{x}^j, \dots, \mathbf{V}^H \mathbf{x}^j))) \end{aligned} \quad (12)$$

где $f_j(t) = \frac{1}{2} (O^j - f_j(\mathbf{u} \cdot G(\mathbf{V}^1 \mathbf{x}^j, \mathbf{V}^2 \mathbf{x}^j, \dots, \mathbf{V}^H \mathbf{x}^j)))^2$, $j = 1, 2, \dots, J$, $t \in \mathbf{R}$ и $h = 1, 2, \dots, H$. Градиенты функции ошибок по отношению к $\mathbf{u}$ и $\mathbf{v}_i^{(h)}$ соответственно даются выражением

$$\begin{aligned} E_{\mathbf{u}}(\mathbf{w}) &= - \sum_{j=1}^J (O^j - y^j) f'(\mathbf{u} \cdot G(\mathbf{V}^1 \mathbf{x}^j, \mathbf{V}^2 \mathbf{x}^j, \dots, \mathbf{V}^H \mathbf{x}^j)) G(\mathbf{V}^1 \mathbf{x}^j, \mathbf{V}^2 \mathbf{x}^j, \dots, \mathbf{V}^H \mathbf{x}^j) \\ &= \sum_{j=1}^J f'_j(\mathbf{u} \cdot G(\mathbf{V}^1 \mathbf{x}^j, \mathbf{V}^2 \mathbf{x}^j, \dots, \mathbf{V}^H \mathbf{x}^j)) G(\mathbf{V}^1 \mathbf{x}^j, \mathbf{V}^2 \mathbf{x}^j, \dots, \mathbf{V}^H \mathbf{x}^j) \end{aligned} \quad (13)$$

$$\begin{aligned} E_{\mathbf{v}_i} &= - \sum_{j=1}^J (O^j - y^j) f'(\mathbf{u} \cdot G(\mathbf{V}^1 \mathbf{x}^j, \mathbf{V}^2 \mathbf{x}^j, \dots, \mathbf{V}^H \mathbf{x}^j)) u_i \sigma'(\mathbf{v}_i \cdot \mathbf{x}^j) \mathbf{x}^j \\ &= - \sum_{j=1}^J f'_j(\mathbf{u} \cdot G(\mathbf{V}^1 \mathbf{x}^j, \mathbf{V}^2 \mathbf{x}^j, \dots, \mathbf{V}^H \mathbf{x}^j)) u_i \sigma'(\mathbf{v}_i \cdot \mathbf{x}^j) \mathbf{x}^j, \end{aligned} \quad (14)$$

где

$$y^j = f(\mathbf{u} \cdot G(\mathbf{V}^1 \mathbf{x}^j, \mathbf{V}^2 \mathbf{x}^j, \dots, \mathbf{V}^H \mathbf{x}^j)), \quad i = 1, 2, \dots, n; \quad j = 1, 2, \dots, J, \quad h = 1, 2, \dots, H. \quad (15)$$

Запишем

$$E_{\mathbf{v}}(\mathbf{w}) = (E_{\mathbf{v}_1}(\mathbf{w})^T, E_{\mathbf{v}_2}(\mathbf{w})^T, \dots, E_{\mathbf{v}_n}(\mathbf{w})^T), \quad (16)$$

$$E_{\mathbf{w}}(\mathbf{w}) = (E_{\mathbf{u}}(\mathbf{w})^T, E_{\mathbf{v}}(\mathbf{w})^T). \quad (17)$$

Сначала рассмотрим случай, когда обучающие выборки подаются в сеть в фиксированном порядке (ОГМ-ФП) в процессе обучения. Следовательно, начиная с произвольного начального предположения $\mathbf{w}^0$, приступим к его итерационному уточнению по формулам

$$\mathbf{u}^{mJ+j+1} = \mathbf{u}^{mJ+j} + \Delta_j \mathbf{u}^{mJ+j}, \quad (18)$$

$$\mathbf{v}_i^{(h)mJ+j+1} = \mathbf{v}_i^{(h)mJ+j} + \Delta_j \mathbf{v}_i^{(h)mJ+j}, \quad (19)$$

где

$$\begin{aligned} \Delta_k \mathbf{u}^{mJ+j} &= \alpha_m (O^k - y^{mJ+j,k}) f'(\mathbf{u}^{mJ+j} \cdot G^{mJ+j,k}) G^{mJ+j,k} \\ &= -\alpha_m f'_k(\mathbf{u}^{mJ+j} \cdot G^{mJ+j,k}) G^{mJ+j,k}, \end{aligned} \quad (20)$$

$$\begin{aligned} \Delta_k \mathbf{v}_i^{mJ+j} &= \alpha_m (O^k - y^{mJ+j,k}) f'(\mathbf{u}^{mJ+j} \cdot G^{mJ+j,k}) \\ &\quad \times u_i^{mJ+j} \sigma'(\mathbf{v}_i^{(h)mJ+j} \cdot \mathbf{x}^k) \mathbf{x}^k \\ &= -\alpha_m f'_k(\mathbf{u}^{mJ+j} \cdot G^{mJ+j,k}) u_i^{mJ+j} \sigma'(\mathbf{v}_i^{(h)mJ+j} \cdot \mathbf{x}^k) \mathbf{x}^k, \end{aligned} \quad (21)$$

$$\begin{aligned} G^{mJ+j,k} &= G(\mathbf{V}^1 \mathbf{x}^k, \mathbf{V}^2 \mathbf{x}^k, \dots, \mathbf{V}^H \mathbf{x}^k), \quad y^{mJ+j} = f(\mathbf{u}^{mJ+j} \cdot G^{mJ+j,k}), \\ m &\in \mathbb{N}; \quad i = 1, 2, \dots, n; \quad j, k = 1, 2, \dots, J, \quad h = 1, 2, \dots, H. \end{aligned} \quad (22)$$

Здесь параметр α_m является скоростью обучения, значение которой может изменяться после каждого цикла эпохи обучения.

Также можно выбрать обучающие образцы в специальном стохастическом порядке (ОГМ-ССП) следующим образом: Для m -го обучающего эпохи, пусть

множества $\{\mathbf{x}^{m,1}, \mathbf{x}^{m,2}, \dots, \mathbf{x}^{m,J}\}$ является стохастической перестановкой исходного множества $\{\mathbf{x}^1, \mathbf{x}^2, \dots, \mathbf{x}^J\}$. Подобно (18) и (19), весовые коэффициенты обновляются итеративно следующим образом

$$\mathbf{u}^{mJ+j+1} = \mathbf{u}^{mJ+j} + \Delta_j \mathbf{u}^{mJ+j}, \quad (23)$$

$$\mathbf{v}_i^{(h)mJ+j+1} = \mathbf{v}_i^{(h)mJ+j} + \Delta_j \mathbf{v}_i^{(h)mJ+j}, \quad (24)$$

где

$$\begin{aligned} \Delta_k^m \mathbf{u}^{mJ+j} &= \alpha_m (O^k - y^{mJ+j,m,k}) f'(\mathbf{u}^{mJ+j} \cdot G^{mJ+j,m,k}) G^{mJ+j,m,k} \\ &= -\alpha_m f'_k(\mathbf{u}^{mJ+j} \cdot G^{mJ+j,m,k}) G^{mJ+j,m,k}, \end{aligned} \quad (25)$$

$$\begin{aligned} \Delta_k^m \mathbf{v}_i^{(h)mJ+j} &= \alpha_m (O^k - y^{mJ+j,m,k}) f'(\mathbf{u}^{mJ+j} \cdot G^{mJ+j,m,k}) u_i^{mJ+j} \sigma'(\mathbf{v}_i^{(h)mJ+j} \cdot \mathbf{x}^{m,k}) \mathbf{x}^{m,k} \\ &= -\alpha_m f'_k(\mathbf{u}^{mJ+j} \cdot G^{mJ+j,m,k}) u_i^{mJ+j} \sigma'(\mathbf{v}_i^{(h)mJ+j} \cdot \mathbf{x}^{m,k}) \mathbf{x}^{m,k}, \end{aligned} \quad (26)$$

$$G^{mJ+j,m,k} = G(\mathbf{V}^{1,mJ+j} \mathbf{x}^{m,k}, \mathbf{V}^{2,mJ+j} \mathbf{x}^{m,k}, \dots, \mathbf{V}^{H,mJ+j} \mathbf{x}^{m,k}),$$

$$y^{mJ+j,m,k} = f(\mathbf{u}^{mJ+j} \cdot G^{mJ+j,m,k}),$$

$$m \in \square; i = 1, 2, \dots, n; j, k = 1, 2, \dots, J.$$

(27)

Напомним, что ОГМ-ФП и ОГМ-ССП также называются циклическим обучением и почти циклическим обучением, соответственно.

4. ОСНОВНЫЕ РЕЗУЛЬТАТЫ

Для $\forall \mathbf{x} \in \square^n$ рассмотрим Евклидовую норму на $\square^n$

$$\|\mathbf{x}\| = \sqrt{\sum_{i=1}^n x_i^2}, \quad \mathbf{x} = (x_1, \dots, x_n).$$

Пусть $\Omega_0 = \{\mathbf{w} \in \Omega: E_{\mathbf{w}}(\mathbf{w}) = 0\}$, набор стационарных точек от функции ошибок $E(\mathbf{w})$, где $\Omega \in \square^{n(p+1)}$ - является ограниченной областью, удовлетворяющей (A3) ниже. Пусть $\Omega_{0,s} \in \square$ проекция от Ω_0 на s -координатную ось

$$\Omega_{0,s} = \{\mathbf{w}_s \in \square: \mathbf{w} = \{w_1, w_2, \dots, w_{n(p+1)}\}^T \in \Omega_0\}, \quad (28)$$

где $s = 1, 2, \dots, n(p+1)$. Для анализа сходимости алгоритма понадобятся следующие предположения.

(A1) $\sigma'(t)$ и $f'(t)$ Липшицево непрерывно на любом ограниченном замкнутом интервале;

(A2) $\alpha_m > 0$, $\sum_{m=0}^{\infty} \alpha_m = \infty$, $\sum_{m=0}^{\infty} \alpha_m^2 < \infty$;

(A3) Существует ограниченное открытое множество $\Omega \in \square^n$ такой, что $\{\mathbf{w}^m\} \subset \Omega$ ($m \in N$);

(A3') Существует ограниченное открытое множество $\Omega' \in \square^n$ такой, что $\{\mathbf{u}^m\} \subset \Omega'$ ($m \in N$), и производная функции активации σ в (1) равномерно ограничена и Липшицева непрерывна на $\square$.

(A4) $\Omega_{0,s}$ не содержит никакой внутренней точки для каждого $s = 1, 2, \dots, n(p+1)$.

Теорема 4.1. Предположим, что условия (A1)-(A3) выполняются (действительны). Тогда, начиная с произвольного начального значения $\mathbf{w}^0$,

обучающая последовательность $\{\mathbf{w}^m\}$, определенная формулами (18) и (19) или (23) и (24), удовлетворяет следующей слабой сходимости

$$\lim_{m \rightarrow \infty} \|E_{\mathbf{w}}(\mathbf{w}^m)\| = 0. \quad (29)$$

Более того, если предположения (A1)-(A4) верны, тогда имеет место сильная сходимость: Существует такое $\mathbf{w}^* \in \Omega_0$, что

$$\lim_{m \rightarrow \infty} \mathbf{w}^m = \mathbf{w}^*. \quad (30)$$

Сделаем три замечания о результате сходимости:

(1) Утверждаем, что слабая сходимость остается в силе, если функция активации σ скрытого слоя является широко используемой сигмоидальной функцией и предположения (A3') (вместо (A3)) и (A2) действительны. Это связано с тем, что сигмоидальная функция σ равномерно ограничена на $\mathbb{R}$ и что (46) справедливо, даже если весовые векторы $\mathbf{v}_i^{(h)} = 1, 2, \dots, n$ не ограничены.

(2) При численном анализе итерационного метода для класса нелинейных задач часто требуется ограничивать итерационную последовательность, чтобы доказать ее сходимость. Это то, что здесь делается в условиях (A3) и (A3'). Упомянем, что веса будут автоматически ограничены в обучении сети с помощью штрафного термина.

(3) Для сильной сходимости условие (A4) на Ω_0 позволяет ему быть конечным множеством, счетно бесконечным множеством, нигде не плотным множеством или даже некоторым несчетным плотным множеством. Это расслабленное состояние значительно упрощает проверку сильной сходимости на практике.

5. ДОКАЗАТЕЛЬСТВО

Для удобства изложения здесь подробно представляется доказательство сходимости для ОГМ-ФП в следующем разделе 5.1. Затем в разделе 5.2 кратко указывается, как можно применить результат на ОГМ-ССП.

5.1. Анализ сходимости для ОГМ-ФП

Сначала мы представим четыре полезные леммы для анализа сходимости.

Лемма 5.1. Пусть $\phi(x)$ - функция, определенная на ограниченном замкнутом интервале $[a, b]$, такая, что $\phi'(x)$ непрерывна по Липшицу с константой Липшица $K > 0$. Тогда $\phi'(x)$ дифференцируема почти всюду в $[a, b]$ и

$$|\phi''(x)| \leq K. \quad (31)$$

Более того, существует константа $C > 0$ такая, что

$$\phi(x) \leq \phi(x_0) + \phi'(x_0)(x - x_0) + C(x - x_0)^2, \quad \forall x_0, x \in [a, b]. \quad (32)$$

Доказательство. Поскольку $\phi'(x)$ непрерывна по Липшицу на $[a, b]$, $\phi'(x)$ абсолютно непрерывна и производная $\phi''(x)$ существует почти всюду на $[a, b]$. Следовательно, для почти любого $x \in [a, b]$,

$$\phi''(x) = \left| \lim_{h \rightarrow 0} \frac{\phi'(x+h) - \phi'(x)}{h} \right| = \lim_{h \rightarrow 0} \frac{\phi'(x+h) - \phi'(x)}{h} \leq K. \quad (33)$$

Используя интегральное разложение Тейлора, выводиться, что

$$\begin{aligned}\phi(x) &= \phi(x_0) + \phi'(x_0)(x-x_0) + (x-x_0)^2 \int_0^1 (1-t) \phi''(x_0 + t(x-x_0)) dt \\ &\leq \phi(x_0) + \phi'(x_0)(x-x_0) + (x-x_0)^2 \int_0^1 K(1-t) dt \\ &= \phi(x_0) + \phi'(x_0)(x-x_0) + C(x-x_0)^2, \\ &\left(C = \frac{K}{2}, x_0, x \in [a, b] \right). \quad \square\end{aligned}\tag{34}$$

Лемма 5.2. Предположим, что скорость обучения α_m удовлетворяет (A2) и последовательность $\{a_m\} (m \in \mathbb{N})$ удовлетворяет $a_m \geq 0$, $\sum_{m=0}^{\infty} \alpha_m a_m^\beta < \infty$ и $|a_{m+1} - a_m| \leq \mu \alpha_m$ для некоторых положительных констант β и μ . Тогда

$$\lim_{m \rightarrow \infty} a_m = 0. \tag{35}$$

Доказательство. Согласно (A2), известно, что $\alpha_m \rightarrow 0$ так как $m \rightarrow \infty$. Утверждаем, что $\lim_{k \rightarrow \infty} \inf_{m > k} a_m = 0$. В противном случае, если $a^* \equiv \lim_{k \rightarrow \infty} \inf_{m > k} a_m \in (0, \infty]$, тогда по определению нижнего предела существует целое число $M > k$ такое, что $a_m \geq \frac{a^*}{2} > 0$ для $m \geq M$, что приводит к

$$\sum_{m=0}^{\infty} \alpha_m a_m^\beta \geq \left(\frac{a^*}{2}\right)^\beta \sum_{m \geq M} \alpha_m = \infty. \tag{36}$$

Это противоречит $\sum_{m=0}^{\infty} \alpha_m a_m^\beta < \infty$ и подтверждает утверждению. Далее утверждаем, что $\lim_{k \rightarrow \infty} \sup_{m > k} a_m = 0$. В противном случае, существует $\delta \in (0, \infty]$ такой что $\lim_{k \rightarrow \infty} \sup_{m > k} a_m = \delta$. Тогда для любого $0 < \varepsilon < \delta$, можем выбрать две подпоследовательности $\{a_{i_k}\}$ и $\{a_{j_k}\}$ из $\{a_m\}$, удовлетворяющие

- (1) $a_{i_k} \in \left(0, \frac{\varepsilon}{2}\right)$, $a_{j_k} \in (\varepsilon, \delta)$;
- (2) $i_k + 1 < j_k < i_{k+1}$;
- (3) $a_{i_k} \in \left[\frac{\varepsilon}{4}, \frac{\varepsilon}{2}\right]$. (Это может быть сделано, потому что $\lim_{k \rightarrow \infty} \inf_{m > k} a_m = 0$ и $\lim_{k \rightarrow \infty} \sup_{m > k} a_m = \delta$, и $|a_{m+1} - a_m| \leq \mu \alpha_m \rightarrow 0$ так как $m \rightarrow \infty$.)

Для любого $i_k < m < j_k$, имеем $a_m \in [\frac{\varepsilon}{4}, \varepsilon]$. Таким образом, можно сделать вывод, что

$$\begin{aligned}\frac{\varepsilon}{2} &\leq |a_{j_k} - a_{i_k+1}| \leq |a_{j_k} - a_{i_k-1}| + \dots + |a_{j_k+2} - a_{i_k+1}| \\ &\leq \mu \sum_{m=i_k+1}^{j_k-1} \alpha_m \leq \mu \sum_{m=i_k+1}^{j_k} \alpha_m.\end{aligned}$$

Следовательно, для всех достаточно больших целых чисел k

$$\sum_{m=0}^{j_k} \alpha_m a_m^\beta \geq \sum_{m=i_k+1}^{j_k} \alpha_m a_m^\beta \geq \left(\frac{\varepsilon}{4}\right)^\beta \sum_{m=i_k+1}^{j_k} \alpha_m \geq \frac{2}{\mu} \left(\frac{\varepsilon}{4}\right)^{\beta+1}.$$

Но это противоречит $\sum_{m=0}^{\infty} \alpha_m a_m^\beta < \infty$ и подразумевает второе утверждение. Наконец, два приведенных выше утверждения вместе явно приводят к желаемой оценке (35).

Лемма 5.3. Пусть $\{b_m\}$ ограниченная последовательность, удовлетворяющая

$\lim_{m \rightarrow \infty} (b_{m+1} - b_m) = 0$. Запишем $\gamma_1 = \lim_{n \rightarrow \infty} \inf_{m > n} b_m$, $\gamma_2 = \lim_{n \rightarrow \infty} \sup_{m > n} b_m$ и $S = \{a \in \mathbb{R} : \text{Существует подпоследовательность } \{b_{i_k}\} \text{ из } \{b_m\} \text{ такой что } b_{i_k} \rightarrow a \text{ так как } k \rightarrow \infty\}$. Тогда имеем

$$S = [\gamma_1, \gamma_2]. \quad (37)$$

Доказательство. Очевидно что $\gamma_1 \leq \gamma_2$ и $S = [\gamma_1, \gamma_2]$. Если $\gamma_1 = \gamma_2$, тогда (37) следует просто из равенства $\lim_{m \rightarrow \infty} b_m = \gamma_1 = \gamma_2$. Рассмотрим случай $\gamma_1 < \gamma_2$ и докажем, что $S \supseteq [\gamma_1, \gamma_2]$.

Для $\forall a \in (\gamma_1, \gamma_2)$, существует $\varepsilon > 0$ такой что $(a - \varepsilon, a + \varepsilon) \subseteq (\gamma_1, \gamma_2)$. Обращая внимание, $\lim_{m \rightarrow \infty} (b_{m+1} - b_m) = 0$, можно заметить, что b_m перемещается между (лежит между) γ_1 и γ_2 с очень маленьким темпом для всего достаточно большого m . Следовательно, должно быть бесконечное количество точек последовательности b_m попадающие в $(a - \varepsilon, a + \varepsilon)$.

Отсюда следует, что $a \in S$ и следовательно, $(\gamma_1, \gamma_2) \subseteq S$. Более того, $(\gamma_1, \gamma_2) \subseteq S$ немедленно приводит к $[\gamma_1, \gamma_2] \subseteq S$. Это завершает доказательство.

Пусть последовательность $\{\mathbf{w}^{mJ+j}\}$ ($m \in \mathbb{N}$, $j = 1, 2, \dots, J$, $h = 1, 2, \dots, H$) генерируется по (18) и (19). Вводим следующие обозначения:

$$R^{m,j} = \Delta_j \mathbf{u}^{mJ+j} - \Delta_j \mathbf{u}^{mJ}, \quad (38)$$

$$r_i^{m,j} = \Delta_j \mathbf{v}_i^{mJ+j} - \Delta_j \mathbf{v}_i^{mJ}, \quad (39)$$

$$\mathbf{d}^{m,l} = \mathbf{u}^{mJ+l} + \mathbf{u}^{mJ} = \sum_{j=1}^l \Delta_j \mathbf{u}^{mJ+j} = \sum_{j=1}^l \Delta_j \mathbf{u}^{mJ} + \sum_{j=1}^l R^{m,j}, \quad (40)$$

$$\mathbf{h}_i^{m,l} = \mathbf{v}_i^{(h)mJ+l} + \mathbf{v}_i^{(h)mJ} = \sum_{j=1}^l \Delta_j \mathbf{v}_i^{(h)mJ+j} = \sum_{j=1}^l \Delta_j \mathbf{v}_i^{(h)mJ} + \sum_{j=1}^l r_i^{m,j}, \quad (41)$$

$$\boldsymbol{\psi}^{m,l,j} = \mathbf{G}^{mJ+l,j} - \mathbf{G}^{mJ,j}, \quad (42)$$

$m \in \mathbb{N}$; $j = 1, 2, \dots, J$; $l = 1, 2, \dots, J$, $i = 1, 2, \dots, n$.

Тогда (18) и (19) можно переписать как

$$\mathbf{u}^{mJ+j} = \mathbf{u}^{mJ} + \sum_{k=1}^j (\Delta_k \mathbf{u}^{mJ} + R^{m,k}), \quad (43)$$

$$\mathbf{v}_i^{(h)mJ+j} = \mathbf{v}_i^{(h)mJ} + \sum_{k=1}^j (\Delta_k \mathbf{v}_i^{(h)mJ} + r_i^{m,k}). \quad (44)$$

Пусть константы C_1 и C_2 определены по формуле (см. предположение (A3))

$$\max_{1 \leq j \leq J} \{\|\mathbf{x}^j\|, |O^j|\} = C_1, \quad \sup_{m \in \mathbb{N}} \|\mathbf{w}^m\| = C_2. \quad (45)$$

По предположению (A1), $f'_j(t)$ также удовлетворяет условию Липшица для $j = 1, 2, \dots, J$. Кроме того, $\sigma(t)$, $f(t)$ и $f_j(t)$ равномерно непрерывны на любом ограниченном замкнутом интервале.

Лемма 5.4. Пусть условия (A1) и (A3) действительны, и пусть последовательность $\{\mathbf{w}^{mJ+j}\}$ генерируется по (18) и (19). Тогда существуют константы $C_3 - C_7$ такие, что

$$\|\mathbf{G}^{mJ+j,k}\| \leq C_3, \quad (46)$$

$$\|d^{m,l}\| \leq C_4 \alpha_m, \quad \|\psi^{m,l,j}\| \leq C_5 \alpha_m, \quad (47)$$

$$\|R^{m,l}\| \leq C_6 \alpha_m^2, \quad \|r_i^{m,j}\| \leq C_7 \alpha_m^2, \quad (48)$$

где $m \in \square$; $j, k = 1, 2, \dots, J$; $l = 1, 2, \dots, J$, $i = 1, 2, \dots, n$.

Доказательство. Согласно (45) имеем

$$|\mathbf{v}_i^{(h)mJ+j} \cdot \mathbf{x}^k| \leq \|\mathbf{v}_i^{(h)mJ+j}\| \|\mathbf{x}^k\| \leq C_1 C_2 \equiv D_1. \quad (49)$$

Таким образом, существует положительная константа $C_{3,1}$ такая, что

$$\max_{|t| \leq D_1} |\sigma(t)| = C_{3,1}, \quad (50)$$

$$\|G^{mJ+j,k}\| = \|G(\mathbf{V}^{1,mJ+j} \mathbf{x}^k, \mathbf{V}^{2,mJ+j} \mathbf{x}^k, \dots, \mathbf{V}^{H,mJ+j} \mathbf{x}^k)\| \leq \sqrt{n} C_{3,1} \equiv C_3. \quad (51)$$

Из (36) и (42) следует, что

$$|\mathbf{u}^{mJ+j} \cdot G^{mJ+j,k}| \leq \|\mathbf{u}^{mJ+j}\| \|G^{mJ+j,k}\| \leq C_2 C_3 \equiv D_2. \quad (52)$$

Тогда существует положительная константа $C_{4,1}$ такая, что

$$\max_{|t| \leq D_2} |f'_j(t)| = C_{4,1}. \quad (53)$$

Кроме того, комбинация (A1), (20), (46) и (49) дает

$$\|d^{m,l}\| = \|\mathbf{u}^{mJ+l} - \mathbf{u}^{mJ}\| = \left\| \sum_{j=1}^l \Delta_j \mathbf{u}^{mJ+j} \right\| \leq C_4 \alpha_m, \quad (54)$$

где $C_4 = JC_{4,1}C_3$.

Используя (49), находим, что

$$\max_{|t| \leq D_1} |\sigma'(t)| = C_{5,1}, \quad (55)$$

где $C_{5,1}$ - положительная константа. Более того, можно заметить, что

$$\begin{aligned} \|\psi^{m,l,j}\| &= \|G^{mJ+l,j} - G^{mJ,j}\| \leq \max_{1 \leq j \leq n} |\sigma(t_i)| \|\mathbf{x}^j\| \sum_{i=1}^n \|h_i^{m,l}\| \\ &\leq \max_{1 \leq j \leq n} |\sigma'(t_i)| \|\mathbf{x}^j\| \sum_{i=1}^n \sum_{k=1}^l \|\Delta_k \mathbf{v}_i^{mJ+k}\| \leq C_5 \alpha_m, \end{aligned} \quad (56)$$

где $C_5 = n l C_{4,1} C_{5,1} \max_{1 \leq i \leq n} |\sigma'(t_i)| \|\mathbf{x}^j\| \sup_{m \in \square} \|\mathbf{w}^n\| \max_{1 \leq k \leq j} \|\mathbf{x}^k\|$, в котором

$t_i = \mathbf{v}_i^{mJ} \cdot \mathbf{x}^j + \theta_i (\mathbf{v}_i^{(h)mJ+l} - \mathbf{v}_i^{(h)mJ}) \cdot \mathbf{x}^j$, $\theta_i \in (0,1)$, и $|t_i| \leq |\mathbf{v}_i^{(h)mJ} \cdot \mathbf{x}^j| + |(\mathbf{v}_i^{(h)mJ+l} - \mathbf{v}_i^{mJ}) \cdot \mathbf{x}^j| \leq 3C_1 C_2$. В

силу (A1), видим, что $|\sigma'(t_i)|$ ($i = 1, 2, \dots, n$) ограничено.

Объединяя Липшицевую непрерывность $f'_j(t)$, (45) и (46), имеем

$$\begin{aligned} |f'_j(\mathbf{u}^{mJ+j} \cdot G^{mJ+j,j}) - f'_j(\mathbf{u}^{mJ} \cdot G^{mJ+j,j})| &\leq L |\mathbf{u}^{mJ+j} \cdot G^{mJ+j,j} - \mathbf{u}^{mJ} \cdot G^{mJ+j,j}| \\ &\leq L \|d^{m,j}\| \|G^{mJ+j,j}\| \leq L C_3 \|d^{m,j}\|, \end{aligned} \quad (57)$$

$$\begin{aligned} |f'_j(\mathbf{u}^{mJ} \cdot G^{mJ+j,j}) - f'_j(\mathbf{u}^{mJ} \cdot G^{mJ,j})| &\leq L |\mathbf{u}^{mJ} \cdot G^{mJ+j,j} - \mathbf{u}^{mJ} \cdot G^{mJ,j}| \\ &\leq L \|\mathbf{u}^{m,j}\| \|\psi^{m,j,j}\| \leq L C_2 \|\psi^{m,j,j}\|, \end{aligned} \quad (58)$$

где $L > 0$ - константа Липшица.

По определению $R^{m,j}$, получим

$$\begin{aligned} R^{m,j} &= \Delta_j \mathbf{u}^{mJ+j} - \Delta_j \mathbf{u}^{mJ} \\ &= -\alpha_m (f'_j(\mathbf{u}^{mJ+j} \cdot G^{mJ+j,j}) G^{mJ+j,j} - f'_j(\mathbf{u}^{mJ} \cdot G^{mJ,j}) G^{mJ,j}) \end{aligned}$$

$$\begin{aligned}
 &= -\alpha_m (f'_j(\mathbf{u}^{mJ+j} \cdot G^{mJ+j,j}) \boldsymbol{\psi}^{m,j,j} \\
 &+ (f'_j(\mathbf{u}^{mJ+j} \cdot G^{mJ+j,j}) - f'_j(\mathbf{u}^{mJ} \cdot G^{mJ+j,j})) G^{mJ,j} \\
 &+ (f'_j(\mathbf{u}^{mJ} \cdot G^{mJ+j,j}) - f'_j(\mathbf{u}^{mJ} \cdot G^{mJ,j})) G^{mJ,j}).
 \end{aligned} \quad (59)$$

Следовательно, из (46), (47), (57) и (58) следует, что

$$\|R^{m,j}\| \leq \alpha_m (LC_3^2 \|d^{m,l}\| + (C_{4,1} + LC_2 C_3) \|\boldsymbol{\psi}^{m,j,j}\|) \leq C_6 \alpha_m^2, \quad (60)$$

где $C_6 = \max\{LC_3^2 C_4, (C_{4,1} + LD_2) C_5\}$.

Точно так же можно показать существование константы $C_7 > 0$ такой, что

$$\|r_i^{m,j}\| \leq C_7 \alpha_m^2. \quad (61)$$

Следующая лемма раскрывает почти монотонность функции ошибок в процессе обучения.

Лемма 5.5. Пусть последовательность $\{\mathbf{w}^{mJ+j}\}$ генерируется по (18) и (19). При предположениях (A1) и (A3) выполняется

$$E(\mathbf{w}^{(m+1)J}) \leq E(\mathbf{w}^{mJ}) - \alpha_m \|E_{\mathbf{w}}(\mathbf{w}^{mJ})\|^2 + C_8 \alpha_m^2, \quad (m = 0, 1, \dots) \quad (62)$$

где $C_8 > 0$ - константа, не зависящая от m и α_m .

Доказательство. В силу предположения (A1) и Леммы 5.1 знаем, что производная

$$\boldsymbol{\sigma}''(\mathbf{v}_i^{(1)mJ} \cdot \mathbf{x}^j + t(h_i^{(1)mJ} \cdot \mathbf{x}^j), \mathbf{v}_i^{(2)mJ} \cdot \mathbf{x}^j + t(h_i^{(2)mJ} \cdot \mathbf{x}^j), \dots, \mathbf{v}_i^{(H)mJ} \cdot \mathbf{x}^j + t(h_i^{(H)mJ} \cdot \mathbf{x}^j))$$

интегрируема почти всюду на $[0, 1]$ и

$$\begin{aligned}
 &f'_j(\mathbf{u}^{mJ} \cdot G^{mJ,j}) \mathbf{u}^{mJ} \cdot \boldsymbol{\psi}^{m,J,j} \\
 &= f'_j(\mathbf{u}^{mJ} \cdot G^{mJ,j}) \sum_{i=1}^n u_i^{mJ} \boldsymbol{\sigma}'(\mathbf{v}_i^{mJ} \cdot \mathbf{x}^j) h_i^{m,J} \cdot \mathbf{x}^j \\
 &+ f'_j(\mathbf{u}^{mJ} \cdot G^{mJ,j}) \sum_{i=1}^n u_i^{mJ} (h_i^{m,J} \cdot \mathbf{x}^j)^2 \\
 &\times \int_0^1 (1-t) \boldsymbol{\sigma}''(\mathbf{v}_i^{(1)mJ} \cdot \mathbf{x}^j + t(h_i^{(1)mJ} \cdot \mathbf{x}^j), \mathbf{v}_i^{(2)mJ} \cdot \mathbf{x}^j + t(h_i^{(2)mJ} \cdot \mathbf{x}^j), \dots, \mathbf{v}_i^{(H)mJ} \cdot \mathbf{x}^j + t(h_i^{(H)mJ} \cdot \mathbf{x}^j)) dt. \quad (63)
 \end{aligned}$$

В силу Леммы 5.1, (20), (21) и (63) существует константа $C_9 > 0$ такая, что

$$\begin{aligned}
 &f_j(\mathbf{u}^{(m+1)J} \cdot G^{(m+1)J,j}) \leq f_j(\mathbf{u}^{mJ} \cdot G^{mJ,j}) + f'_j(\mathbf{u}^{mJ} \cdot G^{mJ,j}) (\mathbf{u}^{(m+1)J} \cdot G^{(m+1)J,j} - \mathbf{u}^{mJ} \cdot G^{mJ,j}) \\
 &+ C_9 (\mathbf{u}^{(m+1)J} \cdot G^{(m+1)J,j} - \mathbf{u}^{mJ} \cdot G^{mJ,j})^2 \\
 &= f_j(\mathbf{u}^{mJ} \cdot G^{mJ,j}) \\
 &+ f'_j(\mathbf{u}^{mJ} \cdot G^{mJ,j}) (d^{mJ} \cdot G^{mJ,j} + \mathbf{u}^{mJ} \cdot \boldsymbol{\psi}^{m,J,j} + d^{mJ} \cdot \boldsymbol{\psi}^{m,J,j}) \\
 &+ C_9 (\mathbf{u}^{(m+1)J} \cdot G^{(m+1)J,j} - \mathbf{u}^{mJ} \cdot G^{mJ,j})^2 \\
 &= f_j(\mathbf{u}^{mJ} \cdot G^{mJ,j}) - \frac{1}{\alpha_m} \Delta_j \mathbf{u}^{mJ} \cdot d^{m,J} - \frac{1}{\alpha_m} \sum_{i=1}^n (\Delta_j \mathbf{v}_i^{mJ} \cdot h_i^{m,J}) \\
 &+ f'_j(\mathbf{u}^{mJ} \cdot G^{mJ,j}) \sum_{i=1}^n u_i^{mJ} (h_i^{m,J} \cdot \mathbf{x}^j)^2 \cdot \int_0^1 (1-t) g''(\mathbf{v}_i^{mJ} \cdot \mathbf{x}^j + t(h_i^{mJ} \cdot \mathbf{x}^j)) dt \\
 &+ f'_j(\mathbf{u}^{mJ} \cdot G^{mJ,j}) d^{m,J} \cdot \boldsymbol{\psi}^{m,J,j} + C_9 (\mathbf{u}^{(m+1)J} \cdot G^{(m+1)J,j} - \mathbf{u}^{mJ} \cdot G^{mJ,j})^2. \quad (64)
 \end{aligned}$$

Суммируя (64) от $j=1$ до $j=J$ и упоминая (12)-(14), (40) и (41), получим

$$\begin{aligned}
 E(\mathbf{w}^{(m+1)J}) &\leq E(\mathbf{w}^{mJ}) - \frac{1}{\alpha_m} \left(\left\| \sum_{j=1}^J \Delta_j \mathbf{u}^{mJ} \right\|^2 + \sum_{i=1}^n \left\| \sum_{j=1}^J \Delta_j \mathbf{v}_i^{(h)mJ} \right\|^2 \right) + \delta_m \\
 &= E(\mathbf{w}^{mJ}) - \alpha_m \left(\left\| E_{\mathbf{u}}(\mathbf{w}^{mJ}) \right\|^2 + \sum_{i=1}^n \left\| E_{\mathbf{v}_i}(\mathbf{w}^{mJ}) \right\|^2 \right) + \delta_m \\
 &= E(\mathbf{w}^{mJ}) - \alpha_m \left\| E_{\mathbf{w}}(\mathbf{w}^{mJ}) \right\|^2 + \delta_m,
 \end{aligned}
 \tag{65}$$

где

$$\begin{aligned}
 \delta_m &= -\frac{1}{\alpha_m} \sum_{j=1}^J \Delta_j \mathbf{u}^{mJ} \cdot \sum_{j=1}^J R^{m,j} - \frac{1}{\alpha_m} \sum_{i=1}^n \left(\sum_{j=1}^J \Delta_j \mathbf{v}_i^{(h)mJ} \cdot \sum_{j=1}^J \Delta_j \mathbf{r}_i^{mJ} \right) \\
 &\quad + \sum_{j=1}^J \sum_{i=1}^n u_i^{mJ} f'_j(\mathbf{u}^{mJ} \cdot G^{mJ,j}) (h_i^{(h)mJ} \cdot \mathbf{x}^j)^2 \cdot \int_0^1 (1-t) \sigma''(\mathbf{v}_i^{(h)mJ} \cdot \mathbf{x}^j + t(h_i^{(h)mJ} \cdot \mathbf{x}^j)) dt \\
 &\quad + \sum_{j=1}^J f'_j(\mathbf{u}^{mJ} \cdot G^{mJ,j}) d^{m,j} \cdot \psi^{m,j} + C_9 (\mathbf{u}^{(m+1)J} \cdot G^{(m+1)J,j} - \mathbf{u}^{mJ} \cdot G^{mJ,j})^2.
 \end{aligned}$$

Теперь из (45) и (46) следует, что

$$\begin{aligned}
 \|G^{mJ,j}\| &= \|G(\mathbf{V}^{1,mJ} \mathbf{x}^j, \mathbf{V}^{2,mJ} \mathbf{x}^j, \dots, \mathbf{V}^{H,mJ} \mathbf{x}^j)\| \leq C_3, \\
 |\mathbf{u}^{mJ} \cdot G^{mJ,j}| &\leq \|\mathbf{u}^{mJ}\| \cdot \|G^{mJ,j}\| \leq C_2 C_3 = D_2.
 \end{aligned}
 \tag{66}$$

По формулам (20), (51) - (51) и (60) первый член δ_m можно оценить следующим образом.

$$\left\| -\frac{1}{\alpha_m} \sum_{j=1}^J \Delta_j \mathbf{u}^{mJ} \cdot \sum_{j=1}^J R^{m,j} \right\| \leq \frac{1}{\alpha_m} \sum_{j=1}^J \|\Delta_j \mathbf{u}^{mJ}\| \cdot \sum_{j=1}^J \|R^{m,j}\| \leq C_{8,1} \alpha_m^2,
 \tag{67}$$

где $C_{8,1} = J^2 C_3 C_{4,1} C_6 = J C_4 C_6$.

Аналогичные оценки для других членов δ_m могут быть получены с соответствующими константами $C_{8,t} > 0$ для $t = 2, \dots, 5$. Наконец, желаемая оценка (62) доказывается установкой $C_8 = \sum_{t=1}^5 C_{8,t}$.

Теперь можно перейти к доказательству теореме сходимости.

Доказательство теоремы 5.1 для ОГМ-ФП. Доказательство разделено на две части, касающиеся (29) и (30) соответственно.

Доказательство (20). В силу (A2) и Леммы 4.5 заключаем, что

$$\sum_{m=0}^{\infty} \alpha_m \|E_{\mathbf{w}}(\mathbf{w}^{mJ})\|^2 = \sum_{m=0}^{\infty} \alpha_m \left(\|E_{\mathbf{u}}(\mathbf{w}^{mJ})\|^2 + \|E_{\mathbf{v}}(\mathbf{w}^{mJ})\|^2 \right) < \infty,
 \tag{68}$$

$$\sum_{m=0}^{\infty} \alpha_m \|E_{\mathbf{u}}(\mathbf{w}^{mJ})\|^2 < \infty.
 \tag{69}$$

Используя интегральное разложение Тейлора, делаем вывод, что

$$\begin{aligned}
 &f'_j(\mathbf{u}^{(m+1)J} \cdot G^{(m+1)J,j}) G^{(m+1)J,j} - f'_j(\mathbf{u}^{mJ} \cdot G^{mJ,j}) G^{mJ,j} \\
 &= f'_j(\mathbf{u}^{(m+1)J} \cdot G^{(m+1)J,j}) \psi^{m,j} \\
 &\quad + \left(f'_j(\mathbf{u}^{(m+1)J} \cdot G^{(m+1)J,j}) - f'_j(\mathbf{u}^{mJ} \cdot G^{mJ,j}) \right) G^{mJ,j} \\
 &\quad + \left(f'_j(\mathbf{u}^{mJ} \cdot G^{(m+1)J,j}) - f'_j(\mathbf{u}^{mJ} \cdot G^{mJ,j}) \right) G^{mJ,j}
 \end{aligned}$$

$$\begin{aligned}
 &= f'_j \left(\mathbf{u}^{(m+1)J} \cdot G^{(m+1)J,j} \right) \psi^{m,J,j} \\
 &+ \left(d^{m,J} \cdot G^{(m+1)J,j} \right) G^{mJ,j} \cdot \int_0^1 (1-t) f'_j \left(\mathbf{u}^{mJ} \cdot G^{(m+1)J,j} + t \left(d^{m,J} \cdot G^{(m+1)J,j} \right) \right) dt \\
 &+ \left(\mathbf{u}^{mJ} \cdot \psi^{m,J,j} \right) G^{mJ,j} \cdot \int_0^1 (1-t) f''_j \left(\mathbf{u}^{mJ} \cdot G^{mJ,j} + t \left(\mathbf{u}^{mJ} \cdot \psi^{m,J,j} \right) \right) dt
 \end{aligned} \quad (70)$$

Заметим (A2) и пусть $\alpha_c > 0$ верхняя граница $\{\alpha_m\}_{m=0}^\infty$. Из (45)-(47) следует, что

$$\begin{aligned}
 &\left| \mathbf{u}^{mJ} \cdot G^{(m+1)J,j} + t \left(d^{m,J} \cdot G^{(m+1)J,j} \right) \right| \leq \left(\|\mathbf{u}^{mJ}\| + \|d^{m,J}\| \right) \|G^{(m+1)J,j}\| \\
 &\leq (C_2 + C_4 \alpha_c) C_3, \quad t \in (0,1),
 \end{aligned} \quad (71)$$

$$\begin{aligned}
 &\left| \mathbf{u}^{mJ} \cdot G^{mJ,j} + t \left(\mathbf{u}^{m,J} \cdot \psi^{m,J,j} \right) \right| \leq \|\mathbf{u}^{mJ}\| \left(\|G^{mJ,j}\| + \|\psi^{m,J,j}\| \right) \\
 &\leq C_2 (C_3 + C_5 \alpha_c) = D_2 + C_2 C_5 \alpha_c, \quad t \in (0,1).
 \end{aligned} \quad (72)$$

Согласно (71), (72) и доказательству Леммы 5.1 существуют положительные константы $C_{10}, C_{11} > 0$, такие, что

$$\left| \int_0^1 (1-t) f'_j \left(\mathbf{u}^{mJ} \cdot G^{(m+1)J,j} + t d^{m,J} \cdot G^{(m+1)J,j} \right) dt \right| \leq C_{10}, \quad (73)$$

$$\left| \int_0^1 (1-t) f''_j \left(\mathbf{u}^{mJ} \cdot G^{mJ,j} + t \mathbf{u}^{m,J} \cdot \psi^{m,J,j} \right) dt \right| \leq C_{11}. \quad (74)$$

По (52) получаем $\left| \mathbf{u}^{(m+1)J} \cdot G^{(m+1)J,j} \right| \leq C_2 C_3 = D_2$. Используя (13) и (46), (47), (53), (73) и (74), и суммирование (70) от $j=1$ до $j=J$, можно делать вывод, что

$$\begin{aligned}
 &\left\| E_u(\mathbf{w}^{(m+1)J}) - E_u(\mathbf{w}^{mJ}) \right\| \leq \left\| E_u(\mathbf{w}^{(m+1)J}) - E_u(\mathbf{w}^{mJ}) \right\| \\
 &\leq C_{10} J \max_{\substack{1 \leq j \leq J \\ m \in \mathbb{N}}} \left(\|G^{(m+1)J,j}\| \|G^{mJ,j}\| \|d^{m,J}\| \right) \\
 &+ \left(J C_{4,1} + C_{11} J \max_{\substack{1 \leq j \leq J \\ m \in \mathbb{N}}} \left(\|\mathbf{u}^{mJ}\| \|G^{mJ,j}\| \right) \right) \max_{\substack{1 \leq j \leq J \\ m \in \mathbb{N}}} \left(\|\psi^{m,J,j}\| \right) \leq C_{12} \alpha_m,
 \end{aligned} \quad (75)$$

где $C_{12} = J C_3^2 C_4 C_{10} + J C_{4,1} C_5 + J D_2 C_5 C_{11}$. Объединение (68), (75) и Лемму 5.2 приводит к $\lim_{m \rightarrow \infty} \|E_u(\mathbf{w}^{mJ})\| = 0$.

Как и в доказательстве (75), существует положительная константа C_{13} такая, что

$$\left\| E_u(\mathbf{w}^{mJ+j}) - E_u(\mathbf{w}^{mJ}) \right\| \leq C_{13} \alpha_m. \quad (76)$$

Поскольку

$$\begin{aligned}
 &\left\| E_u(\mathbf{w}^{mJ+j}) \right\| \leq \left\| E_u(\mathbf{w}^{mJ+j}) - E_u(\mathbf{w}^{mJ}) \right\| + \left\| E_u(\mathbf{w}^{mJ}) \right\| \\
 &\leq C_{13} \alpha_m + \left\| E_u(\mathbf{w}^{mJ}) \right\|,
 \end{aligned} \quad (77)$$

Получим $\lim_{m \rightarrow \infty} \|E_u(\mathbf{w}^{mJ+j})\| = 0$ для $j=1, 2, \dots, J$. Аналогичным образом, выводиться

$\lim_{m \rightarrow \infty} \|E_{v_i}(\mathbf{w}^{mJ+j})\| = 0$ для $n=1, \dots, n, j=1, 2, \dots, J$, и

$$\lim_{m \rightarrow \infty} \|E_w(\mathbf{w}^{mJ+j})\| = 0, \quad j=1, 2, \dots, J. \quad (78)$$

Это сразу дает

$$\lim_{m \rightarrow \infty} \|E_w(\mathbf{w}^m)\| = 0. \quad \square \quad (79)$$

Доказательство (21). Согласно (A3) последовательность $\{\mathbf{w}^m\} (m \in \mathbb{N})$ имеет подпоследовательность $\{\mathbf{w}^{m_k}\} (k \in \mathbb{N})$ что, сходиться, $\mathbf{w}^* \in \Omega_0$.

Из (29) и непрерывности $E_{\mathbf{w}}(\mathbf{w})$ следует, что

$$\|E_{\mathbf{w}}(\mathbf{w}^*)\| = \lim_{k \rightarrow \infty} \|E_{\mathbf{w}}(\mathbf{w}^{m_k})\| = \lim_{m \rightarrow \infty} \|E_{\mathbf{w}}(\mathbf{w}^m)\| = 0. \quad (80)$$

Это означает, что $\mathbf{w}^*$ является стационарной точкой $E(\mathbf{w})$. Следовательно, $\{\mathbf{w}^m\}$ имеет хотя бы одну точку накопления, и каждая точка накопления должна быть стационарной точкой.

Далее, доказываем, что $\{\mathbf{w}^m\}$ имеет ровно одну точку накопления. Предположим, противоположное, что $\{\mathbf{w}^m\}$ имеет как минимум две точки накопления $\bar{\mathbf{w}} \neq \tilde{\mathbf{w}}$. Запишем $\mathbf{w}^m = (w_1^m, w_2^m, \dots, w_{n(p+1)}^m)^T$. Из (18)-(21) легко видеть, что $\lim_{m \rightarrow \infty} \|\mathbf{w}^{m+1} - \mathbf{w}^m\| = 0$, или эквивалентно $\lim_{m \rightarrow \infty} |w_i^{m+1} - w_i^m| = 0$ для $i = 1, 2, \dots, n(p+1)$. Без ограничения общности будем считать, что первые компоненты $\bar{\mathbf{w}}$ и $\tilde{\mathbf{w}}$ не равны друг другу, то есть $\bar{w}_1 \neq \tilde{w}_1$. Для любого вещественного числа $\lambda \in (0, 1)$, пусть $w_1^\lambda = \lambda \bar{w}_1 + (1 - \lambda) \tilde{w}_1$. По Лемме 5.3 существует подпоследовательность $\{w_1^{m_{k_1}}\}$ из $\{w_1^m\}$ сходящийся к w_1^λ так как $k_1 \rightarrow \infty$. Из-за ограниченности $\{w_2^{m_{k_1}}\}$, существует сходящаяся подпоследовательность $\{w_2^{m_{k_2}}\} \subset \{w_2^{m_{k_1}}\}$. Определим $w_2^\lambda = \lim_{k_2 \rightarrow \infty} w_2^{m_{k_2}}$. Повторяя эту процедуру, получим убывающие подпоследовательности $\{m_{k_i}\} \supset \{m_{k_1}\} \supset \dots \supset \{m_{k_{n(p+1)}}\}$ с $w_i^\lambda = \lim_{k_i \rightarrow \infty} w_i^{m_{k_i}}$ для каждого $i = 1, 2, \dots, n(p+1)$. Запишем, $\mathbf{w}^\lambda = (w_1^\lambda, w_2^\lambda, \dots, w_{n(p+1)}^\lambda)^T$. Тогда можно считать, что $\mathbf{w}^\lambda$ является точкой накопления $\{\mathbf{w}^m\}$ для любого $\lambda \in (0, 1)$. Но это означает, что $\Omega_{0,1}$ имеет внутренние точки, что противоречит (A4). Таким образом, $\mathbf{w}^*$ должно быть единственной точкой накопления $\{\mathbf{w}^m\}_{m=0}^\infty$. Это завершает доказательство сильной сходимости.

5.2. Анализ сходимости для ОГМ-ССП

Пусть последовательность $\{\mathbf{w}^{mJ+j}\} (m \in \mathbb{N}, j = 1, 2, \dots, J)$ генерируется по (23) и (24), и пусть

$$R^{m,j} = \Delta_j^m \mathbf{u}^{mJ+j} - \Delta_j^m \mathbf{u}^{mJ}, \quad (81)$$

$$r_i^{m,j} = \Delta_j^m \mathbf{v}_i^{mJ+j} - \Delta_j^m \mathbf{v}_i^{mJ}, \quad (82)$$

$$\mathbf{d}^{m,l} = \mathbf{u}^{mJ+l} - \mathbf{u}^{mJ} = \sum_{j=1}^l \Delta_j^m \mathbf{u}^{mJ+j} = \sum_{j=1}^l \Delta_j^m \mathbf{u}^{mJ} + \sum_{j=1}^l R^{mJ+j}, \quad (83)$$

$$h_i^{(h)m,l} = \mathbf{v}_i^{(h)mJ+l} - \mathbf{v}_i^{(h)mJ} = \sum_{j=1}^l \Delta_j^m \mathbf{v}_i^{(h)mJ+j} = \sum_{j=1}^l \Delta_j^m \mathbf{v}_i^{(h)mJ} + \sum_{j=1}^l r_i^{m,j}, \quad (84)$$

$$\psi^{m,l,j} = G^{mJ+l,m,j} - G^{mJ,m,j}, \quad (85)$$

$$m \in \mathbb{N}; j = 1, 2, \dots, J; l = 1, 2, \dots, J, i = 1, 2, \dots, n, h = 1, 2, \dots, H.$$

Очевидно, что на Леммы 5.1-5.3 не влияют новые определения. Вместо Леммы 5.4 и 5.5 теперь имеем следующие две леммы.

Лемма 5.6. Пусть условия (A1) и (A3) действительны, и пусть последовательность $\{\mathbf{w}^{mJ+j}\}$ генерируется по (23) и (24). Тогда выполняются следующие оценки с теми же константами $C_3 - C_7$, что и в Лемме 5.4:

$$\|G^{mJ+j,m,k}\| \leq C_3, \quad (86)$$

$$\|d^{m,l}\| \leq C_4 \alpha_m, \quad \|\psi^{m,l,j}\| \leq C_5 \alpha_m, \quad (87)$$

$$\|R^{m,l}\| \leq C_6 \alpha_m^2, \quad \|r_i^{m,j}\| \leq C_7 \alpha_m^2, \quad (88)$$

где $m \in \mathbb{N}$; $j, k = 1, 2, \dots, J$; $l = 1, 2, \dots, J$, $i = 1, 2, \dots, n$.

Доказательство. Согласно (45) имеем

$$|\mathbf{v}_i^{(h)mJ+j} \cdot \mathbf{x}^{m,k}| \leq \|\mathbf{v}_i^{(h)mJ+j}\| \max_{1 \leq k \leq J} \|\mathbf{x}^k\| \leq C_1 C_2 \equiv D_1. \quad (89)$$

Таким образом, существует положительная константа $C_{3,1}$ такая, что

$$\max_{t \in D_1} |\sigma(t)| = C_{3,1}, \quad s \quad (90)$$

$$\|G^{mJ+j,m,k}\| = \|G(\mathbf{V}^{1,mJ+j} \mathbf{x}^{m,k}, \mathbf{V}^{2,mJ+j} \mathbf{x}^{m,k}, \dots, \mathbf{V}^{H,mJ+j} \mathbf{x}^{m,k})\| \leq \sqrt{n} C_{3,1} \equiv C_3. \quad (91)$$

Аналогично, (87) и (88) могут быть доказаны после подгонки соответствующих верхних индексов в доказательстве к Лемме 5.4.

Лемма 5.7. Пусть последовательность $\{\mathbf{w}^{mJ+j}\}$ генерируется по (23) и (24). При предположениях (A1) и (A3) выполняется

$$E(\mathbf{w}^{(m+1)J}) \leq E(\mathbf{w}^{mJ}) - \alpha_m \|E(\mathbf{w}^{mJ})\|^2 + C_8 \alpha_m^2, \quad (m = 0, 1, \dots) \quad (92)$$

где $C_8 > 0$ - та же константа, определенная в лемме 5.5.

Доказательство. Как и в доказательстве Леммы 5.6, нужно настроить только некоторые верхние индексы. Например, в соответствии с (63) меняем соответствующие верхние индексы и получаем

$$\begin{aligned} f'_j(\mathbf{u}^{mJ} \cdot G^{mJ,m,j}) \mathbf{u}^{mJ} \cdot \psi^{m,J,j} &= f'_j(\mathbf{u}^{mJ} \cdot G^{mJ,m,j}) \sum_{i=1}^n u_i^{mJ} \sigma'(\mathbf{v}_i^{mJ} \cdot \mathbf{x}^{m,j}) h_i^{m,J} \cdot \mathbf{x}^{m,j} \\ &+ f'_j(\mathbf{u}^{mJ} \cdot G^{mJ,m,j}) \sum_{i=1}^n u_i^{mJ} (h_i^{m,J} \cdot \mathbf{x}^{m,j})^2 \\ &\times \int_0^1 (1-t) \sigma''(\mathbf{v}_i^{(1)mJ} \cdot \mathbf{x}^j + t(h_i^{(1)mJ} \cdot \mathbf{x}^j), \mathbf{v}_i^{(2)mJ} \cdot \mathbf{x}^j + t(h_i^{(2)mJ} \cdot \mathbf{x}^j), \dots, \mathbf{v}_i^{(H)mJ} \cdot \mathbf{x}^j + t(h_i^{(H)mJ} \cdot \mathbf{x}^j)) dt. \end{aligned} \quad (93)$$

Доказательство теоремы 5.1 для ОГМ-ССП. Для доказательства можно использовать Леммы 5.1-5.3 и Леммы 5.6-5.7, чтобы получить результаты слабой и сильной сходимости для ОГМ-ССП точно так же, как в доказательстве Теоремы 5.1 для ОГМ-ФП.

4. ЗАКЛЮЧЕНИЕ

В этой работе было представлено всестороннее исследование слабой и сильной сходимости для трехслойной нейронной сети с множественными весовыми связями. По сравнению с существующими результатами сходимости, соответствующие предположения более смягчены. Анализ сходимости справедлив и для более обширных типов нейронных сетей.

Благодарности. Автор благодарит за поддержку гранта Республики Узбекистан на фундаментальные исследования (грант № IL-5421101773).

ЛИТЕРАТУРА

1. Baldi P., Vershynin R. The capacity of feedforward neural networks. Neural Networks, Vol. 116, 288–311 (2019).

- s2. Ismailov V.E. Approximation by neural networks with weights varying on a finite set of directions. *Journal of Mathematical Analysis and Applications*, Vol. 389, Issue 1, 72–83 (2012).
3. Chen Z., Cao F. The approximation operators with sigmoidal functions. *Computers and Mathematics with Applications*, Vol. 58, Issue 4, 758–765 (2009).
4. Pinkus A. Approximation theory of the MLP model in neural networks. *Acta Numerica*, Vol. 8, 143–195 (1999).
5. Yamashita R., Nishio M., Do R.K., Togashi K. Convolutional neural networks: an overview and application in radiology. *Insights Imaging*, Vol. 9, Issue 4, 611–629 (2018).
6. Werbos P.J. Beyond regression: new tools for prediction and analysis in the behavioral sciences. PhD thesis. Harvard University, Cambridge, MA, (1974).
7. Marakhimov A.R., Khudaybergenov K.K. A Fuzzy MLP Approach for Identification of Nonlinear Systems. *Contemporary Mathematics. Fundamental Directions*, Vol. 65, Issue 1, 44–53 (2019).
8. LeCun Y. Une procedure d'apprentissage pour reseau a seuil asymmetrique. *A la Frontieredel'Intelligence Ariticielle des Sciences de la Connaissance des Neurosciences*, Vol. 85, 599–604 (1985).
9. Rumelhart D.E., Hinton G.E., Williams R.J. Learning representations by backpropagation errors. *Nature*, Vol. 323, 533–536 (1986).
10. Wilson D.R., Martinez T. R. The general inefficiency of batch training for gradient descent learning. *Neural Networks*, Vol. 16, 1429–1451 (2003).
11. Nakama T. Theoretical analysis of batch and on-line training for gradient descent learning in neural networks. *Neurocomputing*, Vol. 73, 151–159 (2009).
12. Zhang H.S., Wu W., Liu F., Yao M.C. Boundedness and convergence of online gradient method with penalty for feedforward neural networks. *IEEE Transactions on Neural Networks*, Vol. 20, 1050–1054 (2009).
13. Bertsekas D.P., Tsitsiklis J.N. *Neuro-dynamic programming*. Athena Scientific, (1996).
14. Wu W., Feng G.R., Li X. Training multilayer perceptrons via minimization of sum of ridge functions. *Advances in Computational Mathematics*, Vol. 17, 331–347 (2002).
15. Wu W., Shao H.M., Qu D. Strong convergence of gradient methods for BP networks training. In *Proceedings of 2005 international conference on neural networks and brains*, 332–334 (2005).
16. Zhang J., Hu J., Liu J. Neural network with multiple connection weights. *Pattern Recognition*. 107 (2020) 107481. doi.org/10.1016/j.patcog.2020.107481
17. Marakhimov, A.R., Khudaybergenov, K.K., Ohundadaev, U.R., Jalelov R. A New Type of Architecture for Neural Networks with Multi-connected Weights in Classification Problems. *Lecture Notes in Networks and Systems*, 2024, 718 LNNS, pp. 105–112.