

підвищенням вимог до кібербезпеки та необхідністю системної оптимізації операційних витрат на утримання фізичної ІТ-інфраструктури. Практичне значення отриманих результатів полягає у можливості адаптації бізнес-стратегій до мінливих умов ринку та підвищенні ефективності прийняття управлінських рішень щодо інвестування в цифрові активи. Варто врахувати, що складний характер процесів вимагає постійного моніторингу та поєднання ітераційних методів обчислень із підходами прогнозування для отримання більш точних оцінок у довгостроковій перспективі.

*Abstract.* The article investigates the dynamics of cloud computing implementation in Ukrainian enterprises using the numerical dichotomy method. A mathematical model in the form of a third-degree polynomial was constructed to identify critical efficiency points in the digitalization process. The study determined that the 20 % market penetration threshold was reached in mid-2022 ( $x \approx 5,458$ ), with a fivefold intensification of growth rates observed over the analyzed period. The results demonstrate the high precision of iterative computational methods for forecasting digital economy indicators and optimizing IT infrastructure costs. The proposed approach provides a reliable basis for strategic decision-making in the context of global digital transformation.

*Keywords:* calculation methods, dichotomy method, cloud technologies, digital transformation.

### СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Метод інтерполяції для прогнозування метрик використання хмарних обчислень в статистичному навчанні / Н. А. Потапова, Л. О. Волонтир, І. П. Частоколенко, М. С. Григоренко. *Наука і техніка сьогодні*. 2024. № 4(32). С. 1192–1205. URL: <http://perspectives.pp.ua/index.php/nts/article/view/10943>
2. Чисельні методи: навчальний посібник / Л. О. Волонтир, О. В. Зелінська, Н. А. Потапова, І. А. Чіков. Вінниця: ВНАУ, 2020. 322 с. URL: <https://r.donnu.edu.ua/handle/123456789/1805>
3. Березовська Л., Кириченко А. Розвиток електронної комерції в Україні та ЄС. *Економіка та суспільство*. 2022. № 42. URL: <https://economyandsociety.in.ua/index.php/journal/article/view/1614/1551>
4. Офіційний сайт Державної служби статистики України. Звіт про використання інформаційно-комунікаційних технологій на підприємствах у 2017–2024 рр. URL: [https://www.ukrstat.gov.ua/operativ/operativ2018/zv/ikt/arh\\_ikt\\_u.html](https://www.ukrstat.gov.ua/operativ/operativ2018/zv/ikt/arh_ikt_u.html)

УДК 004:005:51

## ЗАСТОСУВАННЯ ЧИСЕЛЬНИХ МЕТОДІВ ОПТИМІЗАЦІЇ ДЛЯ НАВЧАННЯ МОДЕЛЕЙ ШТУЧНОГО ІНТЕЛЕКТУ

*А. О. Шевчук, Р. М. Бабаков*

*Анотація.* У статті досліджено роль чисельних методів оптимізації у розвитку систем штучного інтелекту. Проведено порівняльний аналіз трьох найбільш поширених методів: SGD, AdamW та mL-BFGS. Досліджено математичний апарат кожного алгоритму, принципи роботи з градієнтами першого та другого порядків, а також стратегії адаптивності. Особливу увагу приділено механізмам стабілізації апроксимації матриці Гессе в умовах стохастичного шуму. Наведені результати тестування підтверджують ефективність методів другого порядку для прискорення збіжності моделей.

*Ключові слова:* чисельна оптимізація, градієнтний спуск, AdamW, mL-BFGS, збіжність.

**Вступ.** Ефективність навчання сучасних моделей штучного інтелекту безпосередньо залежить від вибору та застосування чисельних методів оптимізації [1]. Дане дослідження зосереджене на аналізі ітераційних процесів мінімізації функції втрат – математичної міри помилки, що відображає розбіжність між прогнозом моделі та реальними даними. В умовах роботи з великими масивами даних (Big Data) виникає потреба у методах, які здатні ефективно маневрувати у багатовимірних ландшафтах параметрів, уникаючи сідлових точок та локальних мінімумів.

Дослідження охоплює еволюцію алгоритмів від базових градієнтних підходів до складних квазі-ньютонівських структур. Основний акцент зроблено на вирішенні проблеми стохастичного шуму, який виникає під час обчислення градієнтів на малих випадкових вибірках даних. Це потребує впровадження механізмів стабілізації, як-от моменти та демпфування, для

забезпечення стійкої збіжності – процесу поступового наближення параметрів моделі до оптимального стану.

Аналіз останніх досліджень [2; 3] свідчить, що велике машинне навчання є особливим середовищем, де традиційні методи нелінійної оптимізації часто поступаються стохастичним підходам через величезні обсяги даних та необхідність швидкої ітерації. Водночас актуальною залишається проблема розробки методів, які б враховували кривизну поверхні втрат (другий порядок) без надмірних витрат пам'яті.

**Метою статті** є математичне обґрунтування та порівняльний аналіз ефективності SGD, AdamW та mL-BFGS у задачах навчання нейромережових моделей.

**Основна частина.** Стохастичний градієнтний спуск SGD є фундаментальним методом, що апроксимує градієнт (напрямок зростання помилки) на основі випадкових зразків. Процес оновлення параметрів  $\theta$  описується рівнянням:

$$\theta_{t+1} = \theta_t - \eta \cdot \nabla_{\theta} J(\theta_t; x_i; y_i),$$

де  $\eta$  – швидкість навчання (learning rate);

$\nabla_{\theta}$  – вектор частинних похідних.

Незважаючи на простоту, SGD залишається стабільним методом, особливо за умови використання накопиченого моменту, який допомагає зберігати напрямок руху в областях із зашумленими градієнтами. Проте основною проблемою методу залишається його чутливість до вибору  $\eta$  та необхідність тривалого часу навчання для досягнення високої точності.

Метод Adam (Adaptive Moment Estimation) став популярним через здатність самостійно регулювати крок для кожного окремого параметра. Алгоритм відстежує перший момент  $m_t$  (середнє значення градієнтів) та другий момент  $v_t$  (оцінка дисперсії).

Дослідження показали, що класична реалізація  $L_2$ -регуляризації [5] в Adam призводить до того, що штраф за великі ваги потрапляє в рухомі середні градієнтів. Це викривляє статистику і не дає регуляризації працювати в належний спосіб. У модифікації AdamW було впроваджено роз'єднане затухання ваг (decoupled weight decay), де штраф застосовується безпосередньо до параметрів після обчислення адаптивного кроку:

$$\theta_{t+1} = \theta_t - \eta \left( \frac{\hat{m}_t}{\sqrt{\hat{v}_t + \epsilon}} \right) - \eta \lambda \theta_t,$$

де  $\lambda$  – коефіцієнт затухання ваг.

Такий підхід забезпечує стабільність регуляризації та дає змогу адаптивним методам конкурувати зі стабільністю класичного SGD на задачах класифікації зображень.

Квазі-ньютонівські методи, як-от L-BFGS, намагаються побудувати апроксимацію оберненої матриці Гессе, використовуючи історію попередніх оновлень. У стандартному вигляді вони вкрай чутливі до шуму, який виникає під час навчання на малих пакетах даних.

Модифікація mL-BFGS (Momentum-based L-BFGS) [4] вирішує цю проблему за допомогою двох ключових механізмів: стабілізації моменту та адаптивного демпфування. По-перше, замість використання миттєвих змін параметрів алгоритм обчислює експоненціально згладжені середні значень параметрів  $M_{\theta,t}$  та градієнтів  $M_{g,t}$ :

$$M_{\theta,t} = \beta M_{\theta,t-1} + (1 - \beta) \theta_t;$$

$$M_{g,t} = \beta M_{g,t-1} + (1 - \beta) g_t,$$

де коефіцієнт  $\beta$  зазвичай дорівнює 0.999.

Це дає змогу фільтрувати стохастичний шум перед оновленням історії L-BFGS.

По-друге, застосовується техніка демпфування (damping), яка коригує вектор зміни градієнта  $\hat{y}_k$ :

$$\hat{y}_k = \tau y_k + (1 - \tau) s_k.$$

Це гарантує, що апроксимована матриця кривизни залишається додатно визначеною, що є необхідною умовою для стабільної збіжності в не опуклих задачах. До того ж mL-BFGS використовує блочно-діагональний підхід, де матриця Гессе апроксимується окремо для різ-

них шарів мережі, що значно знижує витрати пам'яті та дає змогу розподіляти обчислення між різними вузлами системи.

Таблиця 1

### Основні характеристики чисельних методів оптимізації

Критерій	SGD (з моментом)	Adam / AdamW	L-BFGS / mL-BFGS
Порядок методу	Перший	Перший (адаптивний)	Другий (квазі-Ньютонівський)
Пам'ять	$O(d)$	$O(2d)$	$O(md)$
Чутливість до шуму	Середня	Низька	Висока (менша у mL-BFGS)
Швидкість збіжності	Повільна	Швидка на початку	Дуже швидка біля мінімуму

Ефективність методів було перевірено під час навчання архітектури ResNet-18 на наборі даних CIFAR-10 [6]. Результати демонструють суттєву різницю у швидкості збіжності та стабільності значень функції втрат (Loss).

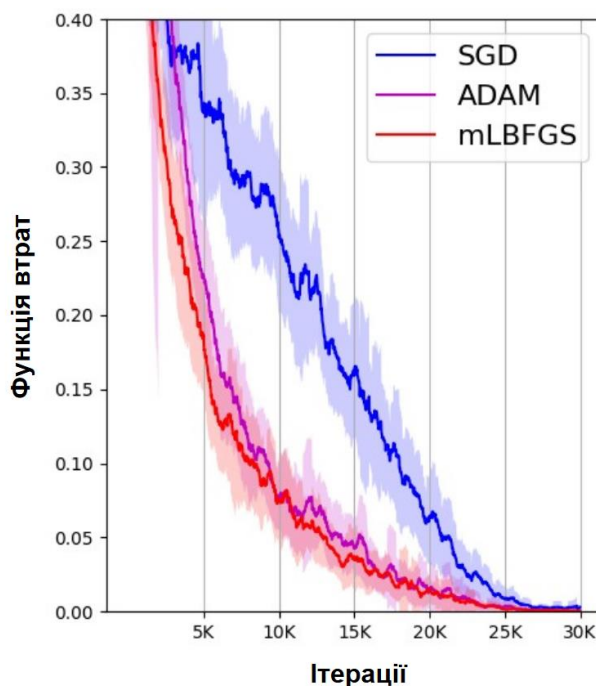


Рис. 1. Динаміка мінімізації функції втрат для оптимізаторів SGD, ADAM та mL-BFGS

На графіку представлена залежність значення функції втрат від кількості ітерацій навчання (до 30 000). Крива SGD (синій колір) демонструє найбільш повільне зниження помилки з великою амплітудою коливань, що підтверджується широкою затіненою областю дисперсії. Оптимізатор ADAM (фіолетовий колір) працює значно стабільніше, забезпечуючи швидкий спуск на початкових етапах. Найбільшу ефективність демонструє метод mL-BFGS (червоний колір), який забезпечує найшвидше падіння функції втрат та найнижчу фінальну похибку. Вже на позначці 15 000 ітерацій mL-BFGS досягає рівня помилки, якого SGD досягає лише після 25 000–30 000 ітерацій. Також помітно, що область дисперсії у mL-BFGS є найвужчою, що свідчить про високу стабільність апроксимації кривизни.

На рис. 2 відображено динаміку зниження помилки навчання залежно від реального часу роботи центрального процесора (CPU Time) в секундах. Цей графік дає змогу оцінити продуктивність алгоритмів, враховуючи обчислювальну складність кожної ітерації.

SGD (синя лінія) демонструє найнижчу ефективність у часовому вимірі; навіть після 300 секунд навчання помилка залишається на рівні вище  $10^{-1}$ . Adam (червона лінія) забезпечує швидкий початковий прогрес протягом перших 50 секунд, проте згодом виходить на плато на рівні помилки  $10^{-2}$ . Найкращі результати демонструє L-BFGS (чорна лінія), який протягом 250 секунд досягає рівня помилки нижче  $10^{-3}$ . Це підтверджує, що попри більшу обчислювальну вартість однієї ітерації, методи другого порядку завдяки точнішому кроку забезпечують значно швидшу wall-clock-збіжність до глибоких мінімумів.

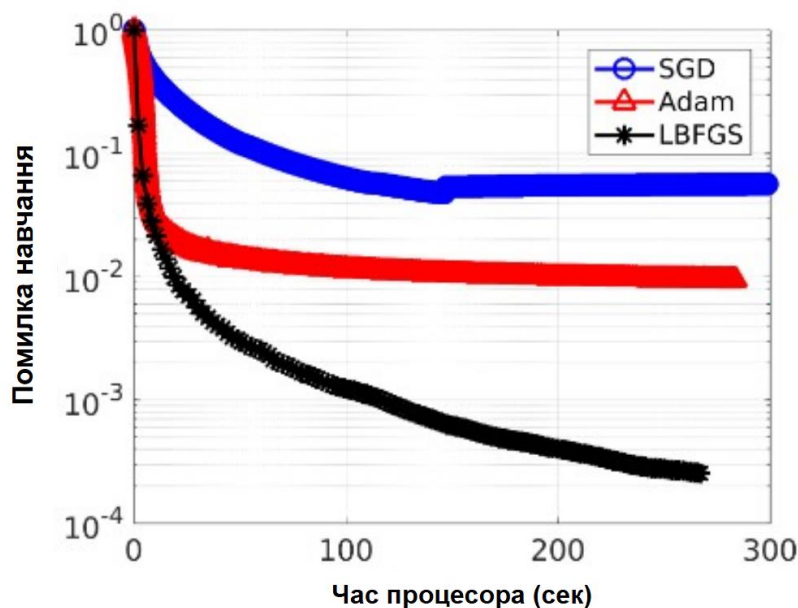


Рис. 2. Порівняння швидкості зниження помилки навчання (Training Error) залежно від реального часу роботи процесора (CPU Time)

**Узагальнення результатів порівняльного аналізу.** Результати порівняльного аналізу свідчать про суттєву перевагу методів другого порядку, зокрема L-BFGS, у швидкості та точності мінімізації функції втрат. Хоча алгоритм SGD є найменш вимогливим до обчислювальних ресурсів пам'яті, він демонструє найнижчу швидкість збіжності та високу амплітуду коливань помилки. Оптимізатор Adam забезпечує стабільне зниження похибки на початкових етапах, проте демонструє схильність до передчасного виходу на плато.

Найвищу ефективність за реальним часом роботи процесора демонструє метод L-BFGS, який досягає значно глибших мінімумів, порівняно з аналогами першого порядку. Головною перевагою цього підходу є здатність точно апроксимувати кривизну поверхні втрат, що мінімізує кількість необхідних ітерацій. Водночас основним обмежуючим фактором для методів другого порядку залишається підвищена обчислювальна складність однієї ітерації та обсяг використовуваної пам'яті. У підсумку, використання L-BFGS є найбільш виправданим для задач, де пріоритетом є досягнення максимальної точності навчання за мінімальний часовий інтервал.

**Перспективні напрями розвитку чисельної оптимізації.** Сучасні дослідження вказують на появу нових методів, що розвивають ідеї адаптивності та другого порядку. Алгоритм Sophia використовує полегшену діагональну оцінку Гессіана і демонструє прискорення навчання мовних моделей у 2 рази, порівняно з AdamW. Метод Lion використовує лише операції взяття знака градієнта, що дає змогу суттєво економити пам'ять GPU та прискорювати ітерації в розподілених системах. Також тривають дослідження щодо застосування квантових оптимізаторів (наприклад, COBYLA), які можуть бути ефективнішими для специфічних типів обчислювальних архітектур у майбутньому.

**Висновки.** Аналіз чисельних методів оптимізації дає змогу підсумувати результати дослідження. Використання методів першого порядку залишається виправданим для задач із низькими обчислювальними ресурсами, проте для масштабних моделей їх збіжність є недостатньою. Наукова новизна роботи полягає в обґрунтуванні ефективності механізму моменту для стабілізації квазі-ньютонівських методів (mL-BFGS), що допомагає використовувати інформацію другого порядку в стохастичних умовах без втрати стійкості. Практична цінність результатів полягає в можливості скорочення часу навчання моделей III вдвічі у разі переходу від адаптивних методів до mL-BFGS, що підтверджено експериментальними даними. Подальші дослідження будуть зосереджені на вивченні впливу автоматичного налаштування гіперпараметрів демпфування на стабільність навчання в умовах змінного розміру батча.

*Abstract.* This article explores the role of numerical optimization methods in the development of artificial intelligence systems. A comparative analysis of three widely used methods – SGD, AdamW, and mL-BFGS – is conducted. The paper examines the mathematical framework of each algorithm, the principles of working with first- and second-order gradients, and adaptability strategies. Particular attention is paid to stabilization mechanisms for approximating the Hessian matrix under stochastic noise conditions. The presented testing results confirm the effectiveness of second-order methods in accelerating model convergence.

*Keywords:* Numerical optimization, gradient descent, AdamW, mL-BFGS, Hessian matrix, convergence, neural networks.

#### СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Чисельні методи: навчальний посібник / Л. О. Волонтир, О. В. Зелінська, Н. А. Потапова, І. А. Чіков. Вінниця: ВНАУ, 2020. 322 с. URL: <https://r.donnu.edu.ua/handle/123456789/1805>
2. Bottou L., Curtis F., Nocedal J. Optimization Methods for Large-Scale Machine Learning. <https://epubs.siam.org/doi/10.1137/16M1080173>. URL: <https://epubs.siam.org/doi/10.1137/16M1080173> (дата звернення: 11.03.2026).
3. Rafati J., Roummel F. M. Quasi-Newton Optimization Methods For Deep Learning Applications. <https://arxiv.org/abs/1909.01994>. URL: <https://arxiv.org/abs/1909.01994> (дата звернення: 11.03.2026).
4. mL-BFGS: A Momentum-based L-BFGS for Distributed Large-Scale Neural Network Optimization – PMC. *PMC*. URL: <https://pmc.ncbi.nlm.nih.gov/articles/PMC12393816/> (дата звернення: 18.03.2026).
5. Venues | OpenReview. URL: <https://openreview.net/pdf/5963886abef941684ffc0cf670297e47fb1e5155.pdf> (дата звернення: 18.03.2026).
6. Niu Y. mL-BFGS: A Momentum-based L-BFGS for Distributed Large-Scale Neural Network Optimization. *researchgate*. URL: [https://www.researchgate.net/publication/372654300\\_mL-BFGS\\_A\\_Momentum-based\\_L-BFGS\\_for\\_Distributed\\_Large-Scale\\_Neural\\_Network\\_Optimization](https://www.researchgate.net/publication/372654300_mL-BFGS_A_Momentum-based_L-BFGS_for_Distributed_Large-Scale_Neural_Network_Optimization) (дата звернення: 19.03.2026).

UDC 528.8:581.5(477.44)

### SPATIOTEMPORAL DYNAMICS OF KNDVI IN VINNYTSIA CITY DURING THE VEGETATION GROWING SEASON: TREND ANALYSIS FOR 1990–2021

*A. O. Chernenko, A. M. Mischenko*

*Summary.* This study examines the spatiotemporal dynamics of urban vegetation in Vinnytsia over a 32-year period (1990–2021) using the Kernel Normalized Difference Vegetation Index (KNDVI) derived from Landsat satellite imagery within the Google Earth Engine (GEE) environment. Annual composite KNDVI maps were generated for the vegetation growing season (May–September) and analysed using pixel-wise linear regression to quantify interannual trends. To capture potential shifts in vegetation dynamics, the analysis was conducted separately for two sub-periods: 1990–2005 and 2005–2021. The results indicate a substantial increase in the proportion of statistically significant positive trends in the latter period, with the dominant pattern shifting from mixed (34.435 negative, 66.575 positive) to strongly positive (11.705 negative, 88.305 positive), suggesting pronounced urban greening. The fully scripted Python/GEE workflow provides a reproducible and scalable alternative to conventional desktop GIS approaches for long-term urban vegetation monitoring.

*Keywords:* KNDVI, urban vegetation, spatiotemporal analysis, Google Earth Engine, Vinnytsia.

**Introduction.** Urban vegetation provides a wide range of ecosystem services, including mitigation of the urban heat island effect, regulation of stormwater runoff, carbon sequestration, improvement of air quality, and support of biodiversity and human well-being [1; 2]. As urban areas continue to expand, the ability to monitor vegetation extent and condition at high spatial and temporal resolution becomes increasingly important for evidence-based urban planning and environmental management.

Remote sensing offers a cost-effective and spatially explicit approach for assessing vegetation dynamics over large areas and extended time periods [3]. Vegetation indices derived from multi-spectral satellite imagery—most notably the Normalized Difference Vegetation Index (NDVI)—have been widely used to monitor phenological dynamics, detect land-cover change, and quantify vegetation responses to climatic variability [4; 3]. However, NDVI is known to saturate in dense canopies and to be sensitive to soil background reflectance, which can limit its performance in heterogeneous urban landscapes [3; 5].

The Kernel Normalized Difference Vegetation Index (KNDVI), introduced by Camps-Valls et al. [5], addresses these limitations by embedding NDVI within a radial basis function kernel. This