

## ОПТИМІЗАЦІЙНИЙ МЕТОД AMSGRAD В БАГАТОШАРОВИХ НЕЙРОННИХ МЕРЕЖАХ

*Сергій Андрійович Свелеба*

ORCID: <https://orcid.org/0000-0002-0823-910X>

Львівський національний університет імені Івана Франка, Львів

*Наталія Андріївна Свелеба*

ORCID: <https://orcid.org/0009-0004-2886-3921>

ПВНЗ «Європейський університет», Львівська філія, Львів

### ОГЛЯД ЛІТЕРАТУРИ

Найпоширенішим способом оптимізації нейронних мереж є метод градієнтного спуску. Градієнтний спуск - це алгоритм оптимізації, який відслідковує негативне значенням градієнта цільової функції, щоб знайти мінімум функції похибки.

Обмеження градієнтного спуску полягає в тому, що даний метод застосовує одну швидкість навчання для всіх вхідних змінних. Розширення градієнтного спуску, як-от алгоритм Adaptive Movement Estimation (Adam), використовує різну швидкість навчання для кожної вхідної змінної, але в результаті цього швидкість навчання може швидко зменшуватися до дуже малих значень [3].

Метод AMSGrad є розширеною версією метода Adam, який намагається покращити властивості конвергенції алгоритму, уникаючи великих різких змін у швидкості навчання для кожної вхідної змінної. Технічно градієнт спуску називають алгоритмом оптимізації першого порядку, оскільки він явно використовує похідну першого порядку цільової функції.

Відомо [2, 5], що алгоритм AMSGrad оновлює експоненціальні середні ковзні градієнта ( $mt$ ) і квадрата градієнта ( $vt$ ), де гіперпараметри  $\beta_1$ ,  $\beta_2$  (величина яких змінюється в інтервалі  $[0, 1)$ ) контролюють експоненціальні швидкості спаду цих середніх ковзних. Самі середні ковзні є оцінками 1-го моменту (середнє) і 2-го моменту (нецентрована дисперсія) градієнта [4].

Таким чином, AMSGrad призводить до не збільшення розміру кроку, що дозволяє уникнути проблем, яких зазнає Adam. Для простоти автори роботи [5] також видаляють крок зміщення, який використовують в Adamі. Повне оновлення AMSGrad без виправлених оцінок можна подати в наступному вигляді:

$$\begin{aligned} mt &= \beta_1(N+1)mt - 1 + (1 - \beta_1(N+1))gt \\ vt &= \beta_2vt - 1 + (1 - \beta_2)g^2t \end{aligned}$$

$$\hat{v}t + \max(\hat{v}t - 1, vt) \\ \theta t + 1 = \theta t - \eta mt / (\sqrt{vt} + \varepsilon)$$

де  $mt$  – середні ковзні градієнта функції похибки,  $vt$  – середні ковзні квадрата градієнта функції похибки,  $gt$  – градієнт функції похибки  $gt = f(x(t-1))$ ,  $f()$  – функція похибки,  $\eta = \alpha$  – швидкість навчання,  $\theta t$  – значення цільової функції на кожному кроці навчання.

Проведені нами дослідження показали, що на відміну від багатошарових нейронних мереж в яких не застосовувались оптимізаційні методи [5], в розглянутій багатошаровій нейронній мережі, при застосуванні оптимізаційного метода AMSGrad зі збільшенням швидкості навчання простежуються каскади переходу до хаотичного стану так і виходу із нього. Їх кількість збільшується зі збільшенням швидкості навчання  $\alpha$ .

Отже метою нашої роботи є проведення дослідження хаотичних станів та процесів входження та виходу із даних станів, тобто механізмів каскадного процесу навчання.

## МАТЕРІАЛИ І МЕТОДИ

Для цього в програмному середовищі Python була написана програма багатошарової нейронної мережі з прихованими шарами для розпізнавання друкованих цифр. Масив кожної цифри складався з набору «0» і «1» розміром  $4 \times 7$ . Вибірка кожної цифри містила набір з 4 можливих спотворень цифри і набір з 3 масивів які не відповідали жодній із цифр. Наприклад для цифри «0» масив значень  $x$ :

```
Numt1=[0,0,0,0,1,1,1,1,1,1,1,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0]
Numt2=[1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1]
Numt3=[0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0]
Num01=[1,1,1,1,1,0,0,1,1,0,0,1,1,0,0,1,1,0,0,1,1,0,0,1,1,1,1]
Num02=[1,1,0,1,1,0,0,1,1,0,0,1,1,0,0,1,1,0,0,1,1,0,0,1,1,1,1]
Num03=[1,1,1,1,1,0,0,1,0,0,0,1,1,0,0,1,1,0,0,1,1,0,0,1,1,1,1]
Num04=[1,1,1,1,1,0,0,1,1,0,0,1,1,0,0,1,1,0,0,1,1,0,0,1,1,1,1]
Num05=[1,1,1,1,1,0,0,1,1,0,1,1,1,0,0,1,1,0,0,1,1,0,0,1,1,1,1]
```

масив значень  $y$ :

```
Num0Y=[[0],[0],[0],[1],[1],[1],[1],[1]]
```

Описана з допомогою даної програми нейронна мережа містила 3 прихованих шарів з 28 нейронів в кожному шарі.

Вибір кількості прихованих шарів і нейронів у кожному із них визначався найменшою похибкою навчання та розпізнавання друкованих цифр. Згідно роботи [8], це трьох шарова нейронна мережа з 28

нейронами в кожному шарі. Вибір значень параметрів  $\beta_1$  і  $\beta_2$  вибирався, як в праці [4]. Функцією активації, згідно праці [1] було вибрано сігмоїду.

Реалізація даного оптимізаційного метода здійснювалась при допомозі наступного коду:

```
for i in range (num - 1):
    layer_errors.append(layer_deltas[i].dot(synapse[num - 1 - i].T))
    layer_deltas.append(layer_errors[i + 1] * sigmoid_output_to_
derivative(layers[num - 1 - i]))
    layer_deltass=layer_errors[i + 1] * sigmoid_output_to_derivative
(layers[num - 1 - i])
    # m(t) = beta1(t) * m(t-1) + (1 - beta1(t)) * g(t)
    m = beta1**(age+1) * m[i-1] + (1.0 - beta1**(age+1)) *
layer_deltass
    # v(t) = beta2 * v(t-1) + (1 - beta2) * g(t)^2
    v = (beta2 * v[i-1]) + (1.0 - beta2) * layer_deltass**2
    # vhat(t) = max(vhat(t-1), v(t))
    vhat = max(max(vhat.reshape(-1,1)), max(v.reshape(-1,1)))
    dd= m / np.sqrt(vhat+ 1e-8)
    d.append(dd)
for i in range (num):
    synapse[num - 1 - i] -= alpha * (layers[num - 1 - i].T.dot(d[i]))
```

Для аналізу функції похибки використовувалась логістична функція наступного виду:

$$x_{n+1} = \alpha - x_n - x_n^2$$

де  $n$  — крок,  $\alpha$  — параметр, який визначає швидкість навчання.

III нерухомі точки:

$$x_{1,2} = -1 \pm (\alpha + 1)^{1/2},$$

власні значення, які можна обчислити наступним чином:

$$r_1 = 1 - 2(\alpha + 1)^{1/2}$$

Вибір даного логістичного відображення обумовлено тим, що воно описує процес подвоєння частоти коливань [7]. В нашому випадку цей процес зумовлений виникненням локальних мінімумів при підході до глобального мінімуму.

## ЕКСПЕРИМЕНТ

На рис.1 наведено, результат роботи даної програми. За умови, коли  $\beta_1=0.9$ ,  $\beta_2=0.999$ , при 100 епохах на рис.1 наведено, залежність значення логістичної функції похибки від параметра  $\alpha$  та Фур'є спектри. Отримана діаграма розгалуження засвідчує, що весь досліджуваний діапазон зміни  $\alpha$  ( $0.000001 \div 0.008$ ) можна розбити на 4 частини: 1) діапазон різкого зменшення величини функції похибки ( $\alpha=0.000001 \div 0.00002$ ) – не донавчання; 2) діапазон мало змінної, монотонної поведінки функції похибки ( $\alpha=0.00002 \div 0.00025$ )–

задовільний процес навчання; 3) діапазон роздвоєння на діаграмі розгалуження ( $\alpha=0.00025\div 0.00047$ ) - процес перенавчання (Рис. 1,б); 4) діапазон хаотичної не монотонної поведінки функції похибки від  $\alpha$  ( $\alpha=0.00047\div 0.008$ ) - поява хаосу.

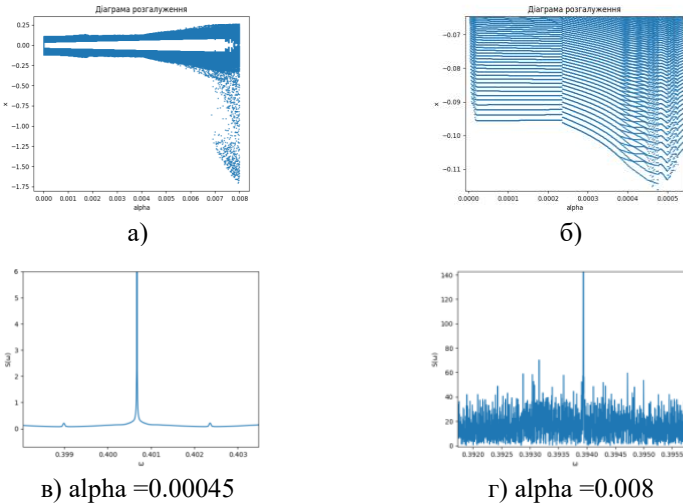


Рис. 1. Діаграму розгалуження (а), (б) (в діапазоні  $\alpha=0\div 0.00055$ ) від швидкості навчання  $\alpha$ , Фур'є спектри (в) - при перенавчанні, та (г) - в умовах хаосу, за умови 100 епох,  $\beta_1=0.9$ ,  $\beta_2=0.999$ , для цифри «0», при застосуванні оптимізаційного метода AMSGrad

Процес перенавчання супроводжується переходом через глобальний мінімум та подвоєнням кількості локальних мінімумів. Дані процеси, особливо добре проявляються при малих значеннях епох (Рис.2). Хоча і тут починає прослідковуватись і процес блокування подвоєння кількості локальних мінімумів, що унеможливує перехід системи в хаотичний стан (рис.2).

Зі збільшенням кількості епох ( $N$ ), зменшення градієнту функції похибки та збільшення величини виразу  $(1 - (\beta_1)^N)$ , за умови малих значеннях  $(1 - \beta_2)$ , зумовлює те, що градієнти стають прорідженими, оскільки співвідношення векторів першого і другого моментів утворюють блочну структуру, яка спеціалізуються на певних шаблонах. Це яскраво спостерігається при малих значеннях епох ( $N=5$  при  $\alpha > 0.0017$ ; (Рис. 2)). Отже процес оптимізації проявляє себе при збільшенні кількості епох, приводячи до зменшення градієнту при наближенні до глобального мінімуму.

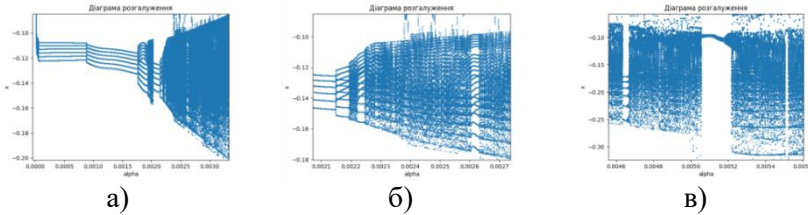


Рис. 2. Діаграму розгалуження від швидкості навчання alpha у збільшеному варіанті, (а),(б), (в), за умови 5 епохах, beta1=0.9, beta2= 0.999, для цифри «0», при застосуванні оптимізаційного метода AMSGrad

Для аналізу блочної структури на діаграмах розгалуження, розглянемо поведінку функції похибки навчання від швидкості навчання при 10 епохах. На Рис. 3, а наведено діаграму розгалуження за умови 10 епох, beta1=0.9, beta2= 0.999, для цифри «0», та в збільшеному варіанті інтервалів існування блочної структури (Рис.3, б - й). Весь діапазон перенавчання, на діаграмі розгалуження, характеризується розбиттям на блоки (Рис.3, б). Розглянемо більш детально кожний інтервал існування блочної структури і проведемо для них аналіз Фур'є спектрів.

Слід зазначити, що метод оптимізації AMSGrad базується на аналізі цільової функції на основі значень розрахованих на попередньому кроці. Тому Фур'є спектри функції похибки розраховувались для незначного інтервалу зміни швидкості навчання в кожному блоці. Перший інтервал, це діапазон  $\alpha = 0.0002 \div 0.00075$  задовільного процесу навчання. При цьому Фур'є спектри засвідчують відсутність гармоніки. На інтервалі  $\alpha = 0.00175 \div 0.018$  простежується процес перенавчання нейронної мережі з появою роздвоєння на діаграмі розгалуження (Рис. 3, в) і другої гармоніки у Фур'є спектрах (Рис.4,а). Рухаючись далі по діаграмі розгалуження в наслідок процесу подвоєння ми попадаємо в інтервал виникнення хаотичного стану ( $\alpha = 0.001825 \div 0.01914$ ) (Рис. 3, г). Фур'є спектри хаотичного стану, згідно Рис.4,в, засвідчують, що в наслідок подвоєння кількості локальних мінімумів і не одноразового проходження через глобальний мінімум виникає хаотичний стан. Даний інтервал існування хаотичного стану характеризується областями «прозорості», де є відсутній хаотичний стан нейронної мережі. Дані області характеризуються відсутністю існування хаотичного стану, а отже і зменшенням кількості локальних мінімумів. Тобто виникаючий хаотичний стан характеризується існуванням блочної структури, яка на думку авторів складається

з областей із сталих значень кількості локальних мінімумів, а усереднене значення по даному блоку дає хаотичне значення. На користь даного припущення виступає той факт, що розглянута система є динамічною, оскільки даний метод оптимізації працює лише за умови коли наступне значення залежить від попереднього. Вхід в область хаотичного стану проходить через процес подвоєння, а поява інтервалів відсутності хаотичних розв'язків в області хаосу супроводжується різким переходом від хаосу до співвірності і навпаки.

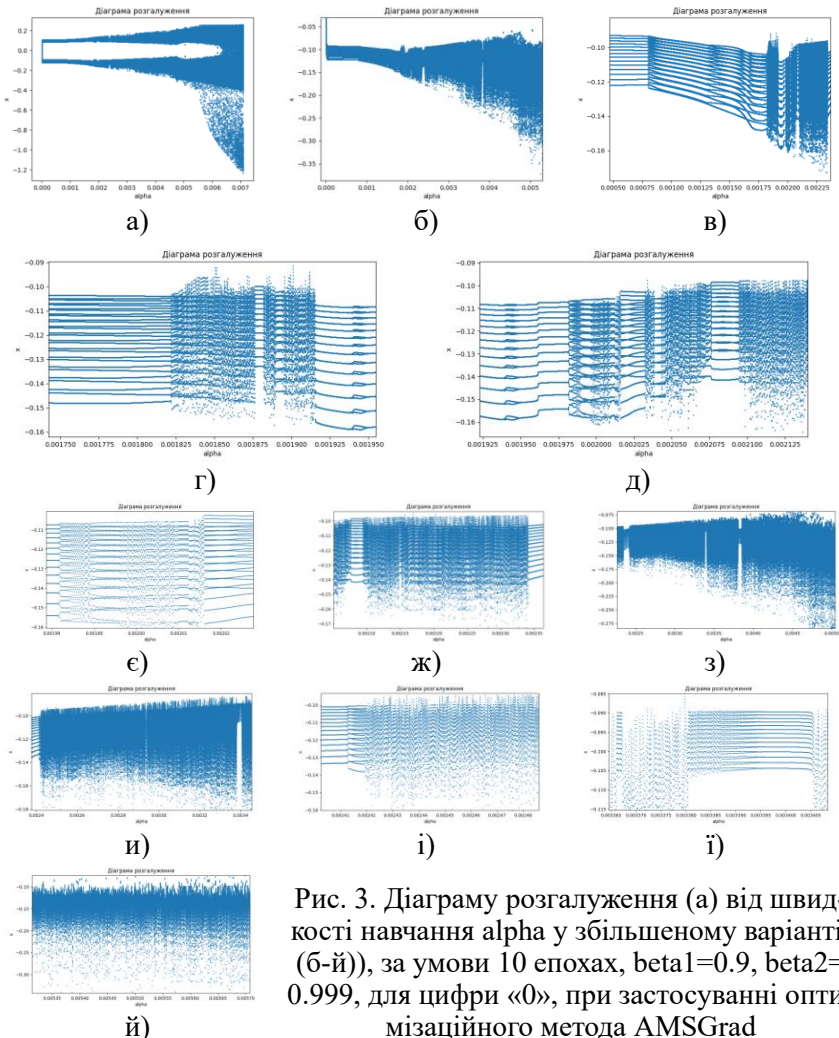


Рис. 3. Діаграму розгалуження (а) від швидкості навчання  $\alpha$  у збільшеному варіанті, (б-й)), за умови 10 епох,  $\beta_1=0.9$ ,  $\beta_2=0.999$ , для цифри «0», при застосуванні оптимізаційного методу AMSGrad

Переходячи до наступної області зміни  $\alpha = 0.001925 \div 0.001975$ , нейронна мережа характеризується відсутністю хаотичного стану (Рис. 3, г, д і Рис. 4, в). В даній області зміни  $\alpha$  Фур'є спектри цільової функції характеризуються присутністю гармоніки (Рис.4,в), а отже процес навчання нейронної мережі супроводжується перенавчанням. Згідно діаграми розгалуження даний стан нейронної мережі в цій області не є хаотичним, але характеризується процесами зародження розгалуження та його зникнення (Рис. 3, г, д). Тому Фур'є спектри цільової функції в даній області зміни швидкості навчання характеризуються появою гармоніки (Рис .4, в), яка за величиною потужності сигналу, є на багато меншою, ніж для області зміни  $\alpha$  де простежується тільки перенавчання даної мережі (Рис. 3, в).

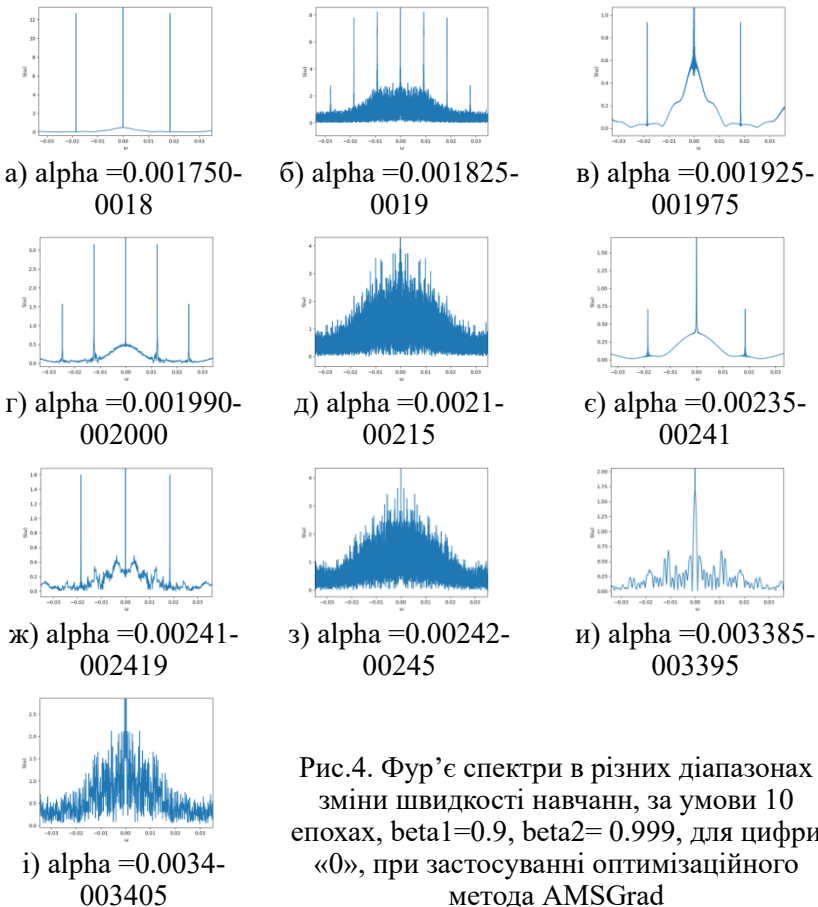


Рис.4. Фур'є спектри в різних діапазонах зміни швидкості навчання, за умови 10 епохах,  $\beta_1=0.9$ ,  $\beta_2= 0.999$ , для цифри «0», при застосуванні оптимізаційного метода AMSGrad

Наступною областю зміни  $\alpha$  є область каскадних перетворень ( $\alpha = 0.001990-0.002000$ ) наведених на Рис.3,є. Діаграма розгалуження засвідчує кілька переходів у хаотичний стан, та виходу із нього. За даних умов Фур'є спектри характеризуються появою вищих гармонік і відсутністю хаотичного стану (Рис.4,г). Для нерівноважних систем, яким притаманний хаотичний стан, відомий такий стан, який виникає при їх зародженні, і характеризується як недорозвинутим хаотичним станом. Такий стан виникає за умови коли дисипативна система характеризується втратою енергії в процесі свого розвитку. Тобто початкова область в фазовому просторі зменшується, і, нарешті, траєкторії від всіх початкових умов, узятих в певній області (басейні тяжіння), сходяться до деякого встановленого в фазовому просторі атрактора [9]. Зазвичай дисипативна динамічна система володіє тільки одним або декількома такими атракторів при певних значеннях параметрів. Відомо, що кількість атракторів зростає зі зменшенням дисипації, але хаотичних між ними мало. Це явище сильно пов'язано з кризами, які атрактор відчуває при зміні параметрів, і викликає зіткненням з межами басейну при зміні параметра нелінійності. В результаті хаотичний атрактор зникає відразу після його появи з каскаду подвоєння періоду. Зміни також відбуваються з басейнами тяжіння і межами басейнів. Перші стають досить малими для більшості співіснуючих атракторів. Тобто можна припустити що в даному інтервалі зміни  $\alpha$  може виникати стан недорозвинутого хаосу, який визначається співвідношенням оновленого експоненціального ковзного середнього градієнта ( $m_1$ ) і квадрата градієнта ( $v_1$ ). До розгляду цього нерозвиненого хаосу вернемося трохи нижче.

Переходячи до наступного інтервалу зміни  $\alpha = 0.00210 \div 0.00234$ , на діаграмі розгалуження простежується перехід до хаотичного стану через процес подвоєння кількості локальних мінімумів та неодноразового переходу через глобальний мінімум (Рис.3,ж). За даних умов, Фур'є спектри засвідчують існування хаотичного стану, однак даний хаотичний стан є відмінний від хаотичного стану який існує при  $\alpha = 0,001825 \div 0,01914$ . Ця відмінність полягає в тому, що при  $\alpha = 0,001825 \div 0,01914$  Фур'є спектри засвідчують співіснування хаотичного стану зі станом, якому притаманне існування лише кількох гармонік. Отже інтервал зміни  $\alpha = 0,001825 \div 0,01914$  можливо характеризується існуванням блочної структури.

Наступний інтервал  $\alpha = 0.00235 \div 0.00241$  (Рис. 3 ж, і) характеризується відсутністю подвоєння на діаграмі розгалуження. Фур'є спектри в даному діапазоні теж засвідчують практично відсутність гармонік (Рис. 4, є). Отже в даному інтервалі зміни  $\alpha$  ( $\alpha$ )



$=0.00235 \div 0.00241$ ) навчання нейромережі не характеризується перенавчанням, тобто проходить задовільний процес навчання.

Наближаючись до наступного інтервалу існування хаотичного стану ( $\alpha = 0.00241 \div 0.002419$ ), на діаграмі розгалуження спостерігається розгалуження (Рис.3, і) з появою гармоніки у Фур'є спектрах (Рис.4, ж).

Подальше збільшення швидкості навчання приводить до виникнення хаотичного стану (Рис. 3, і, ї) в діапазоні  $\alpha = 0.00242 \div 0.00338$ . Фур'є спектри, при даних значеннях  $\alpha$  також засвідчують існування хаотичного стану (Рис. 4, з,  $\alpha = 0.00242-0.00245$ ). Даний хаотичний стан нічим не відрізняється від інших хаотичних станів які виникають при  $\alpha = 0.001825-0.0019$ ,  $\alpha = 0.0021-0.00215$  (Рис.4, б, д). З подальшим збільшенням швидкості навчання, внаслідок зменшення кількості подвоєнь (згідно діаграми розгалуження Рис. 3, ї,  $\alpha = 0.003385-0.003395$ ), можливо завдяки зменшенню кількості локальних мінімумів простежується відсутність хаотичного стану. Фур'є спектрам (Рис. 4, и) даного інтервалу зміни  $\alpha$  притаманна незначна кількість гармонік, які по величині є в кілька разів менша ніж в попередніх вікнах прозорості. Подальше збільшення швидкості навчання супроводжується переходом до хаотичного стану (Рис. 3, й), який притаманний більшому діапазону зміни  $\alpha > 0.004$  (Рис. 3, з). Цікавою особливістю даного хаосу, є те, що його Фур'є спектрам не притаманне існування  $n$ - кратних гармонік. Згідно Рис. 4, і при  $\alpha = 0.0034-0.003405$  Фур'є спектр даного хаотичного стану характеризується меншою кількістю гармонік, що засвідчує існування механізму блокування виникнення локальних мінімумів.

Оскільки розглянутий метод оптимізації AMSGrad, як і Adagrad використовує різну швидкість навчання для кожного параметра  $\theta_i$  на кожному кроці [4], то ми спочатку одержимо оновлення AMSGrad для кожного параметра, яке потім векторизуємо. Для стислості ми використовуємо  $g_t$  для позначення градієнта на кроці часу  $t$ .  $g_{t,i}$  є частковою похідною цільової функції параметра  $\theta_i$  на кроці часу  $t$ . Тобто у своєму правилі оновлення AMSGrad змінює загальну швидкість навчання  $\alpha$  на кожному кроці часу  $t$  для кожного параметра  $\theta_i$  на основі попередніх градієнтів, які були обчислені для  $\theta_{i-1}$ . Отже функцію похибки, на кожному кроці слід розглядати, як набір похибок навчання на кожному нейроні. Похибка навчання на окремо взятому нейроні описується своєю функціональною залежністю, тому похибка навчання при даній швидкості і даному кроці є симбіоз від всіх нейронів, які задіяні в процесі навчання. Оскільки похибка навчання розраховується, як середньоквадратичне відхилення, то дані функціональні залежності похибки навчання на кожному нейроні є в першому

наближенні періодичними функціями з різними періодами. Тобто дані функціональні залежності характеризуються спектром значень хвильових векторів. Дані вектори характеризуються раціональними значеннями (тобто співмірними коливаннями, які визначаються кількістю нейронів в шарі і кількістю прихованих шарів). Сумарне значення хвильового вектора по такому ансамблю періодичностей може приймати, як співмірне так і неспівмірне значення.

Відомо, що в багатошарових нейромережах на корекцію ваг одного нейрона впливають всі нейрони попереднього шару. Оскільки при наближенні до глобального мінімуму цей вплив від швидкості навчання стає не однорідний, що може приводить до його ослаблення, а отже до не однорідного навчання. Ця неоднорідність повинна більше себе проявити при використанні оптимізаційних методів, які базуються на використанні алгоритму оновлює експоненціальних ковзних середніх градієнтів ( $m_t$ ) і квадрата градієнтів ( $v_t$ ) базуючись на попередніх значеннях. Тобто для таких багатошарових нейронних мереж повинна бути притаманна поява блочної структури, яка характеризується співіснуванням, як хаотичного стану так і стану, що характеризується декількома гармоніками. Виходячи із вище наведених результатів (Рис.3) даний стан можливо існує в інтервалі  $\alpha = 0.001982 - 0.001990$ .

## ОБГОВОРЕННЯ

Отже застосування оптимізаційного методу AMSGrad до процесу навчання багатошарової нейронної мережі приводить до виникнення блочної структури. Дана блочна структура характеризує стахостичним процес навчання нейронної мережі. А саме в певному діапазоні зміни швидкості навчання спостерігається перенавчання нейронної мережі, що супроводжується виникненням локальних мінімумів. Збільшення кількості локальних мінімумів приводить до виникнення гармонік вищого порядку. Зростання кількості подвоєн існуючих гармонік спричиняє перехід системи до хаотичного стану. Це виникає за умови, що розглянутий метод навчання нейронної мережі пов'язаний з корекцією ваг кожного нейрона при даній епосі, і на дану корекцію ваг впливають всі нейрони з попереднього шару. Отже в даному випадку функцію похибки навчання на кожному нейрані слід розглядати, як функціональну залежність яка описується набором хвильових векторів від різних періодичностей. В цьому випадку середнє значення похибки навчання по всіх нейронах можна розглядати, як усередненне значення по всіх існуючих періодичностей. А значення хвильового вектора сумарного коливання може приймати, як співмірне

так і неспівмірне значення. Так, як алгоритм AMSGrad оновлює експоненціальні середні ковзні градієнта ( $m_t$ ) і квадрата градієнта ( $v_t$ ), а гіперпараметри  $\beta_1$ ,  $\beta_2$  (величина яких змінюється в інтервалі  $[0, 1)$ ) контролюють експоненціальні швидкості спаду цих середніх ковзних, то даний метод приводить до зміни спектру частот існуючих періодичностей функціональних залежностей похибки навчання кожного нейрона. Тобто проходить корекції швидкості навчання кожного нейрона, яка знімає виродженість даної системи шляхом запобігання процесам перенавчання нейронної системи.

## ПОСИЛАННЯ

1. Bordes A., Bottou L., Gallinari P. SGD-QN: Careful quasi-Newton stochastic gradient descent. *Journal of Machine Learning Research*. 2009. Vol. 10. P. 1737-1754. URL: <https://www.jmlr.org/papers/volume10/bordes09a/bordes09a.pdf>.
2. Olenych Y. et al. Features of deep study neural network. *OPEN-REVIEWHUB*. URL: <https://openreviewhub.org/lea/paper-2019/features-deep-study-neural-network#>.
3. Rudenko O., Bodianskyu E. Artificial neural networks. Kharkiv, Ukraine: SMIT Company, 2006. (Ukrainian).
4. Ruder S. An overview of gradient descent optimization algorithms. *arXiv preprint arXiv:1609.04747*. 2016. URL: <https://ruder.io/optimizing-gradient-descent/index.html#adamax>.
5. Subotin S. Neural networks: theory and practice. Zhytomyr, Ukraine: Publisher O. O. Evenok, 2020. URL: <http://eir.zp.edu.ua/handle/123456789/6800> (Ukrainian).
6. Sveleba S. et al. Хаотичні стани багатошарової нейронної мережі. *Збірник наукових праць "Електроніка та інформаційні технології"*. 2021. №. 16. С. 20-35. <http://dx.doi.org/10.30970/eli.16.3>.
7. Sveleba, S. et al. Multilayer neural networks – as determined systems. *Computational Problems of Electrical Engineering*. 2021. Vol. 11. №2, P. 26–31. <https://doi.org/10.23939/jcpee2021.02.026>.
8. Taranenko Yu. Information entropy of chaos. URL: <https://habr.com/ru/post/447874/>.