

ĐĂNG KÝ ĐỀ TÀI NGHIÊN CỨU KHOA HỌC SINH VIÊN NĂM HỌC
2025 - 2026

1a. TÊN ĐỀ TÀI TIẾNG VIỆT

Tốc độ hội tụ của Gradient Descent trong tối ưu hóa hàm mất mát

1b. TÊN ĐỀ TÀI TIẾNG ANH

Convergence rate of Gradient Descent in optimizing loss functions

2 LĨNH VỰC

- ☒ Theo ngành học|
☐ Khoa học Giáo dục|
☐ Khởi nghiệp|
☐ Khoa học hành vi|
☐ Kinh doanh và quản lý|

3. GIẢNG VIÊN HƯỚNG DẪN

Họ và tên: TS. Lê Quang Minh

Đơn vị: Khoa Công nghệ Thông tin

Di động: 0326996473

Email: minh.lq@ou.edu.vn

4. CHỦ NHIỆM ĐỀ TÀI

Họ và tên: Bùi Thị Yến Nhi

Mã số sinh viên: 2251012103



Khoa: Đào tạo đặc biệt

Năm học: 2022-2026

Di động: 0978407636

Email: 2251012103nhi@ou.edu.vn

5. SINH VIÊN THAM GIA THỰC HIỆN ĐỀ TÀI

	Họ và tên	Mã số sinh viên	Nội dung nghiên cứu dự kiến được giao	Chữ ký
1.	Lê Trần Minh Phúc	2251010047	(Triển khai & Thí nghiệm): Chịu trách nhiệm chính về việc lập trình (code) các thuật toán tối ưu hóa, các hàm test/mô hình, thiết lập và chạy các thí nghiệm (bao gồm cả tuning). Viết chính phần mô tả triển khai trong Phương pháp, đóng gói code	
2.	Nguyễn Công Huy	2251012072	(Lý thuyết & Tổng hợp): Tập trung sâu vào việc đọc hiểu, phân tích, tổng hợp các kết quả lý thuyết về tốc độ hội tụ. Viết chính phần Mở đầu, Cơ sở lý thuyết, Tổng quan LSNC, Thảo luận (kết nối lý thuyết-thực nghiệm), Kết luận.	

6. GIỚI THIỆU Ý TƯỞNG NGHIÊN CỨU

Sự phát triển vượt bậc của Mạng nơ-ron sâu (Deep Neural Networks - DNNs) đã tạo ra những thành tựu đột phá trong nhiều lĩnh vực ứng dụng của Trí tuệ Nhân tạo, từ nhận dạng hình ảnh đến dịch máy tự động [1]. Khả năng học các biểu diễn dữ liệu phức tạp của DNNs phụ thuộc trực tiếp vào hiệu quả của quá trình huấn luyện, trong đó các thuật toán tối ưu hóa đóng vai trò trung tâm trong việc điều chỉnh hàng triệu, thậm chí hàng tỷ tham số mô hình nhằm cực tiểu hóa hàm mất mát. Thuật toán Gradient Descent (GD) cùng các biến thể ngẫu nhiên và thích ứng của nó, như Stochastic Gradient Descent (SGD), SGD with Momentum và Adam, đã trở thành những phương pháp được sử dụng phổ biến nhất trong thực nghiệm nhờ tính đơn giản và hiệu quả tương đối [2], [6]. Tuy nhiên, một thách

thức khoa học nền tảng phát sinh từ bản chất phi lồi (non-convexity) của hàm mất mát trong học sâu. Khác với các bài toán tối ưu hóa lồi truyền thống, hàm mất mát của DNNs thường cực kỳ phức tạp, đặc trưng bởi sự hiện diện của vô số điểm cực tiểu cục bộ, các điểm yên ngựa vốn rất phổ biến trong không gian nhiều chiều [3], và các vùng gần như bằng phẳng nơi gradient có giá trị rất nhỏ [4]. Tính phi lồi này làm mất hiệu lực các yếu tố đảm bảo hội tụ mạnh mẽ của GD trong môi trường lồi, dẫn đến khả năng thuật toán bị mắc kẹt tại các điểm dừng không mong muốn hoặc hội tụ rất chậm.

Trong bối cảnh đó, việc phân tích và hiểu rõ tốc độ hội tụ (convergence rate) của các thuật toán tối ưu hóa trở nên cực kỳ quan trọng. Tốc độ hội tụ, định lượng mức độ nhanh chóng mà thuật toán tiến gần đến một điểm dừng hoặc điểm cực tiểu cục bộ, là yếu tố then chốt quyết định hiệu quả tính toán, thời gian huấn luyện cần thiết và cuối cùng là chất lượng của mô hình học được [19]. Nhằm cải thiện tốc độ hội tụ trên các cảnh quan phi lồi phức tạp, nhiều biến thể của GD đã được đề xuất, từ các phương pháp tích hợp thông tin momentum (như SGD with Momentum [15], [16]) đến các phương pháp có tốc độ học thích ứng (như Adam [12]). Các phương pháp cải tiến này thường được kỳ vọng sẽ điều hướng hiệu quả hơn, thoát khỏi điểm yên ngựa nhanh hơn [8], [9] và đạt được tốc độ hội tụ tổng thể tốt hơn so với các phiên bản GD/SGD cơ bản.

Mặc dù đã có nhiều công trình lý thuyết phân tích tốc độ hội tụ dưới các giả định khác nhau, chẳng hạn như tính trơn Lipschitz (L-smoothness) hay điều kiện Polyak-Łojasiewicz (PL) [6], [10], [14], việc tổng hợp một cách có hệ thống các kết quả lý thuyết này, so sánh chúng và đặc biệt là đối chiếu với hiệu suất hội tụ quan sát được trong thực nghiệm vẫn còn là một khoảng trống cần được làm đầy. Quan trọng hơn, việc đánh giá hiệu quả không chỉ nên dừng lại ở tốc độ hội tụ cuối cùng mà còn cần đi sâu vào phân tích các đặc tính động học (dynamical properties) của quá trình tối ưu hóa. Cụ thể, cách thức mà các cơ chế cải tiến như momentum và tốc độ học thích ứng thực sự định hình quỹ đạo hội tụ, ảnh hưởng đến độ ổn định từng bước, và phản ứng với các cấu trúc hình học khác nhau, đặc biệt dưới sự điều khiển của các siêu tham số đặc trưng (β của Momentum, β_1/β_2 của Adam), là những khía cạnh cần được làm sáng tỏ hơn thông qua các phân tích thực nghiệm chi tiết và so sánh trực tiếp. Việc thiếu một sự đối sánh rõ ràng giữa các đảm bảo lý thuyết về tốc độ, hành vi động học thực tế, và hiệu năng cuối cùng khiến việc đánh giá khách quan lợi ích và cơ chế hoạt động cốt lõi của các thuật toán cải tiến như Momentum và Adam trở nên khó khăn [17], [18].

Vì vậy, nghiên cứu này được đề xuất với mục tiêu kép: thứ nhất, thực hiện một phân tích tổng hợp và so sánh các kết quả lý thuyết về tốc độ hội tụ của thuật

toán GD và các biến thể chính của nó (Momentum, Adam) trong bối cảnh tối ưu hóa hàm mất mát phi lồi; thứ hai, tiến hành các thực nghiệm có kiểm soát để so sánh hiệu suất hội tụ thực tế đồng thời thực hiện một phân tích so sánh chuyên sâu về động lực học hội tụ giữa các thuật toán GD/SGD cơ bản và các biến thể cải tiến Momentum và Adam trên các bài toán mẫu. Nghiên cứu sẽ đặc biệt tập trung vào việc khảo sát hệ thống và trực quan hóa ảnh hưởng của các siêu tham số đặc trưng (β , β_1 , β_2) lên các khía cạnh động học như quỹ đạo, tốc độ tức thời và độ ổn định, nhằm làm rõ sự khác biệt trong cơ chế hoạt động của Momentum và Adam. Thông qua việc kết hợp phân tích lý thuyết, đánh giá thực nghiệm về hiệu suất cuối cùng, và phân tích động học chi tiết được trực quan hóa, đề tài hướng tới cung cấp một cái nhìn toàn diện và sâu sắc hơn về cả tốc độ lẫn quá trình hội tụ, những khía cạnh cốt lõi quyết định hiệu quả của các phương pháp tối ưu hóa trong học sâu hiện đại.

7. MỤC TIÊU CỦA ĐỀ TÀI, ĐỐI TƯỢNG, PHẠM VI

Mục tiêu chính của nghiên cứu này là thực hiện một phân tích lý thuyết kết hợp đánh giá thực nghiệm về hiệu năng hội tụ, bao gồm cả tốc độ và các đặc tính động học, của thuật toán GD và các biến thể tiêu biểu là SGD with Momentum và Adam khi áp dụng vào bài toán tối ưu hóa hàm mất mát, đặc biệt là các hàm phi lồi thường gặp trong huấn luyện Mạng nơ-ron sâu. Để đạt được mục tiêu tổng quát này, nghiên cứu sẽ tập trung vào việc hệ thống hóa và phân tích các giả định toán học nền tảng, như tính trơn L-smoothness hay điều kiện Polyak-Łojasiewicz [10], vốn là cơ sở cho các phân tích tốc độ hội tụ trong môi trường phi lồi. Tiếp theo, nghiên cứu sẽ tiến hành khảo sát, tổng hợp và so sánh một cách có hệ thống các kết quả lý thuyết đã được công bố về tốc độ hội tụ, bao gồm cả tốc độ tiến đến các điểm dừng và tốc độ tiến đến các điểm cực tiểu cục bộ (hoặc điểm dừng bậc hai), cho thuật toán GD, SGD và các biến thể quan trọng như Momentum và Adam [12], làm rõ sự khác biệt về bậc hội tụ (ví dụ: cận tuyến tính so với tuyến tính) dưới các hệ giả định khác nhau [6], [14]. Song song với phần lý thuyết, nghiên cứu sẽ triển khai các thuật toán tối ưu hóa đã chọn, bao gồm ít nhất một thuật toán gradient cơ bản (như GD hoặc SGD) và các biến thể cải tiến là SGD with Momentum [15], [16] và Adam [12], và thiết kế các thí nghiệm để so sánh hiệu suất hội tụ thực tế cũng như phân tích chi tiết các đặc tính động học so sánh của chúng trên các hàm mục tiêu mẫu hoặc các mô hình học máy đơn giản. Quá trình phân tích kết quả thực nghiệm sẽ tập trung vào việc đối chiếu tốc độ hội tụ quan sát được với các dự đoán lý thuyết, đồng thời xem xét ảnh hưởng của việc tinh chỉnh siêu tham số, và đặc biệt là khảo sát, trực quan hóa, lý giải cách thức

các siêu tham số đặc trưng (β của Momentum; β_1 , β_2 của Adam) định hình quỹ đạo và hành vi hội tụ từng bước. Cuối cùng, nghiên cứu sẽ đưa ra những nhận định tổng kết về hiệu quả và ưu điểm tương đối của các phương pháp tối ưu hóa dựa trên cả bằng chứng lý thuyết về tốc độ hội tụ và phân tích thực nghiệm về hiệu năng và động lực học.

Đối tượng nghiên cứu chính của đề tài bao gồm các khía cạnh lý thuyết về tốc độ hội tụ tiệm cận và thực nghiệm về hiệu năng và động lực học hội tụ chi tiết. Về mặt lý thuyết, đối tượng là các định lý, bổ đề, chứng minh toán học và các giả định liên quan trực tiếp đến việc định lượng tốc độ hội tụ của các thuật toán tối ưu hóa dựa trên gradient. Về mặt thực nghiệm, đối tượng là hành vi hội tụ cụ thể, bao gồm cả tốc độ hội tụ tổng thể và các đặc tính động học như quỹ đạo, tốc độ tức thời, độ ổn định, được đo lường thông qua các chỉ số như giá trị hàm mất mát, chuẩn của gradient, và tọa độ tham số theo từng bước lặp (iteration), của các thuật toán được triển khai trên các bài toán xác định.

Phạm vi nghiên cứu được giới hạn như sau: Về lý thuyết, nghiên cứu tập trung vào việc tổng hợp và phân tích các đảm bảo tốc độ hội tụ tiệm cận dưới các giả định phổ biến trong tài liệu về tối ưu hóa phi lồi. Về thuật toán, phạm vi bao gồm GD, SGD, và hai biến thể đại diện là SGD with Momentum và Adam. Về hàm mục tiêu, phân tích lý thuyết áp dụng cho các lớp hàm phi lồi tổng quát thỏa mãn điều kiện nhất định, trong khi phần thực nghiệm sẽ ưu tiên sử dụng các hàm kiểm tra tổng hợp phi lồi 2 chiều (2D synthetic non-convex test functions) có các đặc tính hình học rõ ràng (ví dụ: thung lũng hẹp, điều kiện yếu) để thuận lợi cho việc trực quan hóa và phân tích động lực học chi tiết, mặc dù mô hình học máy đơn giản cũng có thể được xem xét. Phương pháp luận chủ yếu kết hợp giữa tổng quan, phân tích lý thuyết và thực nghiệm mô phỏng tập trung vào đánh giá hiệu năng hội tụ và phân tích động lực học so sánh, không đi sâu vào các kiến trúc DNN phức tạp hay các vấn đề về tổng quát hóa mặc dù mối liên hệ có thể được thảo luận [7], [19].

8. CƠ SỞ LÝ THUYẾT VÀ LỊCH SỬ NGHIÊN CỨU

Nghiên cứu về hiệu năng tối ưu hóa, bao gồm cả tốc độ hội tụ (convergence rate) và các đặc tính động học (dynamics) của các thuật toán tối ưu hóa dựa trên gradient được xây dựng trên nền tảng vững chắc của giải tích đa biến, lý thuyết tối ưu hóa, xác suất thống kê và được định hướng bởi bối cảnh ứng dụng trong học máy. Giải tích cung cấp các công cụ cơ bản như vector gradient (∇f) và ma trận Hessian ($\nabla^2 f$), vốn là chìa khóa để định hướng bước cập nhật và phân tích

độ cong cục bộ của hàm mục tiêu [1]. Đặc biệt, giả định về tính trơn Lipschitz của gradient (L-smoothness), tức là $\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\|$, đóng vai trò cực kỳ quan trọng trong hầu hết các phân tích tốc độ hội tụ, vì nó cho phép giới hạn sự thay đổi của hàm mục tiêu thông qua Bổ đề Suy giảm (Descent Lemma) [6]. Lý thuyết tối ưu hóa phân biệt rõ ràng giữa môi trường lồi và phi lồi. Trong tối ưu hóa lồi, các kết quả kinh điển đã thiết lập các tốc độ hội tụ chuẩn cho GD và SGD, ví dụ như tốc độ $O(\frac{1}{k})$ cho hàm lồi trơn hoặc tốc độ hội tụ tuyến tính (geometric rate) cho hàm lồi mạnh [2], [5]. Những kết quả này đóng vai trò tham chiếu quan trọng khi chuyển sang môi trường phi lồi phức tạp hơn.

Trong bối cảnh phi lồi của học sâu, lịch sử nghiên cứu về tốc độ hội tụ phản ánh nỗ lực đối mặt với các thách thức như điểm yên ngựa và vùng phẳng, những yếu tố không chỉ ảnh hưởng đến tốc độ đạt được điểm dừng mà còn định hình động lực học phức tạp của quá trình tối ưu hóa. Các phân tích ban đầu về tốc độ hội tụ thường chỉ đảm bảo sự hội tụ đến tập các điểm dừng (nơi gradient bằng không), với tốc độ hội tụ của chuẩn gradient bình phương $\|\nabla f(x_k)\|^2$ thường là $O(\frac{1}{k})$ đối với GD dưới giả định L-smoothness [6], [7]. Đối với SGD, tốc độ này thường chậm hơn và phụ thuộc vào cách giảm tốc độ học cũng như mức độ nhiễu của gradient ngẫu nhiên. Sự nhận thức về vai trò cản trở của các điểm yên ngựa [3] lên cả tốc độ lẫn quỹ đạo hội tụ đã thúc đẩy các nghiên cứu sâu hơn về khả năng thuật toán thoát khỏi chúng và hội tụ đến các điểm cực tiểu cục bộ hoặc các điểm dừng bậc hai (SOSPs). Các công trình tiên phong đã chỉ ra rằng GD có thêm nhiễu hoặc các thuật toán ngẫu nhiên có thể thoát khỏi các điểm yên ngựa không suy biến và hội tụ đến SOSPs với tốc độ hiệu quả trong thời gian đa thức [8, 9] đồng thời nhấn mạnh vai trò của sự ngẫu nhiên trong việc định hình động lực học thoát khỏi các điểm tới hạn này, mặc dù việc xác định tốc độ chính xác vẫn là một lĩnh vực nghiên cứu tích cực.

Song song đó, các nhà nghiên cứu đã khám phá các điều kiện cấu trúc đặc biệt trên hàm mất mát có thể đảm bảo tốc độ hội tụ nhanh hơn. Điều kiện Polyak-Łojasiewicz (PL) [10] là một ví dụ nổi bật, cho phép chứng minh tốc độ hội tụ tuyến tính toàn cục cho GD ngay cả khi hàm mục tiêu không lồi, một kết quả rất đáng mong đợi. Đồng thời, các nghiên cứu thực nghiệm và lý thuyết về đặc tính hàm mất mát DNN [4] đã cung cấp những hiểu biết sâu sắc về cấu trúc hình học phức tạp của nó, cho thấy các cực tiểu cục bộ thường có chất lượng tương đương và được kết nối với nhau [11], gợi ý rằng thách thức chính nằm ở việc điều hướng hiệu quả qua các điểm yên ngựa và vùng phẳng thay vì chỉ

tránh các cực tiểu "xấu". Sự thành công thực nghiệm của các phương pháp cải tiến như Momentum [15], [16] và đặc biệt là Adam [12] cũng đã thúc đẩy mạnh mẽ các nỗ lực phân tích lý thuyết về tốc độ hội tụ và hành vi động lực học của chúng. Các phân tích này cho thấy sự phức tạp hơn: ví dụ, cơ chế Momentum không chỉ có thể tăng tốc trong một số trường hợp mà còn được biết đến với khả năng làm muốt quỹ đạo và giảm dao động [15], [16], trong khi Adam, mặc dù hiệu quả thực tế thường thấy trong việc tăng tốc ban đầu và xử lý gradient khác biệt, đã đối mặt với các vấn đề lý thuyết về đảm bảo hội tụ trong một số tình huống [13] và hành vi động lực học ở giai đoạn cuối của nó cũng cần được xem xét cẩn trọng [17], [18]. Tốc độ hội tụ của chúng vẫn là chủ đề so sánh với SGD và với nhau [14]. Các bài báo tổng quan [6], [14] đóng vai trò quan trọng trong việc hệ thống hóa kiến thức và xác định các hướng nghiên cứu về tốc độ hội tụ trong tối ưu hóa cho học sâu. Nghiên cứu này kế thừa các nền tảng lý thuyết và lịch sử đó, nhưng sẽ mở rộng phạm vi phân tích thực nghiệm để bao gồm cả việc so sánh chi tiết các đặc tính động lực học của Momentum và Adam, đặc biệt là khảo sát ảnh hưởng của các siêu tham số đặc trưng (β , β_1 , β_2) lên quá trình hội tụ, bên cạnh việc tổng hợp, so sánh có hệ thống các kết quả về tốc độ hội tụ và đối chiếu chúng thông qua thực nghiệm.

9. PHƯƠNG PHÁP NGHIÊN CỨU

Để đạt được các mục tiêu đề ra, đề tài sẽ áp dụng một phương pháp luận nghiên cứu kết hợp chặt chẽ giữa phân tích lý thuyết và kiểm chứng thực nghiệm. Quá trình nghiên cứu bắt đầu bằng việc thực hiện một nghiên cứu và tổng quan tài liệu hệ thống (Systematic Literature Review) nhằm xác định và thu thập các công trình khoa học cốt lõi công bố các kết quả về tốc độ hội tụ cũng như các phân tích về cơ chế hoạt động và hành vi động lực học của GD, SGD, Momentum, Adam và các biến thể liên quan trong môi trường tối ưu hóa phi lồi. Trọng tâm sẽ đặt vào các định lý, giả định nền tảng (như L-smoothness, điều kiện PL [10]), kỹ thuật chứng minh liên quan đến các đảm bảo tốc độ, và các thảo luận về ảnh hưởng của siêu tham số lên quá trình tối ưu hóa [13], [17], [18]. Tiếp theo, phương pháp phân tích lý thuyết chuyên sâu (In-depth Theoretical Analysis) sẽ được tiến hành để hiểu rõ các định lý, vai trò của từng giả định, các yếu tố ảnh hưởng đến tốc độ hội tụ lý thuyết (hằng số Lipschitz L , phương sai gradient σ^2 , hằng số PL μ , các siêu tham số α , β , β_1 , β_2), và cơ sở lý thuyết cho các đặc tính động lực học dự kiến của Momentum và Adam.

Sau giai đoạn phân tích lý thuyết, phương pháp tổng hợp, so sánh và đánh giá phê bình (Synthesis, Comparison, and Critical Evaluation) sẽ được áp dụng. Giai đoạn này bao gồm việc hệ thống hóa các kết quả tốc độ hội tụ, đối chiếu các đảm bảo lý thuyết giữa các thuật toán khác nhau; ví dụ như, so sánh bậc hội tụ $O(\frac{1}{k})$ với $O(\rho^k)$ dưới các điều kiện tương đương, và đánh giá mức độ chặt chẽ cũng như tính thực tiễn của các giả định. Phần thực nghiệm của nghiên cứu sẽ bao gồm lựa chọn thuật toán và bài toán thí nghiệm, trong đó các thuật toán đại diện cho các phương pháp gradient cơ bản (GD/SGD) và các phương pháp cải tiến (như Momentum [15], [16] hoặc Adam [12]) sẽ được chọn cùng với các hàm kiểm tra tổng hợp phi lồi, ưu tiên loại 2 chiều (2D) để thuận lợi cho việc trực quan hóa và phân tích các đặc tính động học hoặc mô hình học máy đơn giản. Triển khai thuật toán sẽ được thực hiện bằng ngôn ngữ lập trình Python và các thư viện tính toán khoa học phổ biến, với yêu cầu cốt lõi là khả năng thu thập và lưu trữ chi tiết dữ liệu sau mỗi bước lặp (iteration) dưới dạng có cấu trúc, bao gồm giá trị hàm mất mát, chuẩn gradient và tọa độ tham số (đối với hàm 2D), để đảm bảo khả năng phân tích và tái tạo hiệu quả.

Thiết kế và thực hiện thí nghiệm (Experimental Design and Execution) là một phần quan trọng, bao gồm việc xác định các chỉ số đo lường hiệu năng và động lực học hội tụ (giá trị hàm mất mát, chuẩn gradient theo số vòng lặp/thời gian, quỹ đạo tham số) theo số bước lặp, thiết lập môi trường thí nghiệm nhất quán và đặc biệt là thực hiện khảo sát hệ thống ảnh hưởng của các siêu tham số đặc trưng (β cho Momentum; β_1 , β_2 cho Adam) lên quá trình hội tụ, thay vì chỉ tập trung vào tinh chỉnh để tìm giá trị tốt nhất cho một phép đo hiệu suất cuối cùng. Các thí nghiệm sẽ được lặp lại để đánh giá độ ổn định của các hành vi động học quan sát được. Cuối cùng, phân tích kết quả thực nghiệm và đối chiếu lý thuyết sẽ bao gồm việc trực quan hóa chi tiết dữ liệu động học (ví dụ: quỹ đạo 2D, đồ thị loss/gradient norm theo iteration), so sánh định lượng tốc độ hội tụ tổng thể quan sát được, phân tích chuyên sâu các đặc tính động học so sánh (độ mượt - smoothness, tốc độ tức thời - instantaneous rate/update magnitude, dao động - oscillations/fluctuations), và thảo luận về sự tương đồng hay khác biệt giữa kết quả thực nghiệm và các dự đoán từ lý thuyết [6], [14], đồng thời lý giải các hiện tượng động học dựa trên cơ chế hoạt động của Momentum và Adam [12], [15], [16], [17], [18]. Từ đó đưa ra các kết luận về hiệu quả và các đặc tính động học tương đối của các thuật toán. Toàn bộ quá trình và kết quả sẽ được trình bày một cách khoa học thông qua phương pháp mô tả (Descriptive Method) trong báo cáo cuối cùng.

10. ĐÓNG GÓP CỦA NGHIÊN CỨU

Nghiên cứu này, mặc dù không nhằm mục tiêu phát triển một thuật toán tối ưu hóa hoàn toàn mới, dự kiến sẽ đóng góp vào sự hiểu biết và hệ thống hóa kiến thức về các khía cạnh quan trọng của tối ưu hóa trong học sâu, đó là tốc độ hội tụ và các đặc tính động học của quá trình tối ưu. Về mặt học thuật, đóng góp chính nằm ở việc cung cấp một bản tổng hợp, phân tích so sánh và đánh giá có hệ thống các kết quả lý thuyết hiện có liên quan đến tốc độ hội tụ của các thuật toán tối ưu hóa dựa trên gradient phổ biến trong môi trường phi lồi. Công trình sẽ làm rõ mối liên hệ, sự khác biệt về các đảm bảo tốc độ lý thuyết giữa các phương pháp gradient cơ bản và các phương pháp cải tiến, đồng thời thảo luận về vai trò và tính thực tiễn của các giả định toán học nền tảng như L-smoothness hay điều kiện PL [10]. Thông qua phần thực nghiệm, nghiên cứu cung cấp bằng chứng định lượng về hiệu suất hội tụ tương đối của các thuật toán này trên các bài toán cụ thể, minh họa các ưu điểm tiềm năng về tốc độ của các phương pháp cải tiến như Adam [12] so với GD/SGD cơ bản trong các kịch bản được khảo sát, đồng thời, một đóng góp mới quan trọng là cung cấp các phân tích chi tiết và trực quan về động lực học hội tụ so sánh của SGD with Momentum và Adam, đặc biệt là làm sáng tỏ ảnh hưởng của các siêu tham số đặc trưng (β , β_1 , β_2) lên quỹ đạo và hành vi từng bước của chúng.

Một đóng góp quan trọng khác là nỗ lực kết nối giữa lý thuyết và thực hành. Bằng cách đối chiếu các đảm bảo tốc độ hội tụ lý thuyết [6], [14] và cơ chế hoạt động được đề xuất [12], [15], [16] với hành vi hội tụ chi tiết, bao gồm cả các đặc tính động học, quan sát được trong thực nghiệm, nghiên cứu góp phần làm sáng tỏ mức độ phù hợp của các phân tích lý thuyết và cách thức các yếu tố thực tế (như lựa chọn siêu tham số) ảnh hưởng đến hiệu suất và quá trình tối ưu hóa. Điều này có thể cung cấp những hiểu biết hữu ích và khách quan hơn cho các nhà nghiên cứu và kỹ sư thực hành khi lựa chọn, cấu hình và chuẩn đoán hành vi thuật toán tối ưu hóa [17], [18]. Về mặt giáo dục, việc thực hiện đề tài giúp sinh viên nâng cao năng lực nghiên cứu khoa học, kết hợp giữa tư duy phân tích toán học và kỹ năng thực nghiệm (bao gồm cả phân tích dữ liệu động học và trực quan hóa), đồng thời củng cố kiến thức về tối ưu hóa và học máy. Sản phẩm của đề tài, bao gồm báo cáo nghiên cứu với các phân tích động học chi tiết và mã nguồn thực nghiệm, có thể trở thành tài liệu tham khảo để tiếp cận và hữu ích cho cộng đồng học thuật và những người quan tâm đến nền tảng và hành vi thực tế của các phương pháp tối ưu hóa trong học sâu.

11. TÀI LIỆU THAM KHẢO

- [1] I. Goodfellow, Y. Bengio, and A. Courville, “Deep Learning,” www.deeplearningbook.org, 2016. <https://www.deeplearningbook.org>
- [2] S. Boyd and L. Vandenberghe, Convex Optimization. Cambridge University Press, 2004.
- [3] Y. Dauphin, R. Pascanu, C. Gulcehre, K. Cho, S. Ganguli, and Y. Bengio, “Identifying and attacking the saddle point problem in high-dimensional non-convex optimization,” arXiv:1406.2572 [cs, math, stat], Jun. 2014, Available: <https://arxiv.org/abs/1406.2572>
- [4] H. Li, Z. Xu, G. Taylor, C. Studer, and T. Goldstein, “Visualizing the Loss Landscape of Neural Nets,” arXiv:1712.09913 [cs, stat], Nov. 2018, Available: <https://arxiv.org/abs/1712.09913>
- [5] Y. Nesterov, Introductory Lectures on Convex Optimization. Springer Science & Business Media, 2013.
- [6] L. Bottou, F. E. Curtis, and J. Nocedal, “Optimization Methods for Large-Scale Machine Learning,” arXiv.org, Feb. 08, 2018. <https://arxiv.org/abs/1606.04838>
- [7] D. A. Roberts, “SGD Implicitly Regularizes Generalization Error,” arXiv.org, 2021. <https://arxiv.org/abs/2104.04874>
- [8] C. Jin, R. Ge, P. Netrapalli, Kakade, Sham M, and M. I. Jordan, “How to Escape Saddle Points Efficiently,” arXiv.org, 2017. <https://arxiv.org/abs/1703.00887>
- [9] R. Ge, F. Huang, C. Jin, and Y. Yuan, “Escaping From Saddle Points --- Online Stochastic Gradient for Tensor Decomposition,” arXiv.org, 2015. <https://arxiv.org/abs/1503.02101>
- [10] H. Karimi, J. Nutini, and M. Schmidt, “Linear Convergence of Gradient and Proximal-Gradient Methods under the Polyak- Łojasiewicz Condition,” arXiv.org, 2016. <https://arxiv.org/abs/1608.04636>
- [11] F. Draxler, K. Veschgini, M. Salmhofer, and F. A. Hamprecht, “Essentially No Barriers in Neural Network Energy Landscape,” arXiv.org, 2018. <https://arxiv.org/abs/1803.00885>

- [12] D. P. Kingma and J. Ba, “Adam: a Method for Stochastic Optimization,” arXiv.org, Dec. 22, 2014. <https://arxiv.org/abs/1412.6980>
- [13] S. J. Reddi, S. Kale, and S. Kumar, "On the convergence of Adam and beyond," in Proc. Int. Conf. Learn. Represent. (ICLR), 2018, Available: <https://arxiv.org/abs/1904.09237>
- [14] R. Sun, “Optimization for deep learning: theory and algorithms,” arXiv:1912.08957 [cs, math, stat], Dec. 2019, Available: <https://arxiv.org/abs/1912.08957>
- [15] B. T. Polyak, “Some methods of speeding up the convergence of iteration methods,” *USSR Computational Mathematics and Mathematical Physics*, vol. 4, no. 5, pp. 1–17, Jan. 1964, doi: 10.1016/0041-5553(64)90137-5.
- [16] I. Sutskever, J. Martens, G. Dahl, and G. Hinton, “On the importance of initialization and momentum in deep learning,” PMLR. Available: <https://proceedings.mlr.press/v28/sutskever13.html>
- [17] A. C. Wilson, R. Roelofs, M. Stern, N. Srebro, and B. Recht, “The Marginal Value of Adaptive Gradient Methods in Machine Learning,” arXiv.org. Available: <https://arxiv.org/abs/1705.08292>
- [18] L. Liu *et al.*, “On the Variance of the Adaptive Learning Rate and Beyond,” arXiv.org. Available: <https://arxiv.org/abs/1908.03265>
- [19] N. S. Keskar, D. Mudigere, J. Nocedal, M. Smelyanskiy, and P. T. P. Tang, “On Large-Batch Training for Deep Learning: Generalization Gap and Sharp Minima,” arXiv.org. Available: <https://arxiv.org/abs/1609.04836>

12. KẾ HOẠCH NGHIÊN CỨU

Tháng 1-2: Xây dựng nền tảng lý thuyết, thiết kế và triển khai ban đầu.

- **Tuần 1-4:** Nghiên cứu tổng quan & Lý thuyết nền tảng.
 - Tìm hiểu đề tài, xác định phạm vi nghiên cứu.
 - Nghiên cứu sâu về lý thuyết tốc độ hội tụ của các thuật toán GD, SGD, Momentum, Adam và các điều kiện liên quan (L-smoothness, PL...)
 - Chốt danh sách thuật toán và bài toán/mô hình cụ thể cho thực nghiệm.

- **Tuần 5-7:** Phân tích lý thuyết & Thiết kế thực nghiệm.
 - Phân tích, so sánh chi tiết các kết quả tốc độ hội tụ lý thuyết.
 - Thiết kế chi tiết các kịch bản, quy trình và metrics cho thí nghiệm.
 - Bắt đầu triển khai code các thuật toán và hàm/mô hình thử nghiệm.
- **Tuần 8:** Hoàn thiện triển khai & Thử nghiệm ban đầu.
 - Hoàn thiện code các thuật toán cơ bản, chạy thử nghiệm và debug.
 - Chuẩn bị môi trường và dữ liệu/script cho thí nghiệm.

Tháng 3-4: Thực hiện thí nghiệm, phân tích kết quả và hoàn thiện báo cáo.

- **Tuần 9-10:** Hoàn thiện hệ thống thực nghiệm.
 - Hoàn thiện code các thuật toán nâng cao và hệ thống logging kết quả.
 - Đảm bảo toàn bộ code và môi trường sẵn sàng cho thí nghiệm chính thức.
- **Tuần 11-13:** Chạy thí nghiệm.
 - Thực hiện tinh chỉnh siêu tham số (hyperparameter tuning).
 - Chạy các thí nghiệm chính thức với các thiết lập tối ưu để thu thập dữ liệu hội tụ.
- **Tuần 14-15:** Phân tích kết quả & Viết báo cáo.
 - Xử lý, trực quan hóa dữ liệu thực nghiệm.
 - Phân tích, so sánh hiệu suất hội tụ thực tế và đối chiếu với lý thuyết.
 - Viết các phần chính của báo cáo (Kết quả, Thảo luận, Kết luận).
- **Tuần 16(+):** Hoàn thiện & Báo cáo.
 - Hoàn thiện toàn bộ báo cáo/luận văn nghiên cứu.
 - Chuẩn bị tài liệu trình bày và bảo vệ kết quả cuối kỳ.