* 1. **Переваги та недоліки застосування АЗПП для тренування нейронних мереж прямого поширення**

У книзі “Artificial Neural Networks: An Introduction”за  Kevin L. Priddy, Paul E. Keller[8 с.116], наводяться такі переваги алгоритму зворотного поширення помилки як “легке використання, адже не вимагає великої кількості параметрів для налаштування, алгоритм простий для реалізації, застосовний до широкого спектру проблем, популярний та часто застосовується для тренування нейронних мереж прямого поширення так само, як і для деяких рекурентних нейронних мереж”*.* Також автори наводять і недоліки даного алгоритму, а саме “важко сказати яка кількість шарів та нейронів є необхідною, навчання може бути повільним”.

В.А Головко, В.В Краснопрошин у виданні “Нейросетевые технологии оброботки данных” міститься перелік проблем, що можуть виникнути при використанні для навчання багатошарових нейронних мереж АЗПП, тобто “складність вибору потрібного кроку навчання [9]. Так, надто мала швидкість навчання збільшує час навчання і може привести до скочення нейронної мережі до ближнього локального мінімуму функції сумарної квадратичної помилки мережі. Велика швидкість навчання може привести до пропуску глобального мінімуму і зробити процес навчання розбіжним. Градієнтний метод не розрізняє точок локального та глобального мінімумів…”. Однак на противагу даному недоліку автори стверджують, що “у загальному випадку алгоритм зворотного поширення помилки не дозволяє досягти глобального мінімуму функції сумарної квадратичної помилки мережі. Однак це не применшує його плюсів, так як для багатьох практичних завдань точка глобального мінімуму на навчальній вибірці не є оптимальним рішенням задачі. Основне завдання навчання мережі - досягнення сумарної квадратичної помилки, при якій мережа має хорошу узагальнюючу здатністю, при цьому не має значення, який мінімум локальний або глобальний.”

Вплив випадкової ініціалізації вагових коефіцієнтів нейронної мережі на пошук мінімуму функції сумарної квадратичної помилки є ще одним ‘слабким місцем” алгоритму. Дана властивість означає, що при різній ініціалізації синаптичних зв'язків можуть виходити різні рішення задачі. Це характеризує нестійкість алгоритму навчання, коли нейронна мережа в одних випадках може навчатися до необхідної сумарної квадратичної помилки, а в інших ні[10].

Важливою перевагою алгоритму у час потреби швидкої обробки великих даних є те, що через кожен нейрон проходить інформація тільки по зв'язкових з ним нейронах, тому АЗПП легко реалізується на обчислювальних пристроях з паралельною архітектурою.

**2 АНАЛІЗ ДОСЛІДЖЕНЬ АЗПП. ГРАДІЄНТНИЙ СПУСК**

* 1. **Аналіз публікацій, досліджень АЗПП**

Як зазначається, сучасна версія АЗПП (також її називають зворотним режимом автоматичної диференціації) вперше була опублікована у 1970 році фінським студентом-магістром Сеппо Ліннаінмаа [11]. З того часу було проведено багато досліджень даного алгоритму.

Yann le Cun в своїй публікації “A Theoretical Framework for Back-Propagation” [12] презентує математичний фреймворк для вивчення АЗПП. В ньому backpropagation сформульований як оптимізаційна проблема без лінійних обмежень. Описаний метод, як зазначає сам автор “не тільки забезпечує чіткий спосіб диференціювання АЗПП на зразок процедур, але й значно спрощує саме диференціювання”.

Дослідження даного алгоритму включають і внесення до нього модифікацій. Наприклад, у статті “Іmprovements to the backpropagation algorithm” Mircea Petrini описує варіації алгоритму зворотного поширення помилки. Однією з варіацій він прагне зменшення кількості операцій з плаваючою точкою, тому що “АЗПП - це дорогий алгоритм, оскільки пряма реалізація заснована на операціях з плаваючою комою. Так як всі значення між 0 і 1 використовуються, виникають проблеми точності та стійкості, яких, як правило, можна уникнути, використовуючи арифметику з плаваючою комою з 32-бітною або 64-бітовою точкою. Існує кілька можливостей скоротити кількість необхідних операцій з плаваючою комою.” В своїй роботі автор приходить до висновку, що хоча АЗПП і є широко реалізований у найбільш практичних програмах ANN і працює досить добре, він страждає від проблеми повільної збіжності та збіжності до локальних мінімумів. “Це робить додатки, що використовують штучні нейронні мережі, дуже складним при вирішенні великих проблем”. [13]

Про зазначений вище недолік згадує в своїй публікації “A New Modified Back-Propagation Algorithm for Forecasting Malaysian Housing Demand” Nazri Mohd Nawi. Автор представляє “новий підхід до підвищення ефективності навчання алгоритмів BP для прогнозування попиту на житло в одному з штатів Малайзії. Запропонований алгоритм (BPM / AG) адаптивно змінює напрям пошуку на основі градієнта, вводячи значення параметра посилення у функції активації.” [14]. Отже, бачимо не тільки узагальнені варіації алгоритму, а й модифікації, що чітко направлені на вирішення певної задачі.

Наведені публікації лише частина загальної кількості досліджень АЗПП. Через свою популярність і частоту застосування у нейронних мережах алгоритм є дослідженим з різних можливих ракурсів.

* 1. **Проблема повільного навчання мережі**

Як згадувалося раніше(1.5) одним з недоліків застосування АЗПП для тренування нейронних мереж прямого поширення може бути те, що процес навчання відбувається повільно. При значних розмірах набору даних навчання може тривати дуже довго. Дана проблема вимагає вирішення в умовах сьогодення, коли швидка обробка великих даних є дуже важливою.

Шукають шляхи вирішення цієї проблеми автори публікації “Parallel Gradient Descent for Multilayer Feedforward Neural Networks”[15]. Вони представляють паралельний підхід до класифікації з використанням нейронних мереж як класу гіпотези. “У нейронних мереж може бути мільйони параметрів і вивчення оптимального значення всіх параметрів з величезних наборів даних при послідовній реалізації може бути дуже трудомістким завданням. У цій роботі ми реалізували паралельний градієнтний спуск для тренування багатошарової нейронних мереж прямого поширення. Зокрема, ми аналізуємо два види методів паралелізації: (а) паралельна обробка декількох прикладів тренувань через декілька потоків та (б) паралелізація матричних операцій для одного прикладу навчання.” В даній роботі досягнуто значних переваг шляхом паралелізації, а саме прискорення у 10 разів. Таким чином, підсумовуючи вище сказане можна ствердити про можливості успішного розпаралелення алгоритму градієнтного спуску.