Training dynamics in modern neural network optimization

Rashid Turgunbaev
Kokand State University

Abstract: This article examines the critical role of training dynamics in the optimization of modern neural networks. Moving beyond a static analysis of algorithms and converged solutions, we argue that the time-evolving trajectory of parameters and internal representations - the dynamical system defined by the interaction of optimizer, architecture, and data - is fundamental to understanding contemporary deep learning phenomena. We analyze the journey from initialization through convergence, exploring how initial conditions implicitly regularize the optimization path, how stochasticity and loss landscape geometry interact to select flat minima, and how adaptive optimizers alter the fundamental search dynamics. The discussion extends to the co-evolution of internal representations, the emergent macroscopic scaling laws observed in large-scale training, and theoretical lenses such as the Neural Tangent Kernel that provide formal insight into these behaviors. Ultimately, a dynamic perspective is shown to be indispensable for explaining implicit regularization, generalization, and the emergence of capabilities, thereby guiding the development of more efficient, robust, and controllable optimization strategies for advanced artificial intelligence systems.

Keywords: training dynamics, optimization trajectory, implicit regularization, loss landscape, representation learning, stochastic gradient descent

The optimization of neural networks represents the central mechanism through which raw computational architecture acquires functional capability. For decades, the foundational algorithm for this process has been stochastic gradient descent and its numerous variants. While the static description of these algorithms is well-documented in literature, focusing on update rules and convergence guarantees, a profound shift in understanding has emerged from scrutinizing the dynamics of training - the intricate, time-evolving trajectory of parameters and representations throughout the learning process. This article posits that a dynamic, rather than static, perspective is crucial for unraveling the mysteries of modern neural network behavior, explaining phenomena such as implicit regularization, loss landscape navigation, and the emergence of robust internal representations. The journey from initialization to convergence is not a mere incremental refinement but a complex dynamical system exhibiting distinct phases and properties that fundamentally shape the final model.

The training process commences with initialization, a step whose profound influence on dynamics is now widely recognized. The shift from simplistic schemes like uniform or normal initialization to carefully calibrated methods like He or LeCun initialization was motivated by the desire to preserve signal variance across layers in deep networks at the onset of training. This initialization sets the initial conditions for the dynamical system. Crucially, in overparameterized networks - where parameters far exceed the number of training examples - the initialization point lies in a region of high-dimensional space. The gradient flow from this point is not uniquely determined by the loss function alone; the optimization path is heavily biased by the initial coordinates. This bias is a primary source of implicit regularization, where the optimization algorithm, guided by its dynamics, converges to a particular solution among the infinite set of interpolating solutions. The dynamics early in training often exhibit a rapid decline in loss, accompanied by a significant shift in parameters. However, this initial phase is more than just loss reduction. Research has shown that networks often learn simple, generalized patterns first, a dynamic sometimes aligned with the notion of spectral bias where lower-frequency functions are learned faster. The initial trajectory is thus a search directed by gradients but constrained and biased by the geometry of the initialization region.

As training progresses beyond the initial rapid improvement, networks typically enter a prolonged period of gradual loss refinement. The dynamics in this phase are characterized by the interplay between stochasticity from minibatch sampling and the geometry of the loss landscape. The concept of a loss landscape, often visualized as a high-dimensional surface, is inherently linked to dynamics. Optimization algorithms do not merely seek the lowest valley; they navigate a topology riddled with saddle points, flat plateaus, and ravines. The dynamics of stochastic gradient descent exhibit a fascinating property: the noise introduced by minibatch sampling is not merely an impediment but a critical component of effective training. This stochasticity prevents convergence to sharp minima, instead encouraging movement towards wider, flatter basins in the loss landscape. Flat minima are empirically associated with better generalization, as small perturbations in parameters lead to negligible changes in loss, suggesting robustness. The training dynamics effectively perform an approximate Bayesian inference, with the trajectory and final resting point encoding a form of uncertainty. The continuous injection of gradient noise means the parameters do not converge to a fixed point in a traditional sense but rather exhibit persistent fluctuations within a basin. The covariance of these fluctuations relates to the sharpness of the minimum, and thus the dynamics in this late-phase training are essential for selecting generalizable solutions.

A pivotal concept for understanding these dynamics is the learning rate, which acts not as a mere step size but as a central governor of the dynamical system's behavior. It controls the influence of the stochastic noise and determines the trade-off between the speed of descent and the stability of the final solution. A high learning rate can cause the optimization trajectory to oscillate or even diverge, while a too-low rate leads to agonizingly slow progress and potential capture in suboptimal regions. The development of adaptive learning rate methods like Adam or RMSProp can be interpreted as introducing stateful dynamics to the optimization process. These methods maintain per-parameter moving averages of past gradients, effectively creating a friction-like or momentum-based component to the dynamical system. Momentum, in particular, transforms the gradient descent equation into a second-order differential equation akin to a ball rolling down a hillside with inertia. This allows the optimization to traverse flat plateaus more quickly and dampen oscillations in steep ravines, fundamentally altering the training trajectory compared to vanilla gradient descent. The choice of optimizer is thus a choice of dynamical system, each with different properties regarding memory, noise resilience, and convergence speed.

The dynamics of optimization are inextricably linked to the evolving internal representations within the network. As parameters change, so does the function that maps inputs to intermediate activations and finally to outputs. Recent studies probing these representation dynamics reveal a non-monotonic, rich evolution. Features do not solidify immediately but can undergo periods of rapid restructuring followed by consolidation. This is particularly evident in transfer learning, where a pre-trained network undergoes fine-tuning. The dynamics in this setting are constrained, starting from a point of relatively low loss, and the optimization explores a local basin shaped by the pre-trained parameters. The dynamics of feature learning also highlight the difference between modern deep networks and shallow models. In deep networks, gradient flow via backpropagation facilitates a coordinated, though complex, adjustment across layers. The phenomenon of catastrophic forgetting in continual learning scenarios is a direct consequence of unconstrained training dynamics - the natural trajectory of gradient descent when applied to new data erases the representations needed for previous tasks. Techniques like elastic weight consolidation attempt to modify the dynamics by adding a regularization term that penalizes movement away from important parameters for previous tasks, effectively shaping the loss landscape to preserve critical basins.

Furthermore, the scale of models and data has brought to light emergent dynamical phenomena. In large language models and vision transformers, training dynamics exhibit predictable, scalable patterns. The loss during training often follows a power-law decay with respect to computational budget and data volume, a observation that has led to the formulation of scaling laws. These laws are a macroscopic description of the training dynamics, suggesting underlying order in the complex optimization process. The dynamics also reveal phase transitions; for instance, as training computation increases, models may suddenly acquire new capabilities or exhibit sharp improvements on certain tasks, indicating a shift in the internal representation structure. The double descent curve, where test error increases and then decreases as model size or training time grows past the interpolation threshold, is another dynamic phenomenon. It contradicts classical bias-variance trade-off wisdom and is explained by observing the optimization dynamics through the lens of the time at which training is stopped. Early stopping in the overfitting regime can yield the first peak, while continued training allows the dynamics to find a smoother interpolating solution in the second descent.

Theoretical frameworks for analyzing these dynamics are continually evolving. The neural tangent kernel theory provides a powerful lens by approximating the training dynamics of infinitely wide networks as a linear system. In this regime, the NTK remains constant during training, and the dynamics reduce to a convex optimization problem, solvable analytically. While this fails to capture the feature learning dynamics of finite-width networks, it offers a precise mathematical tool for understanding the early stages of training and the convergence properties in ultra-wide networks. Another perspective comes from physics, where training dynamics are analogized to particles moving in a potential field (the loss landscape) under the influence of stochastic forces (minibatch noise). This thermodynamic analogy allows researchers to apply concepts from statistical mechanics, such as temperature (controlled by the learning rate and batch size) and entropy, to analyze the collective behavior of parameters.

In conclusion, the training dynamics of modern neural networks constitute a rich field of study that moves beyond the algorithmic checklist of optimization. It demands a perspective that views training as a trajectory through a high-dimensional space, shaped by initial conditions, governed by differential equations with stochastic forcing, and resulting in the emergent acquisition of complex functions. The dynamics explain why we obtain the models we do, how they generalize, and where their capabilities and limitations arise. Future progress in developing more efficient, robust, and capable AI systems will likely hinge on a deeper, more refined understanding of these dynamics. This includes designing optimizers that better shape the training trajectory, developing principled methods for scheduling learning rates and selecting batch sizes, and formulating new theoretical models that can predict dynamic phenomena like phase transitions and representation evolution. The optimization process is the crucible in which intelligence is forged in artificial neural networks; understanding its fire and flow is key to mastering the alchemy of deep learning.

References

1. Turgunbaev, R. (2025). CORE DATABASE COMPETENCIES FOR LIBRARY STUDENTS. European Review of Contemporary Arts and Humanities, 1(5), 104-111.

2. Jurayev, I. (2025). A SYSTEMATIC PEDAGOGICAL APPROACH TO MAQOM ENCULTURATION IN UNDERGRADUATE VOCAL STUDIES. European Review of Contemporary Arts and Humanities, 1(5), 34-41.

3. Ktaybekova, D. J. (2025). METHODOLOGICAL APPROACHES APPROPRIATE TO AGE CHARACTERISTICS IN THE ACQUISITION OF DUTOR CREATIVITY AND PERFORMANCE SKILLS. European Review of Contemporary Arts and Humanities, 1(5), 9-14.

4. Berdikhanova, D. M. (2025). PEDAGOGICAL MECHANISMS FOR DEVELOPING STUDENTS'ARTISTIC THINKING IN DUTOR PERFORMANCE. European Review of Contemporary Arts and Humanities, 1(5), 4-8.

5. Sattarov, F. I. (2025). THE FORMATION OF COLORS IN NATURE AND THEIR CHARACTERISTICS OF CHANGE. European Review of Contemporary Arts and Humanities, 1(5), 70-73.

6. Kurbanova, M. M. (2025). COMPARATIVE ANALYSIS OF EASTERN AND WESTERN ORCHESTRAL TRADITIONS. European Review of Contemporary Arts and Humanities, 1(5), 55-59.

7. Isaqov, S. S. (2025). NONVERBAL COMMUNICATION AND COHESION IN THE ORCHESTRAL ENSEMBLE. European Review of Contemporary Arts and Humanities, 1(5), 112-115.

8. Narmatova, A. P. (2025). DEVELOPING COUNTRY-SPECIFIC CULTURAL COMPETENCE IN GERMAN LANGUAGE LEARNERS THROUGH LITERARY TEXTS. European Review of Contemporary Arts and Humanities, 1(5), 97-103.

9. Akhrorova, M. (2025). THE CONCEPT OF TRAGIC FATE IN COMPARATIVE LITERATURE: EASTERN AND WESTERN PERSPECTIVES. European Review of Contemporary Arts and Humanities, 1(5), 46-50.

10. qizi Oxunjonova, G. Z. O. (2025). METHODS OF DEVELOPING STUDENTS'SPATIAL THINKING ABILITIES IN DRAWING LESSONS. European Review of Contemporary Arts and Humanities, 1(5), 87-91.

11. qizi Rahimjonova, N. B. (2025). EFFECTIVENESS OF USING DIGITAL TECHNOLOGIES IN DRAWING. European Review of Contemporary Arts and Humanities, 1(5), 78-82.

12. qizi Boyigitova, G. B. (2025). CONNECTIONS AND ITS TYPES. European Review of Contemporary Arts and Humanities, 1(5), 65-69.

13. qizi Xabibullayeva, K. Z. (2025). SHEARING AND ITS IMPORTANCE IN DRAWINGS. European Review of Contemporary Arts and Humanities, 1(5), 116-121.

14. qizi Amoniddinova, I. B. (2025). RESTORATION OF PAINTING. European Review of Contemporary Arts and Humanities, 1(5), 51-54.

15. qizi G'aybullayeva, T. O. (2025). STAGES OF CREATING A COMPLEX STILL LIFE COMPOSITION IN PAINTING. European Review of Contemporary Arts and Humanities, 1(5), 74-77.

16. Turg'unboy qizi Karimjonova, D. (2025). FINE ART OF UZBEKISTAN. European Review of Contemporary Arts and Humanities, 1(5), 29-33.

17. Qurbonova, S. (2025). ANALYSIS OF ORNAMENTATION IN UZBEK MAQOM VOCAL PERFORMANCE. European Review of Contemporary Arts and Humanities, 1(5), 15-19.

18. Muminov, M. (2025). TRANSMISSION OF SHASHMAQOM REPERTOIRE IN CONTEMPORARY MASTER-STUDENT RELATIONSHIPS. European Review of Contemporary Arts and Humanities, 1(5), 60-64.

19. Xakimxo'ja qizi Odilova, M. (2025). MODERN APPROACH AND IMPORTANCE OF DRAWING TEACHING. European Review of Contemporary Arts and Humanities, 1(5), 83-86.

20. Ibragimov, S. (2025). TEACHING ACADEMIC WRITING WITH CHATGPT: OPPORTUNITIES AND LIMITATIONS. European Review of Contemporary Arts and Humanities, 1(5), 92-96.

21. ogli Muqimov, S. Z. (2025). MUSIC AND NEUROPHYSIOLOGY: HOW DOES MUSIC CHANGE BRAIN ACTIVITY?. European Review of Contemporary Arts and Humanities, 1(3), 3-7.

22. oglu Muqimov, S. Z. (2025). INTERPRETING REPETITION AND VARIATION IN DIGITAL MUSIC: FROM ALGORITHMS TO ARTISTIC EXPRESSION. European Review of Contemporary Arts and Humanities, 1(3), 8-13.

23. Turgunbaev, R. (2025). From Perception to Action with Integrated VLA Systems. Technical Science Integrated Research, 1(6), 11-17.

24. Kamilov, K. M., Khakimova, Z. K., Shermatov, A. A., & Jakhongirov, I. J. (2025). Digital monitoring and diagnostic optimization of ocular pathologies in the elderly population. Technical Science Integrated Research, 1(6), 23-28.

25. Uzoqjonov, M., & Mirzakarimova, M. (2025). The role and significance of artificial intelligence in modern society. Technical Science Integrated Research, 1(6), 8-10.

26. Ahmedov, A. B. (2025). The impact of artificial intelligence on production efficiency in the digital economy. Technical Science Integrated Research, 1(6), 3-7.

27. kizi Akhmatjonova, U. L. (2025). Hidden contour line: meaning and use. Technical Science Integrated Research, 1(6), 18-22.

28. Egamberganova, Z. (2025). The digitally empowered library. Technical Science Integrated Research, 1(5), 29-35.

29. Turgunbaev, R. (2025). A project-based learning framework for teaching distributed data processing. Technical Science Integrated Research, 1(5), 3-7.

30. Makhkamova, G. T. (2025). Clinical and diagnostic features of respiratory diseases of combined etiology in children. Technical Science Integrated Research, 1(5), 21-24.

31. Xo‘jjiyevev, M. Y., & Bozorova, F. J. R. (2025). The role of pumpkin seed oil in slowing lipid oxidation in functional beverages and its kinetic models. Technical Science Integrated Research, 1(5), 25-28.

32. Tajibayev, B. J. (2025). Methods of providing psychological assistance to children after an earthquake. Technical Science Integrated Research, 1(5), 36-38.

33. qizi Valijonova, H. S. (2025). Innovative approach to the implementation of bim technologies ın architectural and urban planning projects. Technical Science Integrated Research, 1(5), 12-16.

34. qizi Nabijonova, C. I. (2025). The formation of pilgrimage tourism under the influence of natural factors. Technical Science Integrated Research, 1(5), 17-20.

35. Urazmatov, J., & Raxmatullayev, O. R. (2025). The impact of preferential loans on private entrepreneurship, small business possibilities expansion factor. Technical Science Integrated Research, 1(4), 3-6.

36. To‘rayeva, D. M. (2025). Developing Students’ Communicative Skills through Extra-Linguistic Sources. Technical Science Integrated Research, 1(4), 7-10.

37. Mullayeva, M. K. (2025). Ways to develop speech culture in future teachers through poetic works. Technical Science Integrated Research, 1(4), 11-14.

38. ogli Juraboyev, A. T. (2025). Organization of recreational facilities in the mountainous territories of Uzbekistan. Technical Science Integrated Research, 1(4), 15-19.

39. Ergasheva, N. N., & kizi Ruzikulova, K. N. (2025). Features of the course of iron deficiency conditions and risk factors for their development in children of the first year of life. Academic Journal of Science, Technology and Education, 1(7), 26-29.

40. Shomuradova, D. A. (2025). Effective use of authentic materials in developing professional communicative skills. Academic Journal of Science, Technology and Education, 1(7), 53-55.

41. Ahmedov, A. B. (2025). The emergence of digital platforms in Uzbekistan's economy and their impact on market competition. Academic Journal of Science, Technology and Education, 1(7), 6-10.

42. Ergasheva, N. N., & Ergashaliyev, S. D. (2025). The role of gut microflora in postoperative adaptation and necrotizing enterocolitis risk following congenital intestinal obstruction. Academic Journal of Science, Technology and Education, 1(7), 56-59.

43. Abdusalimova, A. (2025). Improving the effectiveness of fine arts lessons through innovative pedagogical technologies. Academic Journal of Science, Technology and Education, 1(7), 36-39.

44. Asatullayev, R. B., & Latifova, S. N. (2025). The lymphatic system. Academic Journal of Science, Technology and Education, 1(7), 43-44.

45. Jurayeva, M. A. (2025). The formation of Uzbek journalism at the beginning of the 20th century. Academic Journal of Science, Technology and Education, 1(7), 63-65.

46. Xo'jjiyevev, M. Y., & Bozorova, F. J. R. (2025). Antioxidant properties of pumpkin seed oil addition to functional drinks. Academic Journal of Science, Technology and Education, 1(7), 40-42.

47. Yuldashev, K. (2025). Genre and performance style in theater art. Academic Journal of Science, Technology and Education, 1(7), 45-49.

48. oglu Nematov, F. N. (2025). Building effective collaboration between teachers and parents in primary school. Academic Journal of Science, Technology and Education, 1(7), 60-62.

49. Nasirova, G. R. (2025). Understanding structural changes in the respiratory tract in chronic disease. Academic Journal of Science, Technology and Education, 1(6), 82-85.

50. Masharipova, S. A. (2025). Prospects for the use of digitalization in the effective financial management of the chemical industry. Academic Journal of Science, Technology and Education, 1(6), 95-98.

51. Ergasheva, G. N. (2025). Storytelling method in teaching English to preschool children. Academic Journal of Science, Technology and Education, 1(6), 63-65.

52. Kamalova, A. (2025). Shukur Khollmirzaev's essay "That person is a mentor, and I am a disciple", its composition, plot, and system of characters. Academic Journal of Science, Technology and Education, 1(6), 48-50.

53. Mahmudova, D. Q. (2025). Effectiveness of microfertilizers in corn cultivation. Academic Journal of Science, Technology and Education, 1(6), 51-52.

54. Tokhtayev, I., & Ganiyeva, F. (2025). Opportunities for organic potato cultivation. Academic Journal of Science, Technology and Education, 1(6), 66-69.

55. Eraliyeva, M. T. K. (2025). Constitutional rights of citizens in the field of social security and their guarantees. Academic Journal of Science, Technology and Education, 1(6), 53-54.

56. oglu Idiyev, B. B., & Khujakulov, K. R. (2025). Synthesis and kinetic regularities of copolymers based on styrene and nitrogen-containing methacrylic monomers. Academic Journal of Science, Technology and Education, 1(6), 99-104.

57. qizi Haydarova, S. A. (2025). Electromagnetism. Academic Journal of Science, Technology and Education, 1(6), 34-35.

58. qizi Kenjayeva, Z. S. (2025). Advantages of modern methodology in forming phonetic competence in primary school students. Academic Journal of Science, Technology and Education, 1(6), 59-62.

59. qizi Xurramova, S. Q. (2025). Neurolinguistics: Comparative study of language processing in English and Uzbek. Academic Journal of Science, Technology and Education, 1(6), 12-15.

60. qizi Vakilova, S. T. (2025). Technological factors influencing the antioxidant activity of mulberry leaf tea. Academic Journal of Science, Technology and Education, 1(6), 45-47.

61. Santos, A. J. (2025). The Role of Behavioral Insights in Economic Decision Making Education. Academic Journal of Science, Technology and Education, 1(5), 39-42.

62. oglu Abdurakhmonov, D. M. (2025). Fungal diseases of grapes and measures to combat them. Academic Journal of Science, Technology and Education, 1(5), 43-45.