АНОТАЦІЯ

Смородін А.В. Методи навчання нейронних мереж на основі нелінійної динаміки. - Кваліфікаційна наукова праця на правах рукопису.

Дисертація на здобуття наукового ступеня доктора філософії за спеціальністю 122 – Комп'ютерні науки. – Державний університет «Одеська політехніка» МОН України, Одеса, 2021.

У **вступі** обґрунтовано актуальність удосконалення методів навчання нейронних мереж при вирішенні задач класифікації та сегментації об'єктів. Показано можливість скорочення часу навчання нейронних мереж на основі застосування положень нелінійної динаміки при навчанні з використанням алгоритмів градієнтного спуску. Визначено об'єкт, предмет, задачі і методи дослідження; показано зв'язок з науковими програмами та планами; наведено наукову новизну та практичне значення одержаних результатів; висвітлено особистий внесок здобувача.

В **першому розділі** дисертаційної роботи проведено аналіз задач сегментації та класифікації об'єктів на зображеннях із застосуванням нейронних мереж. Показано, що, незважаючи на ефективність існуючих підходів їх вирішення, як правило, побудованих на глибоких нейронних мережах (ГНМ) складної архітектури, існуючі тенденції великих даних (великі обсяги, різнорідність, мінливість) призводять до закономірної проблеми – збільшення обчислювальних ресурсів, які витрачаються на їх навчання та перенавчання. За останні 10 років, за відомостями компанії OpenAI спостерігається збільшення в 300 тис. разів витрачених на навчання ГНМ обчислювальних ресурсів. Аналіз показав, що час навчання ГНМ значною мірою визначається оптимізаційними методами, які використовуються. Тому, важливою науково-практичною задачею, яка вирішується у дисертаційній роботі, є скорочення часу навчання нейронних мереж при вирішенні задач сегментації та класифікації об'єктів шляхом удосконалення оптимізаційних методів, які

для цього використовуються.

На основі аналізу існуючих методів навчання ГНМ, які входять в сучасні бібліотеки розробки нейронних мереж, виявлено, що більшість із них засновані на методі градієнтного спуску, який в міні-пакетному режимі навчання формально виглядає наступним чином:

$$\Theta_{t+1} = \Theta_t - \frac{1}{B}\gamma\nabla_\Theta \sum_{i=1}^{B} L\left(f(\bar{x}_i;\Theta_t), \bar{y}_i\right) \tag{1}$$

де Θ_t – параметри нейронної мережі на момент часу t; B – кількість елементів в одному міні-пакеті тренувальних даних; $L\left(f(\bar{x}_i;\Theta_t), \bar{y}_t\right)$ – функція втрат, що відображає похибку між передбаченими значеннями $f(\bar{x}_i;\Theta_t)$ та очікуваними результатами $\bar{y}_t$; $f(\bar{x}_i;\Theta_t)$ – функція, що позначає відповідні перетворення, які виконуються нейронною мережею з внутрішніми параметрами Θ_t та вхідними параметрами $\bar{x}_i$; γ – коефіцієнт навчання.

Якщо ввести додаткову функцію від параметра ϑ:

$$G(\vartheta) = \vartheta - \frac{1}{B}\gamma\nabla_\vartheta \sum_{i=1}^{B} L\left(f(\bar{x}_i;\vartheta), \bar{y}_i\right),$$

то вираз (1) можна представити у вигляді класичної дискретної динамічної системи (ДДС).

$$\Theta_{t+1} = G(\Theta_t). \tag{2}$$

Таким чином, вираз (2) дозволяє представити метод градієнтного спуску у вигляді ДДС, до якої можна застосувати положення нелінійної динаміки. Показано, що таке представлення дозволяє розробляти модифікації ДДС (2) з урахуванням зворотного зв'язку з запізненням та прогнозуючого зворотного зв'язку.

З урахуванням (1) та (2) запропоновано модифікацію методу градієнтного спуску з урахуванням зворотного зв'язку з запізненням в ДДС:

$$\Theta_{t+1} = (1-\lambda) G\left(\sum_{j=1}^{N} a_j \Theta_{t-jT+T}\right) + \lambda \sum_{j=1}^{N} b_j \Theta_{t-jT+1}. \tag{3}$$

У (3) параметри a_j та b_j повинні задовольняти умовам $\sum_j a_j = \sum_j b_j = 1, \forall \lambda \in \Re$.

Для їх обчислення необхідна розробка відповідного математичного апарату.

З урахуванням (1) і (2) запропоновано модифікацію методу градієнтного спуску з урахуванням прогнозуючого зворотного зв'язку в ДДС:

$$\Theta_{t+1} = G(\Theta_t) + u_t, \tag{4}$$

де прогнозуючий зворотній зв'язок $u_t = (a_1 - 1)G(\Theta_t) + \sum\limits_{j=2}^{N} a_j G^{((j-1)T+1)}(\Theta_t)$, $G^{(1)}(\Theta) = G(\Theta)$, $G^{(k)}(\Theta) = G(G^{(k-1)}(\Theta))$, $\sum\limits_{j=1}^{N} a_j = 1$, $k \in \mathbb{N}$.

Таким чином, удосконалено метод градієнтного спуску на основі положень нелінійної динаміки (**перший пункт новизни наукового дослідження**), що дозволило розробити модифікації методу з урахуванням зворотного зв'язку з запізненням та прогнозуючого зворотного зв'язку в ДДС та застосувати його як при навчанні нейронних мереж з метою зменшення часу навчання так і для вирішення інших оптимізаційних задач.

В другому розділі розроблено метод пошуку параметрів для запропонованих модифікацій методу градієнтного спуску з урахуванням зворотного зв'язку з запізненням та прогнозуючого зворотного зв'язку в ДДС.

Запропонований метод пошуку параметрів реалізується наступною послідовністю:

Етап 1. Рішення оптимізаційної задачі на класі комплексних поліномів.

При вирішенні оптимізаційної задачі параметри комплексних поліномів необхідно вибирати таким чином, щоб множина $\left(\bar{\mathbb{C}} \setminus \Phi(\bar{\mathbb{D}})\right)^*$ на комплексній площині була найбільш розтягнутою в потрібному напрямку, наприклад, у від'ємному напрямку дійсної осі. Така задача розглядається на класі однолистних функцій та типово-дійсних функцій по Рогозинському. Застосовані означення: $\bar{\mathbb{C}}$ – розширена комплексна площина, $\hat{\mathbb{D}}$ – замкнутий центральний одиничний круг, зірочкою позначена операція інверсії $(z)^* = \frac{1}{z}$, а функція $\Phi(z) = (1-\lambda)^T \frac{z\left(\sum\limits_{j=0}^{N-1} a_j z^j\right)^T}{\left(1-\lambda\sum\limits_{j=1}^{N} b_j z^j\right)^T}$.

Рішення зводиться до розв'язання екстремальних задач, для яких знаходяться екстремальні значення, та екстремайзери.

1. Для полінома $F(z) = \sum\limits_{j=1}^{N} a_j z^j$ з дійсними коефіцієнтами та schlicht- нормуванням $a_1 = 1$. Знайти $\sup\limits_{a_2,\ldots,a_N}\left(\inf\limits_{z\in\mathbb{D}}\{\Re(F(z)) : \Im(F(z)) = 0\}\right)$.

   Екстремальне значення є таким $-\frac{1}{4}\sec^2\frac{\pi}{N+2}$, а відповідний екстремайзер

   $$F(z) = \frac{1}{U_N'\left(\cos\frac{\pi}{N+2}\right)}\sum_{j=1}^{N} U_{N-j+1}'\left(\cos\frac{\pi}{N+2}\right)U_{j-1}\left(\cos\frac{\pi}{N+2}\right)z^j$$

   є єдиним та однолистним, $U_j(x)$ - поліноми Чебишова другого роду.

2. У класі T_N усіх типово дійсних поліномів ступеню N schlicht нормуванням $a_1 = 1$ знайти величину $J_N = \max\limits_{p(z)\in T_N}\sup\limits_{z\in\mathbb{D}}\{|F_n(z)|\}$. Екстремальне значення визначено як $J_N = \begin{cases}\frac{1}{4\sin^2\frac{\pi}{2(N+2)}}, & N - \text{непарне}\\ \frac{1}{4\sin^2\vartheta}, & N - \text{парне}\end{cases}$, де ϑ мінімальний позитивний корінь рівняння $(N+3)\cos(N+1)\vartheta + (N+1)\cos(N+3)\vartheta = 0$. Наведено алгоритм обчислення коефіцієнтів екстремайзера.

3. У класі однолистних з T-круговою симетрією поліномів $F_N^{(T)} =$ $= z\left(1 + \sum\limits_{j=2}^{N} a_j z^{(j-1)T}\right)$ знайти асимптотичну оцінку величини

$\max\limits_{z\in\mathbb{D}} |F_N^{(T)}(z)|$. Екстремальне асимптотичне значення при $N \to \infty$, $\max\limits_{z\in\mathbb{D}} |F_N^{(T)}| \sim \frac{1}{2^{2/T} c_T} N^{2/T}$, де $c_T = \pi^{\frac{2}{T}-1}\Gamma^2\left(\frac{1}{T}+\frac{1}{2}\right)$, $\Gamma(\xi)$ - гама-функція.

Экстремайзером $F^{(0)}(z) = z + \sum\limits_{j=2}^{N} a_j^{(0)} z^{T(j-1)+1}$, є

$$a_j^{(0)} = \left(1 - \frac{(j-1)T}{1+(N-1)T}\right) \prod_{k=1}^{j-1} \frac{\sin\frac{\pi(2+T(j-1))}{2+T(N-1)}}{\sin\frac{\pi T j}{2+T(N-1)}}, \; j = 2, \ldots, N.$$

Коефіцієнти екстремайзерів, які екстремально розтягують або стискають центральний одиничний круг на комплексній площині, використовуються для чисельного моделювання з метою визначення допустимих параметрів.

Етап 2. Проведення чисельного моделювання з використанням варіацій допустимих параметрів, з метою визначення параметрів a_j, b_j та λ удосконаленого методу градієнтного спуску (3), (4).

Етап 3. Експериментальна перевірка знайдених параметрів для різних завдань.

Розроблений метод пошуку параметрів для запропонованих модифікацій методу градієнтного спуску з урахуванням зворотного зв'язку з запізненням та прогнозуючого зворотного зв'язку в ДДС на основі розв'язання оптимізаційних задач на класі комплексних поліномів, який дозволяє забезпечити їх реалізацію для широкого кола оптимізаційних задач, є **другим пунктом наукової новизни**.

У **третьому розділі** набули подальшого розвитку методи навчання нейронних мереж шляхом запропонованих модифікацій методу градієнтного спуску. Проведено їх дослідження для різних архітектур нейронних мереж.

Розроблено на основі (3) метод навчання нейронних мереж шляхом модифікації методу градієнтного спуску з урахуванням зворотного зв'язку з запізненням в ДДС при заданих параметрах $T = 1$, $N = 2$:

$$\Theta_{t+1} = ((1-\lambda)a + \lambda b)\Theta_t + (1 - ((1-\lambda)a + \lambda b))\Theta_{t-1} - (1-\lambda)\frac{1}{B}\gamma\nabla_\Theta \sum_{i=1}^{B} L(f(\bar{x}_i; a\Theta_t + (1-a)\Theta_{t-1}), \bar{y}_i), \quad (5)$$

де параметри a, b и λ визначаються у відповідності з наведеним вище методом пошуку параметрів.

Розроблено на основі (4) метод навчання нейронних мереж шляхом модифікації методу градієнтного спуску з урахуванням прогнозуючого зворотного зв'язку в ДДС при заданих параметрах $T = 1$, $N = 2$:

$$\begin{cases} y_t = G(\Theta_t) \\ x_{t+1} = G(a + (1-a)y_t) \end{cases} \quad (6)$$

або

$$\begin{cases} y_t = G(\Theta_t) \\ x_{t+1} = aG(\Theta_t) + (1-a)G(y_t) \end{cases} . \quad (7)$$

Ці схеми було покладено в основу створення низки нових схем реалізації методу градієнтного спуску. Проведено їх тестування на прикладах.

1. Розв'язана оптимізаційна задача $\min\limits_{x\in\mathbb{R}} \max\limits_{y\in\mathbb{R}} [\Phi(x,y)]$ для функції $\Phi(x,y) = 2x^2 + y^2 + 4xy + \frac{4}{3}y^2 - \frac{1}{4}y^4$, для яких відомі 3 критичні точки (точка z_1 - точка локального мінімуму, а точки z_0 и z_2 - сідлові точки) в яких $\nabla\Phi(x,y) = 0$, а саме $z_0 = (0,0)$; $z_1 = (-2-\sqrt{2}, 2+\sqrt{2})$; $z_2 = (-2+\sqrt{2}, 2-\sqrt{2})$.

   Проведено експерименти з побудови басейнів тяжіння до екстремальних точок. Обчислення проводилися для невеликих областей локалізації екстремальних точок з накладеною на них дрібною сіткою за допомогою розробленої на мові програмування GoLang багатопотокової програми. Результати аналізу показали, що модифікації (6) і (7) призводять до збіжності до всіх трьох точок, включаючи сідлові (рис. 1), та потребують менш обчислювальних ресурсів, ніж методи оптимізації, які вимагають обчислення Гессіану. Здатність збіжності до сідлових точок є ключовою перевагою, завдяки чому можливо значно поліпшити навчання GAN мереж (generative adversarial networks).

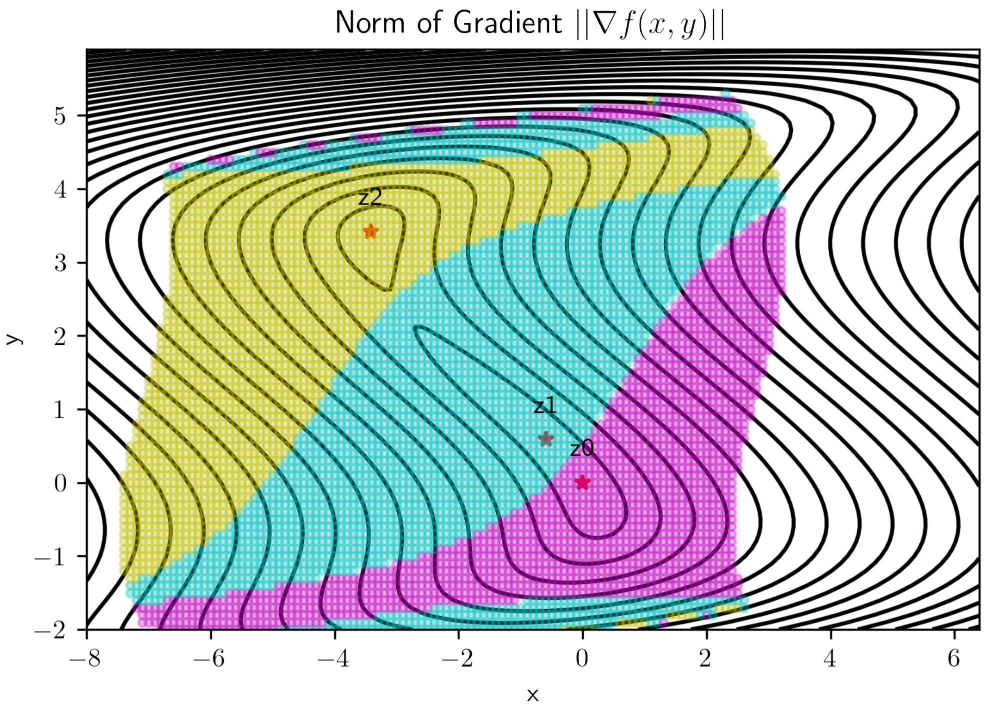

Рисунок 1: Контурний графік $\|\Phi(x, y)\|$ (блакитним кольором позначено басейн тяжіння до точки z_1, жовтим до z_2, пурпурним до z_0

2. Розв'язана оптимізаційна задача для типових тестових функцій, однією з яких є функція Розенброка $f(x, y) = (1 - x)^2 + 100(y - x^2)^2$, яка має глобальний мінімум у точці $(x, y) = (1, 1)$.

Проведено порівняльний аналіз результатів пошуку екстремуму функції Розенброка, яку було надано у вигляді нейронної мережі, досліджуваним та найпопулярнішими методами для навчання глибоких нейронних мереж (SGD (Stochastic Gradient Decent) алгоритмом з моментом та методом Нестерова, алгоритмом Adam) з використанням бібліотеки PyTorch компанії Facebook, що містить градієнтні алгоритми оптимізації. Численні експерименти показали перевагу розробленого методу (New SGD algorithm with control) як у швидкості збіжності, так і в більш плавній поведінці функції втрат (рис. 2).

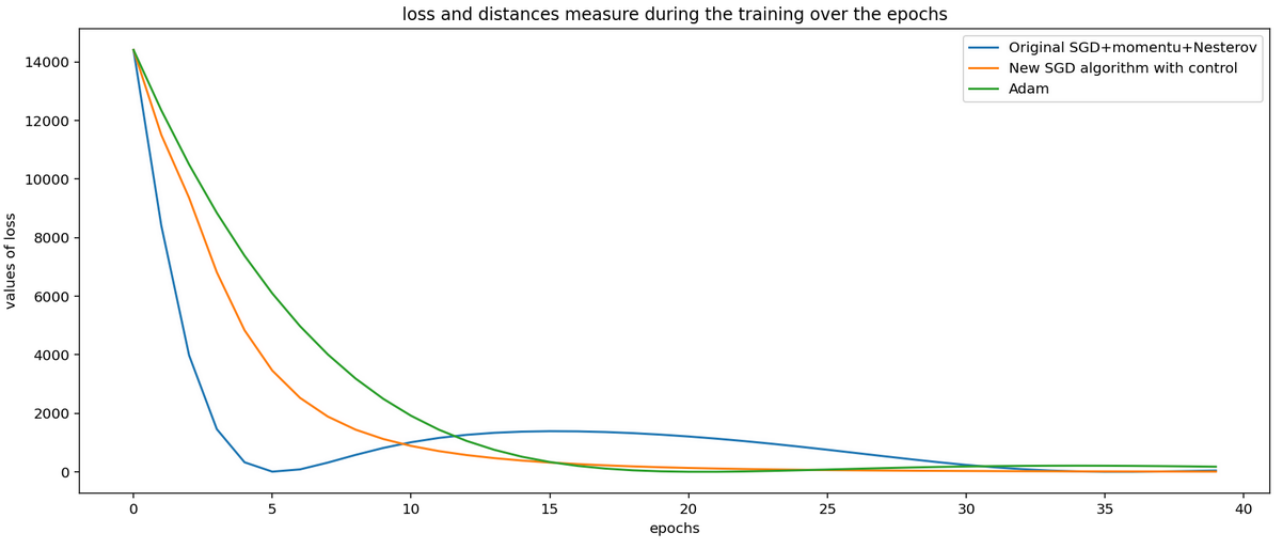

Рисунок 2: Графік функції втрат при пошуку екстремуму функції

3. Навчання багатошарового персептрона для розпізнавання рукописних цифр.

В якості навчального та тестового наборів даних був використаний MNIST (60,000 зображень помічених вручну). Аналіз функції помилки навчання багатошарового перцептрона при розпізнаванні рукописних цифр показав, що розроблений метод забезпечує скорочення часу навчання (до трьох разів) при меншому рівні помилки (рис. 3).

Рисунок 3: Графік функції помилки навчання розпізнавання рукописних цифр

Таким чином, набули подальшого розвитку методи навчання нейронних мереж на основі запропонованих модифікацій методу градієнтного спуску, що дозволило зменшити час навчання нейронних мереж різних архітектур при вирішенні задач сегментації та класифікації об'єктів (**третій пункт новизни наукового дослідження**).

У **четвертому розділі** розроблена програмна реалізація запропонованих методів навчання нейронних мереж на основі нелінійної динаміки та проведено їх апробацію при побудові підсистеми сегментації та класифікації ортопантомограм для стоматологічної інтелектуальної інформаційної системи.

При побудові підсистеми вирішувалась задача перетворення ортопантомограм в *одонтограми*, які є спеціалізованою формою представлення інформації лікарю про стан кожного зуба пацієнта, з застосуванням згорткової нейронної мережі.

На першому етапі відбувалось формування навчальної вибірки. Оскільки створення необхідної навчальної вибірки є неможливим у зв'язку з високою ціною отримання панорамних зображень, необхідністю їх фільтрації від сторонніх шумів і складністю проведення початкової ручної сегментації стоматологами, використано метод аугментації навчальної вибірки з використанням як геометричних, так і піксельних перетворень. Це дозволило сформувати 300 тис. зображень на основі початкового набору з 53 зображень і 32 бінарних масок. Всі зображення були збережені як багатовимірний тензор PyTorch в спеціальному внутрішньому форматі.

Для вирішення задачі сегментації обрано архітектуру гібридної згорткової нейронної мережі на основі попередньо навченого мережевого кодера ResNet-18 і декодера з мережі U-Net. Навчання цієї мережі передбачає налаштування близько 18 мільйонів параметрів. В якості функції помилки запропоновано використовувати лінійну комбінацію крос-ентропії і коефіцієнта Дайса, що добре зарекомендувала себе при навчанні згорткових нейронних мереж. Навчання проводилося з використанням бібліотеки PyTorch з різними модифікаціями методу градієнтного спуску. Досліджено використання оперативної пам’яті GPU, мінімальне значення функції втрати (Минимум) і значення функції втрати (Тест) на тестовому наборі даних (таблиця 1).

Таблиця 1: Порівняльний аналіз алгоритмів навчання

Назва	**GPU пам’ять (MiB)**	**Минимум**	**Тест**
RMSProps	3503	1.3766	0.5385
Adagrad	3571	0.0502	0.5269
Adadellta	3459	0.5520	0.5560
Adam	3459	**0.0474**	1.189
SGD	**3450**	0.2197	0.5365
New method (3)	3471	0.2673	**0.2049**

Аналіз показав, що запропонований метод, за помилкою навчання має аналогічні результати на навчальній виборці, а на тестовій спостерігається її зменшення. Крім того при однакових витратах оперативної пам’яті GPU для розроблених методів функція втрати в два рази менше на тестових даних в порівнянні з SGD і майже в 7 разів в порівнянні з алгоритмом Adam ([18]).

Графіки функції втрати під час навчання нейронної мережі (рис. 4), показали, що New method (3) майже вдвічі швидше мінімізує функцію втрати в порівнянні з SGD.

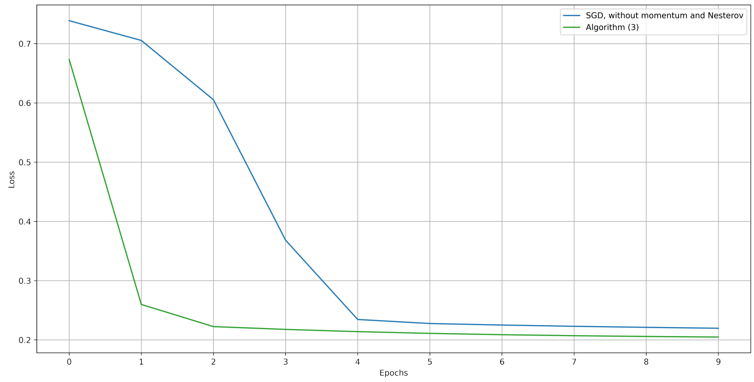

Рисунок 4: Графік порівняння швидкості мінімізації функції втрати алгоритмів SGD (синій) та методу (3) (зелений) від часу навчання

У **висновках** сформульовано отримані в ході роботи над дисертацією науково-практичні результати.

Ключові слова: методи навчання нейронних мереж, оптимізація, згорткові нейронні мережі, сегментація, класифікація.

СПИСОК ПУБЛІКАЦІЙ ЗДОБУВАЧА

Основні матеріали дисертаційної роботи і головні результати проведених досліджень були опубліковані в сьоми наукових роботах (4 статті – у науково-метричних базах Scopus, 3 першого квартиля Q1; 3 – у фахових збірниках, 3 без співавторів), а також у одинадцяті апробаційних матеріалах.

Наукові праці, в яких опубліковані основні наукові результати дисертації

1. Dmitrishin D., Smorodin A., Stokolos A. On a family of extremal polynomials. Comptes Rendus Mathematique. 2019. 357(7). P. 591-596. DOI:https://doi.org/10.1016/j.crma.2019.06.010. (Scopus, Q1).

2. Dmitrishin D., Smorodin A., Stokolos A., Tohaneanu M. Symmetrization of Suffridge polynomials and approximation of T-symmetric Koebe functions. Journal of
Mathematical Analysis and Applications. 2021. 503(3). DOI:https://doi.org/10.1016/j.jmaa.2021.125350. (Scopus, Q1).

3. Dmitrishin D., Smorodin A., Stokolos A. An extremal problem for polynomials. Applied and Computational Harmonic Analysis. 2021. 56. P. 283-305. DOI:https://doi.org/10.1016/j.acha.2021.08.008. (Scopus, Q1)

4. Smorodin A. Predictive control methods in tasks of searching saddle points. Proceedings of Odessa Polytechnic University. 2020. V. 3(62). P. 80-90. DOI:10.15276/opu.3.62.2020.10. URL:https://pratsi.op.edu.ua/articles/show/20968

5. Smorodin A. The use of control theory methods in training neural networks on the example of teeth recognition on panoramic x-ray images//Automation of Technological and Business Processes. 2021. V. 13(2). P. 36-40. URL:http://dx.doi.org/10.15673/atbp.v13i2.2055

6. Smorodin, A.V. Використання методів теорії управління при вивченні нейронних мереж на прикладі рукописного тексту. Прикладні аспекти інформаційних технологій. V. 4(3). 2021. P. 243–249. URL:http://aait.ccs.od.ua/index.php/journal/article/view/119

7. Dillies, J., Dmitrishin, D., Smorodin, A., Stokolos A. (2022). On the Koebe Quarter

Theorem for Polynomials. Proceedings of the International Geometry Center, V. 14(3). 2022. P. 219-230. https://doi.org/10.15673/tmgc.v14i3.2057. (Scopus, Q4)

Наукові праці, які засвідчують апробацію матеріалів дисертації

8. Dmitrishin D., Smorodin A., Stokolos A. An extremal problem for polynomials. Harmonic Analysis and Partial Differential Equations. Canadian Mathematical Society. 2021.

9. Dmitrishin D., Smorodin A., Stokolos A. Extremal polynomials and geometric complex analysis. Algebraic and Geometric Methods of Analysis. International Online Conference. 2020.

10. Smorodin A. On question by Dimitrov. Analysis Seminar. Odessa. I. I. Mechnikov National University. 2018.

11. Смородин А. Методы прогнозирующего управления в задачах поиска седловых точек и обучения нейронных сетей. /Актуальні питання фізико-математичних та технічних наук: теоретичні та прикладні дослідження. 2021. P. 13-23. II Міжнародна науково-практична інтернет-конференція. URL: https://openscilab.org/wp-content/uploads/2021/05/aktualni-pitannja-fiziko-matematichnih-ta-tehnichnih-nauk-teoretichni-ta-prikladni-doslidzhennja_2021_05_21.pdf

12. Смородін А. Методи прогнозного управління для задач оптимізації. Теоретико-практичні проблеми використання математичних методів та комп'ютерно-орієнтованих технологій в освіті та науці. 2021. III Всеукраїнська науково-практична онлайн-конференція.

13. Dmitrishin D., Stokolos A., Smorodin A. On a new family of extremal polynomials. The 35th annual Southeastern Analysis. Meeting, University of Alabama, 2019.

14. Dmitrishin D., Stokolos A., Smorodin A. An extremal problem for polynomials. CMS 75th +1 Anniversary Summer Meeting, Montreal, 2021

15. Dmitrishin D., Skrinnik I., Smorodin A., Stokolos A. Dimitrov's question for the polynomials of degree 1,2,3,4,5,6. 2018. arXiv:1808.08636. URL: https://arxiv.org/abs/1808.08636.

16. Dmitrishin D., Smorodin A., Stokolos A. Estimating the Koebe radius for polynomials. 2018. arXiv: 1805.06927. URL: https://arxiv.org/abs/1805.06927v1

17. Dmitrishin D., Smorodin A., Stokolos. On C. Michel's hypothesis about the modulus of typically real polynomials. 2020. arXiv: 2005.12432. URL: https://arxiv.org/abs/2005.12432

18. Dillies J., Dmitrishin D., Smorodin A., Stokolos. On the Koebe Quarter Theorem for Polynomials. 2019. arXiv: 1904.11039. URL: https://arxiv.org/abs/1904.11039

ABSTRACT

Smorodin A.V. Methods of neural networks training on the basis of nonlinear dynamics. – Qualification scientific work on the rights of the manuscript.

PhD degree dissertation – Doctor of Philosophy in specialty 122 – Computer Science.– Odessa Polytechnic State University Ministry of Education and Science of Ukraine of Ukraine, Odessa, 2021.

In the **introduction** the relevance of improving the methods of neural networks training to solve the problems of classification and segmentation is proved. The possibility of reducing the training time of neural networks is shown based on the application of provisions of nonlinear dynamics when using gradient descent algorithms. Object, subject, tasks and methods of research are determined; the connection with scientific programs and plans is shown; scientific novelty and practical significance of the obtained results are given; the applicant's personal contribution is highlighted.

The **first section** of the dissertation analyzes the problems of segmentation and classification of objects in images using neural networks. It is shown that despite the effectiveness of existing approaches to their solution usually based on deep neural networks (DNN) of complex architecture, existing trends of big data (large volumes, diversity, variability) lead to a natural problem – an increase in computing resources spent on their training and retraining (over the past 10 years OpenAI has seen a 300,000-fold increase in computing resources spent on DNN training). The analysis showed that the training time of DNN is largely determined by the optimization methods used. Therefore, the important scientific and practical task, which is solved in this dissertation, is to reduce the training time of neural networks in solving the problems of segmentation and classification by improving the optimization methods used for this purpose.

Based on the analysis of existing methods of training DNN which are included in modern libraries of neural networks development, it was found that most of them are

based on the method of gradient descent, which in the mini-batch learning mode looks like this:

$$\Theta_{t+1} = \Theta_t - \frac{1}{B}\gamma\nabla_{\Theta}\sum_{i=1}^{B} L\left(f(\bar{x}_i;\Theta_t), \bar{y}_i\right) \tag{1}$$

where Θ_t – neural network parameters at the time t, B – number of items in one mini training data package, $L\left(f(\bar{x}_i;\Theta_t), \bar{y}_t\right)$ – loss function, which reflects the error between the predicted values $f(\bar{x}_i;\Theta_t)$ and expected results $\bar{y}_t$, $f(\bar{x}_i;\Theta_t)$ – function that denotes the corresponding transformations performed by the neural network with internal parameters Θ_t and input parameters learning $\bar{x}_i$; γ – learning rate.

If one enters an additional function from the parameter ϑ:

$$G(\vartheta) = \vartheta - \frac{1}{B}\gamma\nabla_{\vartheta}\sum_{i=1}^{B} L\left(f(\bar{x}_i;\vartheta), \bar{y}_i\right)$$

then expression (1) can be represented as the classic discrete dynamic system (DDS).

$$\Theta_{t+1} = G(\Theta_t). \tag{2}$$

Thus, the expression (2) allows to represent a method of gradient descent in the form of the DDS, to which the position of nonlinear dynamics can be applied. It is shown that such presentation allows to develop modifications of the DDS (2) considering the delayed and predictive feedback.

Taking into account (1) and (2), a modification of the gradient descent method is proposed considering the delayed feedback in the DDS:

$$\Theta_{t+1} = (1-\lambda)G\left(\sum_{j=1}^{N} a_j\Theta_{t-jT+T}\right) + \lambda\sum_{j=1}^{N} b_j\Theta_{t-jT+1}. \tag{3}$$

In (3) parameters a_j and b_j must meet the conditions $\sum_j a_j = \sum_j b_j = 1, \forall\lambda \in \Re$.

To calculate them, it is necessary to develop a suitable mathematical apparatus.

Taking into account (1) and (2), a modification of the gradient descent method is proposed considering the predictive feedback in the DDS:

$$\Theta_{t+1} = G(\Theta_t) + u_t, \tag{4}$$

where the predictive feedback $u_t = (a_1 - 1)G(\Theta_t) + \sum\limits_{j=2}^{N} a_j G^{((j-1)T+1)}(\Theta_t)$, $G^{(1)}(\Theta) = G(\Theta)$, $G^{(k)}(\Theta) = G(G^{(k-1)}(\Theta))$; $\sum\limits_{j=1}^{N} a_j = 1$, $k \in \mathbb{N}$.

Thus, the method of gradient descent based on nonlinear dynamics (**the first point of novelty of the scientific research**) was improved which allowed to develop modifications of the method considering the delayed and predictive feedback in the DDS and apply it in neural networks training in order to reduce the training time and to solve other optimization problems.

In the second section the method of search of parameters for the offered modifications of the gradient descent method considering the delayed and predictive feedback in the DDS is developed.

The proposed method of searching for parameters is implemented in the following sequence:

Stage 1. Solution of the optimization problem on the class of complex polynomials.

When solving the optimization problem, the parameters of complex polynomials must be selected so that the set $\left(\bar{\mathbb{C}} \setminus \Phi(\bar{\mathbb{D}})\right)^*$ on the complex plane was the most stretched in the desired direction, for example, in the negative direction of the real axis. Such problem is considered in the class of univalent and typically-real Rogozinski functions. Applied definitions: $\bar{\mathbb{C}}$ - extended complex plane, $\hat{\mathbb{D}}$ - closed central unit circle, an asterisk indicates an inversion operation $(z)^* = \frac{1}{z}$, and the function $\Phi(z) = (1-\lambda)^T \frac{z\left(\sum\limits_{j=0}^{N-1} a_j z^j\right)^T}{\left(1-\lambda\sum\limits_{j=1}^{N} b_j z^j\right)^T}$.

The solution comes down to solving optimization problems for which extreme values and extremizers were found.

1. For polynomial $F(z) = \sum\limits_{j=1}^{N} a_j z^j$ with valid coefficients and schlicht-rationing $a_1 = 1$. Find $\sup\limits_{a_2,\ldots,a_N}\left(\inf\limits_{z\in\mathbb{D}}\{\Re(F(z)) : \Im(F(z)) = 0\}\right)$.

   The extreme value is $-\frac{1}{4}\sec^2\frac{\pi}{N+2}$, and the corresponding extremizer

$$F(z) = \frac{1}{U_N'\left(\cos\frac{\pi}{N+2}\right)} \sum_{j=1}^{N} U_{N-j+1}'\left(\cos\frac{\pi}{N+2}\right) U_{j-1}\left(\cos\frac{\pi}{N+2}\right) z^j$$

   is single and univalent, $U_j(x)$ - Chebyshev polynomial of the second kind.

2. In class T_N ll typically valid polynomials of degree N with schlicht-rationing $a_1 = 1$ find value $J_N = \max\limits_{p(z)\in T_N}\sup\limits_{z\in\mathbb{D}}\{|F_n(z)|\}$. Extreme value is defined as $J_N = \begin{cases} \frac{1}{4\sin^2\frac{\pi}{2(N+2)}}, N-\text{odd} \\ \frac{1}{4\sin^2\vartheta}, N-\text{even} \end{cases}$, where ϑ minimum positive root of the equation $(N+3)\cos(N+1)\vartheta + (N+1)\cos(N+3)\vartheta = 0$. The algorithm for calculating extremizer coefficients is given.

3. In the class of univalent T-fold symmetry polynomials $F_N^{(T)} =$ $= z\left(1 + \sum\limits_{j=2}^{N} a_j z^{(j-1)T}\right)$ find an asymptotic estimate of the value $\max\limits_{z\in\mathbb{D}}|F_N^{(T)}(z)|$. Extreme asymptotic value at $N\to\infty$, $\max\limits_{z\in\mathbb{D}}|F_N^{(T)}| \sim \frac{1}{2^{2/T}c_T}N^{2/T}$, where $c_T = \pi^{\frac{2}{T}-1}\Gamma^2\left(\frac{1}{T}+\frac{1}{2}\right)$, $\Gamma(\xi)$ - gamma-function.

   Extremizer $F^{(0)}(z) = z + \sum\limits_{j=2}^{N} a_j^{(0)} z^{T(j-1)+1}$, is

$$a_j^{(0)} = \left(1 - \frac{(j-1)T}{1+(N-1)T}\right) \prod_{k=1}^{j-1} \frac{\sin \frac{\pi(2+T(j-1))}{2+T(N-1)}}{\sin \frac{\pi T j}{2+T(N-1)}}, j = 2, \ldots, N.$$

The extremizer coefficients that extremely stretch or compress the central unit circle on the complex plane are used for numerical simulations to determine the allowable parameters.

Stage 2. Carrying out of numerical simulations using variations of the allowable parameters to determine the parameters a_j, b_j and λ of the improved gradient descent method (3), (4).

Stage 3. Experimental verification of the parameters found for various problems.

The method of searching for parameters for the proposed modifications of the gradient descent method considering the delays and predictive feedback in the DDS based on the solution of optimization problems in the class of complex polynomials, which allows their implementation for a wide range of optimization problems, is **the second point of scientific novelty**.

In **the third section** methods of neural network training were further developed based on the proposed modifications of the gradient descent method. The research for different architectures of neural networks was carried out.

The method of neural networks training based on the modification of the gradient descent method considering the delayed feedback in the DDS is developed on the basis of (3) at the specified parameters $T = 1$, $N = 2$:

$$\Theta_{t+1} = ((1-\lambda)a + \lambda b)\Theta_t + (1 - ((1-\lambda)a + \lambda b))\Theta_{t-1} - (1-\lambda)\frac{1}{B}\gamma\nabla_\Theta \sum_{i=1}^{B} L(f(\bar{x}_i; a\Theta_t + (1-a)\Theta_{t-1}), \bar{y}_i), \quad (5)$$

where parameters a, b and λ are determined in accordance with the above method of searching for parameters.

The method of neural networks training based on the modification of the gradient descent method considering the predictive feedback in the DDS is developed on the basis of (4) at the specified parameters $T = 1$, $N = 2$:

$$\begin{cases} y_t = G(\Theta_t) \\ x_{t+1} = G(a + (1-a)y_t) \end{cases} \tag{6}$$

or

$$\begin{cases} y_t = G(\Theta_t) \\ x_{t+1} = aG(\Theta_t) + (1-a)G(y_t) \end{cases} . \tag{7}$$

These schemes were the basis for the creation of a number of new schemes for the implementation of the gradient method descent. They were tested on examples.

1. The optimization problem $\min_{x\in\mathbb{R}} \max_{y\in\mathbb{R}} [\Phi(x,y)]$ for function $\Phi(x,y) = 2x^2 + y^2 + 4xy + \frac{4}{3}y^2 - \frac{1}{4}y^4$, for which 3 critical points are known (point z_1 - point of the local minimum, and the points z_0 and z_2 - saddle points) in which $\nabla\Phi(x,y) = 0$, namely $z_0 = (0,0)$; $z_1 = (-2-\sqrt{2}, 2+\sqrt{2})$; $z_2 = (-2+\sqrt{2}, 2-\sqrt{2})$.

   Experiments were carried out to build pools of attraction to extreme points. The calculations were performed for small areas of localization of extreme points with a fine grid superimposed on them using a multi-threaded program developed in the GoLang programming language. The results of the analysis showed that modifications (6) and (7) lead to convergence to all three points, including saddles (Fig. 1), and require less computational resources than optimization methods that require the calculation of Hessian. The ability to convergence to saddle points is the key advantage making it possible to significantly improve the training of GAN networks (generative adversarial networks).

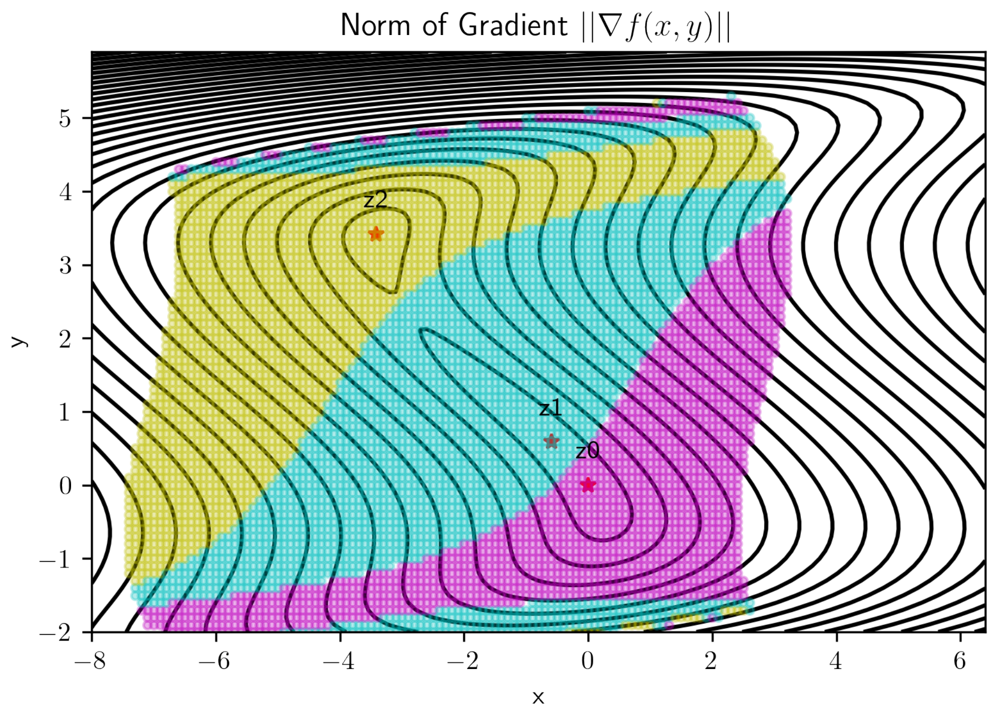

Figure 1: Contour graph $\|\Phi(x, y)\|$ (blue indicates the pool of attraction to the point z_1, yellow to z_2, purple to z_0

2. The optimization problem for typical test functions, one of which is the Rosenbrock function $f(x, y) = (1 - x)^2 + 100(y - x^2)^2$, which has a global minimum at the point of $(x, y) = (1, 1)$ is solved.

   A comparative analysis of the results of the search for the extreme of the Rosenbrock function, which was provided as a neural network, researched and the most popular methods for deep neural networks training (StoGD (Stochastic Gradient Decent) algorithm with momentum? and the Nesterov method, the Adam algorithm), using Facebook's PyTorch library, which contains all gradient optimization algorithms was performed. Numerous experiments showed the advantage of the developed method (New SGD algorithm with control) both in the rate of convergence and in the smoother behavior of the loss function (Fig. 2).

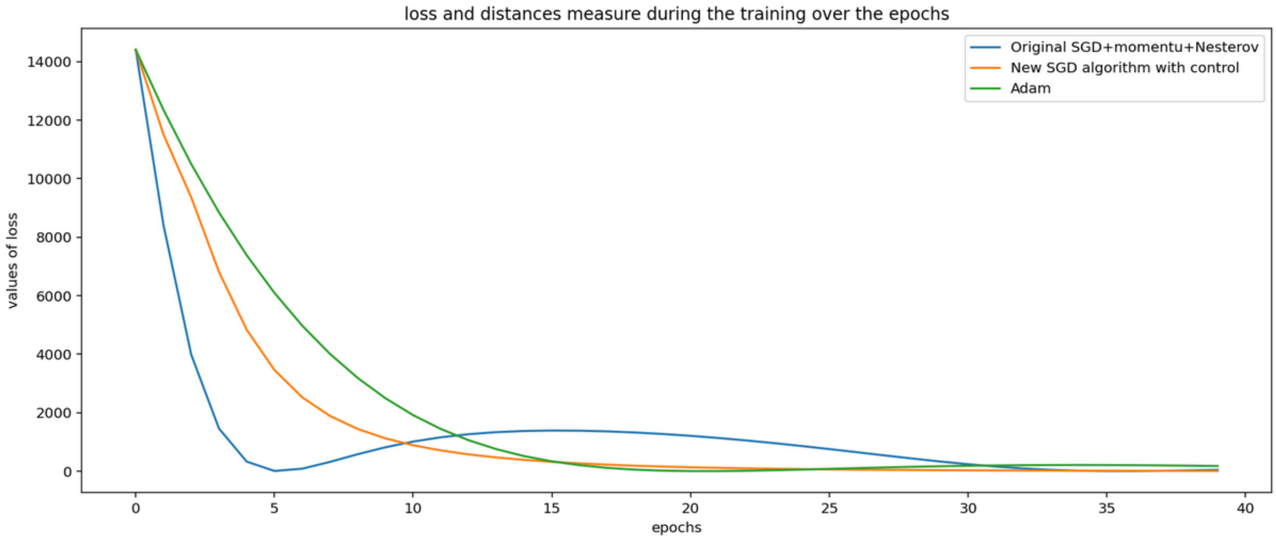

Figure 2: Graph of loss function when searching for an extreme function

3. Training multilayer perceptron for handwriting recognition.

MNIST (60,000 manually marked images) was used as a training and test dataset. Analysis of the multilayer perceptron learning error function in handwriting recognition showed that the developed method reduces the training time (up to three times) at a lower error rate (Fig. 3).

Figure 3: Graph of handwriting recognition learning error function

Thus, the methods of neural networks training were further developed based on the proposed modifications of the gradient descent method, which reduced the training time of neural networks of different architectures in solving problems of segmentation and classification of objects (**the third point of novelty of the scientific research**).

In **the fourth section** the software realization of the offered methods of neural networks training on the basis of nonlinear dynamics was developed and their approbation was carried out to build the subsystem of segmentation and classification of orthopantomograms for the dental intelligent information system.

When building the subsystem, the problem of converting *orthopantomograms* into odontograms, which are a specialized form of presenting information to a doctor about the condition of each patient's tooth, using a convolutional neural network, was solved.

At the first stage the training sample was formed. Since obtaining the necessary training sample is impossible due to the high cost of acquiring panoramic images, the need to clean them of extraneous noise and the difficulty of the initial manual segmentation by dentists, a method of training sample augmentation was developed using both geometric and pixel transformations. This allowed creation of 300,000 images based on the initial set of 53 images and 32 binary masks. All images were saved as a multi-dimensional PyTorch tensor in a special internal format.

To solve the segmentation problem the architecture of the hybrid convolutional neural network was chosen based on the previously trained ResNet-18 network encoder and the decoder from the U-Net network. Training of this network involves the configuration of about 18 million parameters. As an error function, it is proposed to use a linear combination of cross-entropy and the Dice coefficient, which proved itself well in the training of convolutional neural networks. The training was conducted using the PyTorch library with various modifications of the gradient descent method. The use of GPU RAM, the minimum value of the loss function (Minimum) and the value of the loss function on the test dataset (Test) were studied (Table 1).

Table 1: Comparative analysis

Name	GPU memory (MiB)	Minimum	Test
RMSProps	3503	1.3766	0.5385
Adagrad	3571	0.0502	0.5269
Adadellta	3459	0.5520	0.5560
Adam	3459	**0.0474**	1.189
SGD	**3450**	0.2197	0.5365
New Method (3)	3471	0.2673	**0.2049**

The analysis showed that the proposed method, according to the learning error, has similar results in the training sample, and there is a decrease in the test. In addition, at the same cost of GPU RAM for the developed methods, the loss function is twice less on the test data compared to SGD and almost 7 times less compared to the Adam algorithm ([18]).

The graphs of the loss function during the neural network training (Fig 4) showed that the new method (3) minimizes the loss function relative to the SGD almost twice as fast.

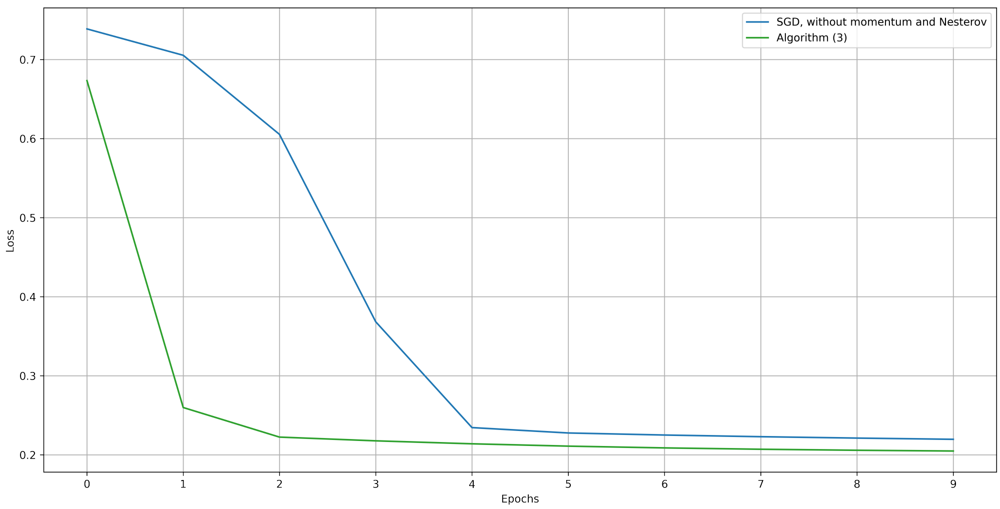

Figure 4: Graph comparing the minimization rate of the loss function of SGD algorithms (blue) and the New Method (3) (green) from the time of training

The conclusions formulate the scientific and practical results obtained during the work on the dissertation.

Keywords: methods of neural networks training, optimization, convolutional neural networks, segmentation, classification.

LIST OF THE APPLICANT'S PUBLICATIONS

The main materials of the dissertation and the main results of the research were published in 7 scientific papers (4 articles were published in the scientific and metric databases Scopus, three in the first quartile Q1; three articles were published in professional collections, 3 without co-authors) and in eleven approbation materials.

Scientific works in which the main scientific results of the dissertation are published

1. Dmitrishin D., Smorodin A., Stokolos A. On a family of extremal polynomials. Comptes Rendus Mathematique. 2019. 357(7). P. 591-596. DOI:https://doi.org/10.1016/j.crma.2019.06.010. (Scopus, Q1).

2. Dmitrishin D., Smorodin A., Stokolos A., Tohaneanu M. Symmetrization of Suffridge polynomials and approximation of T-symmetric Koebe functions. Journal of Mathematical Analysis and Applications. 2021. 503(3). DOI:https://doi.org/10.1016/j.jmaa.2021.125350. (Scopus, Q1).

3. Dmitrishin D., Smorodin A., Stokolos A. An extremal problem for polynomials. Applied and Computational Harmonic Analysis. 2021. 56. P. 283-305. DOI:https://doi.org/10.1016/j.acha.2021.08.008. (Scopus, Q1)

4. Smorodin A. Predictive control methods in tasks of searching saddle points. Proceedings of Odessa Polytechnic University. 2020. V. 3(62). P. 80-90. DOI:10.15276/opu.3.62.2020.10. URL:https://pratsi.op.edu.ua/articles/show/20968

5. Smorodin A. The use of control theory methods in training neural networks on the example of teeth recognition on panoramic x-ray images//Automation of Technological and Business Processes. 2021. V. 13(2). P. 36-40. URL:`http://dx.doi.org/10.15673/atbp.v13i2.2055`

6. Smorodin, A.V. Using control theory methods in the neural networks study on the example of handwritten text. Information technology applied aspects. V. 4(3). 2021. P. 243–249. URL:`http://aait.ccs.od.ua/index.php/journal/article/view/119`

7. Dillies, J., Dmitrishin, D., Smorodin, A., Stokolos A. (2022). On the Koebe Quarter Theorem for Polynomials. Proceedings of the International Geometry Center, V. 14(3). 2022. P. 219-230. `https://doi.org/10.15673/tmgc.v14i3.2057`. (Scopus, Q4)

Scientific works that certify the approbation of the dissertation materials

8. Dmitrishin D., Smorodin A., Stokolos A. An extremal problem for polynomials. Harmonic Analysis and Partial Differential Equations. Canadian Mathematical Society. 2021.

9. Dmitrishin D., Smorodin A., Stokolos A. Extremal polynomials and geometric complex analysis. Algebraic and Geometric Methods of Analysis. International Online Conference. 2020.

10. Smorodin A. On question by Dimitrov. Analysis Seminar. Odessa. I. I. Mechnikov National University. 2018.

11. Smorodin A. Predictive control methods to solve problems of finding saddle points and neural networks training. Current issues of physical, mathematical

and engineering sciences: theoretical and applied research. 2021. P. 13-23. II International scientific and practical Internet conference. URL: https://openscilab.org/wp-content/uploads/2021/05/aktualni-pitannja-fiziko-matematichnih-ta-tehnichnih-nauk-teoretichni-ta-prikladni-doslidzhennja_2021_05_21.pdf

12. Smorodin A. Forecontrol methods at optimization problems solving. Theoretical and practical problems of using mathematical methods and computer-based technologies in education and science. 2021. III All-Ukrainian scientific-practical online conference.

13. Dmitrishin D., Stokolos A., Smorodin A. On a new family of extremal polynomials. The 35th annual Southeastern Analysis. Meeting, University of Alabama, 2019.

14. Dmitrishin D., Stokolos A., Smorodin A. An extremal problem for polynomials. CMS 75th +1 Anniversary Summer Meeting, Montreal, 2021

15. Dmitrishin D., Skrinnik I., Smorodin A., Stokolos A. Dimitrov's question for the polynomials of degree 1,2,3,4,5,6. 2018. arXiv:1808.08636. URL: https://arxiv.org/abs/1808.08636.

16. Dmitrishin D., Smorodin A., Stokolos A. Estimating the Koebe radius for polynomials. 2018. arXiv: 1805.06927. URL: https://arxiv.org/abs/1805.06927v1

17. Dmitrishin D., Smorodin A., Stokolos. On C. Michel's hypothesis about the modulus of typically real polynomials. 2020. arXiv: 2005.12432. URL: https://arxiv.org/abs/2005.12432

18. Dillies J., Dmitrishin D., Smorodin A., Stokolos. On the Koebe Quarter Theorem for Polynomials. 2019. arXiv: 1904.11039. URL: https://arxiv.org/abs/1904.11039