American Journal of Data Science and Machine Learning
ISSN: - applied

Deep Neural Networks as Dynamical Systems: Expressivity, Controllability, and High Dimensional Learning

[1] Adrian Kovarik
[1] Department of Applied Mathematics, Charles University, Czech Republic

Received: 16th Oct 2025 | Received Revised Version: 27th Oct 2025 | Accepted: 02th Nov 2025 | Published: 15th Nov 2025

Volume 01 Issue 01 2025 | Crossref DOI: 10.64917/ajdsml/V01I01-005

Abstract

Deep neural networks have evolved from heuristic pattern recognition tools into mathematically grounded systems whose theoretical understanding spans approximation theory, geometry, dynamical systems, and optimal control. This article develops a unified theoretical framework that interprets classical and modern neural architectures through the lenses of geometric separability, statistical learning, and dynamical systems theory. Drawing exclusively from foundational and contemporary contributions in machine learning, control theory, and high dimensional geometry, we provide a comprehensive synthesis of the evolution from perceptrons and support vector machines to residual networks and neural ordinary differential equations.

The study begins by examining early geometric formulations of classification, including linear threshold units and shattering properties, and situates these within modern capacity analysis. We then analyze multilayer feedforward networks in terms of universal approximation and storage capacity, addressing both width and depth considerations. Special attention is devoted to the power of depth and residual connections, emphasizing the reinterpretation of deep networks as discretized dynamical systems.

A central contribution of this work is an integrative exploration of neural ordinary differential equations and their mean field optimal control formulations. We explain how continuous depth models unify discrete architectures and reveal new insights into controllability, interpolation, and long time behavior. The mean field perspective connects parameter learning with population level dynamics, clarifying the role of measure theoretic interpolation and turnpike phenomena in training trajectories.

We further investigate geometric and topological perspectives, including manifold learning and invertible architectures, demonstrating how controllability conditions determine expressive power in neural ODE frameworks. Stability considerations and identity preserving structures are analyzed to explain empirical success in deep training regimes. Optimization landscape properties, stochastic gradient methods, and automatic differentiation are contextualized within this broader dynamical view.

Finally, the article synthesizes classical statistical learning theory with modern transformer based dynamics and cluster formation in self attention systems, positioning deep learning as a theory of measure evolution under learned flows. By integrating insights from approximation theory, control, geometry, and statistical learning, this work provides a publication ready theoretical narrative that clarifies both the mathematical foundations and the emerging research directions of deep learning as a dynamical systems discipline.

Keywords: Deep learning, neural ordinary differential equations, universal approximation, mean field optimal control, expressivity, geometric separability.

Cite This Article: Adrian Kovarik. 2025. Deep Neural Networks as Dynamical Systems: Expressivity, Controllability, and High Dimensional Learning. American Journal of Data Science and Machine Learning 1, 01, 25-30. https://doi.org/10.64917/ajdsml/V01I01-005

1. Introduction

The development of neural networks reflects a gradual deepening of mathematical insight into learning systems. Early models such as the perceptron introduced by Rosenblatt in 1958 framed learning as a probabilistic process of linear separation in feature space (Rosenblatt, 1958). The perceptron represented a fundamental breakthrough by demonstrating that classification could be automated through parameter adaptation. Yet its limitations were equally formative, as it was restricted to linearly separable data.

The geometric interpretation of linear classifiers was further elaborated through the study of systems of linear inequalities. Cover demonstrated the combinatorial properties of linear threshold systems and quantified the number of linearly inducible orderings in Euclidean space (Cover, 1965; Cover, 1967). These results linked learning capacity with geometric dimension and laid the groundwork for understanding shattering and separability.

Parallel developments in statistical pattern recognition were initiated by Fisher, who formalized linear discriminant analysis as a method for maximizing class separability under Gaussian assumptions (Fisher, 1936). This statistical approach connected classification with variance decomposition and probabilistic modeling, influencing later margin based methods.

The 1990s saw the emergence of support vector networks, which unified geometric margin maximization with convex optimization principles (Cortes and Vapnik, 1995). These methods emphasized structural risk minimization and introduced a principled framework for generalization in high dimensional settings. Meanwhile, combinatorial investigations into shattering complexity and arrangements of hyperplanes deepened theoretical understanding of capacity and separation (Freimer et al., 1991; Boland and Urrutia, 1995; Houle, 1993).

The curse and blessing of dimensionality, articulated by Donoho, reframed high dimensional spaces as simultaneously challenging and rich with structure (Donoho, 2000). This duality remains central to modern deep learning: high dimensional parameter spaces enable expressive models but complicate generalization analysis.

The renaissance of deep learning was catalyzed by multilayer architectures trained with gradient based methods. LeCun and colleagues demonstrated the effectiveness of convolutional networks for document recognition (LeCun et al., 1998), and later surveys synthesized progress in hierarchical representation learning (LeCun et al., 2015). The universal approximation capabilities of multilayer perceptrons were rigorously characterized, including bounds on hidden neuron counts and storage capacity (Huang and Huang, 1991; Huang, 2003).

A pivotal theoretical development was the recognition that depth confers representational advantages. Eldan and Shamir proved that certain functions require exponentially many neurons in shallow networks but can be efficiently represented by deeper ones (Eldan and Shamir, 2015). Lin and Jegelka further demonstrated that residual architectures with minimal hidden width can achieve universal approximation (Lin and Jegelka, 2018). These results indicated that architectural structure, not merely parameter count, governs expressive power.

The reinterpretation of deep networks as discretizations of differential equations provided a transformative conceptual shift. Haber and Ruthotto analyzed stable architectures through continuous dynamical systems analogies (Haber and Ruthotto, 2017). He and colleagues introduced residual learning frameworks that implicitly approximate identity mappings, facilitating training in very deep models (He et al., 2016). Hardt and Ma emphasized the importance of identity components in preserving gradient flow (Hardt and Ma, 2017).

The formal introduction of neural ordinary differential equations established a continuous depth limit for residual networks (Chen et al., 2018). In this paradigm, layer evolution becomes the trajectory of a controlled dynamical system. E proposed viewing machine learning through dynamical systems theory, advocating continuous time formulations (E, 2017). Subsequent work framed deep learning as a mean field optimal control problem, linking parameter dynamics with control trajectories (E et al., 2018).

Recent advances have extended this framework to measure theoretic interpolation and long time control behavior.

Cuchiero and collaborators established universal interpolation properties for controlled ODEs (Cuchiero et al., 2020). Geshkovski and Zuazua analyzed turnpike phenomena in optimal control for deep residual networks (Geshkovski and Zuazua, 2022). Variational and measure theoretic approaches to neural ODE training have clarified existence and optimality conditions (Bonnet et al., 2023; Isobe and Okumura, 2024).

Simultaneously, approximation and controllability analyses have been refined. Li, Lin, and Shen investigated deep learning through dynamical approximation theory (Li et al., 2022). Cheng and colleagues studied interpolation and controllability of deep neural networks, highlighting connections between expressivity and control reachability (Cheng et al., 2025). Elamvazhuthi and co authors characterized universal approximation on manifolds using geometric controllability conditions (Elamvazhuthi et al., 2023).

The theoretical landscape has further expanded to include invertible architectures (Ishikawa et al., 2023), optimization landscape geometry in deep convolutional networks (Nguyen and Hein, 2018), sparsity in long time control of neural ODEs (Esteve Yague and Geshkovski, 2023), and cluster formation in self attention dynamics (Geshkovski et al., 2023). Measure to measure interpolation using transformer architectures suggests a generalization of flow based learning beyond classical ODE frameworks (Geshkovski et al., 2024).

This article synthesizes these developments into a coherent theoretical narrative. The central thesis is that deep neural networks are best understood as controlled dynamical systems operating in high dimensional geometric spaces, where expressivity is governed by separability, controllability, and measure evolution. The literature gap addressed here lies in the absence of an integrated exposition connecting classical geometric learning theory with modern mean field and dynamical systems formulations. By bridging these domains, we aim to clarify foundational principles and outline future research trajectories.

2. Methodology

The methodology of this research is theoretical and integrative. It consists of a systematic reinterpretation of established results under a unifying dynamical systems perspective. Rather than introducing new empirical experiments, the study develops conceptual synthesis through detailed comparative analysis of foundational works.

First, classical geometric learning models are re examined. The perceptron is interpreted as a static linear separator whose capacity corresponds to hyperplane arrangements in Euclidean space (Rosenblatt, 1958; Cover, 1965). We analyze separability through combinatorial and geometric arguments, incorporating k separability notions and weak separation algorithms (Duch, 2006; Houle, 1993). The methodology here is interpretive, relating combinatorial shattering results to expressive capacity.

Second, multilayer networks are analyzed through universal approximation and storage capacity bounds. Upper and lower bounds on hidden units are discussed in relation to functional interpolation capabilities (Huang and Huang, 1991; Huang, 2003). The depth versus width trade off is explored using complexity results from Eldan and Shamir and Lin and Jegelka, emphasizing constructive representational hierarchies (Eldan and Shamir, 2015; Lin and Jegelka, 2018).

Third, we adopt a dynamical systems methodology. Residual networks are described as forward Euler discretizations of continuous flows (He et al., 2016; Haber and Ruthotto, 2017). Neural ODE frameworks are then examined as limits of infinite depth architectures (Chen et al., 2018). The training process is reformulated as an optimal control problem where parameters act as control functions (E et al., 2018; Li et al., 2018).

Fourth, mean field formulations are incorporated to transition from finite parameter models to distributional dynamics (Bonnet et al., 2023; Isobe and Okumura, 2024). The methodology involves conceptual translation from particle based parameterization to measure valued evolution, clarifying interpolation properties (Cuchiero et al., 2020).

Fifth, controllability and manifold approximation analyses are synthesized. Geometric control conditions are used to determine reachable representations (Elamvazhuthi et al., 2023; Cheng et al., 2025). Invertibility and universal approximation properties of invertible networks are examined to understand bijective transformations (Ishikawa et al., 2023).

Finally, optimization dynamics are contextualized through stochastic gradient methods and automatic differentiation (Kingma and Ba, 2014; Paszke et al., 2017). The optimization landscape perspective of deep convolutional networks is integrated to explain convergence properties (Nguyen and Hein, 2018).

Throughout, the methodological principle is interpretive coherence. Each cited contribution is situated within a single theoretical continuum that progresses from geometric separation to measure evolution.

3. Results

The integrative analysis yields several principal findings.

First, expressive capacity in neural networks is fundamentally geometric. Early linear models established that separability depends on dimension and hyperplane arrangements (Cover, 1967). Multilayer networks extend separability through hierarchical partitioning, effectively increasing the complexity of induced decision regions (Huang, 2003). Depth enables exponential representational compression compared to shallow networks (Eldan and Shamir, 2015).

Second, residual architectures derive their strength from identity preserving dynamics. The inclusion of skip connections stabilizes gradient propagation and approximates continuous time flows (He et al., 2016; Hardt and Ma, 2017). Stability analyses confirm that such architectures correspond to well posed differential systems (Haber and Ruthotto, 2017).

Third, neural ODE frameworks reveal that deep learning is a controlled dynamical system. The continuous formulation unifies depth and time, demonstrating that representation learning corresponds to evolving features along trajectories (Chen et al., 2018; E, 2017). Optimal control perspectives show that training corresponds to minimizing a cost functional over control paths (E et al., 2018).

Fourth, mean field analysis indicates that large scale neural systems approximate measure valued dynamics. Interpolation and approximation can be understood at the distributional level, providing universal interpolation properties (Cuchiero et al., 2020; Bonnet et al., 2023).

Fifth, controllability conditions determine universality on manifolds. When the underlying vector fields satisfy geometric reachability properties, neural ODEs can approximate functions defined on embedded manifolds (Elamvazhuthi et al., 2023). This extends classical universal approximation theorems to geometric domains.

Sixth, optimization landscapes in deep networks exhibit benign structures under certain conditions, explaining empirical trainability (Nguyen and Hein, 2018). Adaptive optimization algorithms such as Adam enhance convergence in high dimensional parameter spaces (Kingma and Ba, 2014).

Finally, recent transformer based and self attention models can be interpreted as measure evolution systems, where clusters emerge dynamically (Geshkovski et al., 2023; Geshkovski et al., 2024). This generalizes the dynamical paradigm beyond ODE based architectures.

4. Discussion

The findings suggest that deep learning theory has converged toward a unified dynamical systems paradigm. Classical geometry established the language of separation and shattering. Modern depth results clarified hierarchical expressivity. Continuous time formulations integrated these insights into flow based interpretations.

A major implication concerns generalization. While classical statistical learning theory emphasized margin and complexity control (Cortes and Vapnik, 1995), the dynamical perspective suggests that stability and controllability play complementary roles. Stable flows may implicitly regularize learning trajectories.

Limitations of the current synthesis include reliance on theoretical frameworks without empirical validation within this article. Additionally, while mean field theory clarifies large scale limits, finite width phenomena require further investigation.

Future research should explore deeper connections between geometric invariant theory and symmetry properties in neural representations (Mumford et al., 1994). The interaction between sparsity in long time control and generalization remains an open question (Esteve Yague and Geshkovski, 2023). Moreover, transformer based measure interpolation frameworks suggest a broader class of flow models that transcend classical ODE formulations (Geshkovski et al., 2024).

5. Conclusion

Deep neural networks have evolved from simple linear separators into sophisticated dynamical systems capable of universal approximation and measure transformation. By integrating geometric separability, approximation theory, optimal control, and mean field analysis, this article provides a comprehensive theoretical account of expressivity and controllability in deep learning. The dynamical systems perspective not only unifies past developments but also illuminates future directions in understanding learning as controlled evolution in high dimensional spaces.

References

1. Bengio, Y., Ducharme, R., Vincent, P., and Janvin, C. A neural probabilistic language model. Journal of Machine Learning Research, 3:1137 to 1155, 2003.
2. Boland, R. P., and Urrutia, J. Separating collections of points in Euclidean spaces. Information Processing Letters, 53(4):177 to 183, 1995.
3. Bonnet, B., Cipriani, C., Fornasier, M., and Huang, H. A measure theoretical approach to the mean field maximum principle for training NeurODEs. Nonlinear Analysis, 227:113161, 2023.
4. Breiman, L. Random forests. Machine Learning, 45(1):5 to 32, 2001.
5. Chen, R. T. Q., Rubanova, Y., Bettencourt, J., and Duvenaud, D. Neural ordinary differential equations. In Proceedings of the 32nd International Conference on Neural Information Processing Systems, 6572 to 6583, 2018.
6. Cheng, J., Li, Q., Lin, T., and Shen, Z. Interpolation, approximation, and controllability of deep neural networks. SIAM Journal on Control and Optimization, 63(1):625 to 649, 2025.
7. Cortes, C., and Vapnik, V. Support vector networks. Machine Learning, 20(3):273 to 297, 1995.
8. Cover, T. M. Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition. IEEE Transactions on Electronic Computers, EC 14(3):326 to 334, 1965.
9. Cover, T. M. The number of linearly inducible orderings of points in d space. SIAM Journal on Applied Mathematics, 15(2):434 to 439, 1967.
10. Cuchiero, C., Larsson, M., and Teichmann, J. Deep neural networks, generic universal interpolation, and controlled ODEs. SIAM Journal on Mathematics of Data Science, 2(3):901 to 919, 2020.
11. Donoho, D. High dimensional data analysis: The curses and blessings of dimensionality. AMS Mathematical Challenges Lecture, 1 to 32, 2000.
12. Duch, W. K separability. In ICANN 06 Proceedings of the 16th International Conference on Artificial Neural Networks, Springer, 188 to 197, 2006.
13. E, W. A proposal on machine learning via dynamical systems. Communications in Mathematical Statistics, 5:1 to 11, 2017.
14. E, W., Han, J., and Li, Q. A mean field optimal control formulation of deep learning. Research in the Mathematical Sciences, 6(1):10, 2018.
15. Elamvazhuthi, K., Zhang, X., Oymak, S., and Pasqualetti, F. Learning on manifolds: Universal approximation properties using geometric controllability conditions for neural ODEs. Proceedings of the 5th Annual Learning for Dynamics and Control Conference, 211:1 to 11, 2023.
16. Eldan, R., and Shamir, O. The power of depth for feedforward neural networks. JMLR Workshop and Conference Proceedings, 49:1 to 34, 2015.
17. Esteve Yague, C., and Geshkovski, B. Sparsity in long time control of neural ODEs. Systems and Control Letters, 172:105452, 2023.
18. Fisher, R. A. The use of multiple measurements in taxonomic problems. Annals of Eugenics, 7(7):179 to 188, 1936.
19. Freimer, R., Mitchell, J. S. B., and Piatko, C. On the complexity of shattering using arrangements. Technical Report, Cornell University, 1991.
20. Geshkovski, B., Letrouit, C., Polyanskiy, Y., and Rigollet, P. The emergence of clusters in self attention dynamics. Advances in Neural Information Processing Systems, 36:57026 to 57037, 2023.
21. Geshkovski, B., Rigollet, P., and Ruiz Balet, D. Measure to measure interpolation using transformers. arXiv:2411.04551, 2024.
22. Geshkovski, B., and Zuazua, E. Turnpike in optimal control of PDEs, ResNets, and beyond. Acta Numerica, 31:135 to 263, 2022.
23. Haber, E., and Ruthotto, L. Stable architectures for deep neural networks. Inverse Problems, 34(1):014004, 2017.
24. Hardt, M., and Ma, T. Identity matters in deep learning. International Conference on Learning Representations, 1:1627 to 1640, 2017.
25. He, K., Zhang, X., Ren, S., and Sun, J. Deep residual learning for image recognition. IEEE Conference on Computer Vision and Pattern Recognition, 770 to 778, 2016.
26. Houle, M. F. Algorithms for weak and wide separation of sets. Discrete Applied Mathematics, 45(2):139 to 159, 1993.
27. Huang, G. B. Learning capability and storage capacity of two hidden layer feedforward networks. IEEE Transactions on Neural Networks, 14(2):274 to 281, 2003.
28. Huang, S. C., and Huang, Y. F. Bounds on the number of hidden neurons in multilayer perceptrons. IEEE Transactions on Neural Networks, 2(1):47 to 55, 1991.
29. Ishikawa, I., Teshima, T., Tojo, K., Oono, K., Ikeda, M., and Sugiyama, M. Universal approximation property of invertible neural networks. Journal of Machine Learning Research, 24(287):1 to 68, 2023.
30. Kingma, D. P., and Ba, J. Adam: A method for stochastic optimization. arXiv:1412.6980, 2014.

31. LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P. Gradient based learning applied to document recognition. Proceedings of the IEEE, 86(11):2278 to 2324, 1998.

32. LeCun, Y., Bengio, Y., and Hinton, G. Deep learning. Nature, 521:436 to 444, 2015.

33. Li, Q., Chen, L., Tai, C., and E, W. Maximum principle based algorithms for deep learning. Journal of Machine Learning Research, 18(165):1 to 29, 2018.

34. Li, Q., Lin, T., and Shen, Z. Deep learning via dynamical systems: An approximation perspective. Journal of the European Mathematical Society, 25(5):1671 to 1709, 2022.

35. Lin, H., and Jegelka, S. ResNet with one neuron hidden layers is a universal approximator. Advances in Neural Information Processing Systems, 31:6169 to 6178, 2018.

36. Massaroli, S., Poli, M., Park, J., Yamashita, A., and Asama, H. Dissecting neural ODEs. Advances in Neural Information Processing Systems, 33:3952 to 3963, 2020.

37. Mumford, D., Fogarty, J., and Kirwan, F. Geometric Invariant Theory. Springer, 1994.

38. Nguyen, Q., and Hein, M. Optimization landscape and expressivity of deep CNNs. Proceedings of the 35th International Conference on Machine Learning, 3730 to 3739, 2018.

39. Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., Lin, Z., Desmaison, A., Antiga, L., and Lerer, A. Automatic differentiation in PyTorch. NIPS 2017 Workshop on Autodiff, 2017.

40. Rosenblatt, F. The perceptron: A probabilistic model for information storage and organization in the brain. Psychological Review, 65(6):386, 1958.