

NTK vs Feature Learning, Spectral Bias, and Functional-Analytic Perspectives

Shreyas Kalvankar

August 25, 2025

1 Feature Learning Beyond the NTK Regime

1.1 Background

The Neural Tangent Kernel (NTK) linearizes neural networks around initialization and predicts training dynamics for infinitely wide networks [Jacot et al., 2018, Lee et al., 2019]. Empirically, however, trained networks often outperform their linearized or fixed-kernel counterparts, suggesting that real networks move beyond the “lazy” regime and engage in non-trivial *feature learning*.

1.2 Main Ideas in the Literature

Beyond Linearization. Bai and Lee [2020] analyze higher-order Taylor approximations of over-parameterized networks to capture behavior beyond NTK. They show that (i) higher-order terms can govern training, and (ii) suitable randomization can decouple models from their NTK linearization and couple them to quadratic (or higher-order) models, with implications for expressivity and sample complexity.

Statistical Perspective on Feature Learning. Zhang et al. [2024] argue that the advantage of deep networks arises from *adaptive feature learning*, which fixed-kernel approaches such as NTK cannot capture. They propose prototypical models (e.g. over-parameterized Gaussian sequence models) to study feature adaptation and its generalization benefits.

1.3 Future Directions / Open Questions

- **Transition to feature learning:** When (as a function of width, depth, learning rate, and data) do higher-order terms dominate over NTK dynamics?
- **Provable gains:** Can we mathematically prove cases where feature learning requires fewer training examples than fixed-kernel methods such as NTK? In other words, can we show that adaptive features always help reduce sample complexity in certain problems?
- **Operator viewpoint:** In NTK, the kernel (similarity between data points) stays fixed, but in real networks it evolves during training. Can we describe this evolving similarity as a changing operator in function space, and analyze its dynamics to better understand how features adapt?

2 Spectral Bias and Frequency Dynamics

2.1 Background

During training, neural networks tend to learn *low-frequency* (low-complexity) components first, and only later fit high-frequency modes, a phenomenon known as *spectral bias* [Rahaman et al., 2019]. This perspective helps explain why overparameterized models can generalize despite their capacity to fit noise.

2.2 Main Ideas in the Literature

Empirical Fourier Analysis. Rahaman et al. [2019] analyze ReLU networks via Fourier tools and show: (i) low frequencies are learned earlier, (ii) they are more robust to parameter perturbations, and (iii) manifold geometry can make high frequencies easier to learn by reparameterization effects.

Rigorous NTK-based Characterization. Cao et al. [2020] decompose training dynamics along the eigenfunctions of the NTK: components associated with larger eigenvalues converge faster, implying that lower-complexity (often low-frequency) components are learned first. For inputs uniform on the sphere, lower-degree spherical harmonics are learned earlier.

2.3 Future Directions / Open Questions

- **Beyond the lazy regime:** Current theory explains spectral bias only when networks behave like fixed kernels (NTK). What happens when networks move beyond this regime and actively change their features during training?
- **Data geometry:** Real data lies on low-dimensional manifolds (e.g. image manifolds). How does the shape of these manifolds affect which frequencies are learned first?
- **Architectural effects:** Do different architectures (CNNs, Transformers, different activations) show different kinds of spectral bias, and if so, why?
- **Practical leverage:** Can spectral bias be exploited to design better training strategies, such as curriculum learning, early stopping, or frequency-based regularizers?

3 Functional Analytic / Operator-Theoretic Perspectives

3.1 Background

Functional analysis characterizes networks as elements of function spaces (RKHS/RKBS), making inductive biases explicit and analyzable via operators, spectra, and norms.

3.2 Main Ideas in the Literature

Inductive Bias of NTKs. Bietti and Mairal [2019] investigate the Reproducing Kernel Hilbert Space (RKHS) associated with the Neural Tangent Kernel (NTK), including convolutional variants. They analyze properties such as smoothness, spectral approximation, and stability to input deformations. Their results show that while NTK-RKHSs have strong approximation capabilities, their smoothness and stability properties are weaker: the kernel maps are not Lipschitz, but satisfy only weaker Hölder-type continuity.

RKBS and Radon-based Norms. Bartolucci et al. [2023] extend the functional analytic study of neural networks from reproducing kernel Hilbert spaces (RKHS) to reproducing kernel Banach spaces (RKBS), focusing on one-hidden-layer ReLU networks. They establish representer theorems using integral representations of such networks, and show that the associated RKBS norm can be characterized via Radon transforms as the total variation of certain measures. This provides an alternative functional framework for studying neural networks, distinct from the Hilbert space setting.

3.3 Future Directions / Open Questions

- **Banach vs Hilbert:** When do RKBS norms better capture implicit biases of gradient descent than RKHS norms?
- **Evolving kernels:** Track NTK eigenspectra during finite-width training and relate to approximation/stability trade-offs.
- **Invariances:** Link functional norms and deformation stability to data invariances (translations/rotations/manifolds).
- **Gradient flows in function space:** View training as flows in RKHS/RKBS and study resulting regularization.

References

- Y. Bai and J. D. Lee. Beyond linearization: On quadratic and higher-order approximation of wide neural networks. In *International Conference on Learning Representations (ICLR)*, 2020. URL <https://openreview.net/pdf?id=rk1lGyBFPH>.
- F. Bartolucci, E. De Vito, L. Rosasco, and S. Vigogna. Understanding neural networks with reproducing kernel banach spaces. *Applied and Computational Harmonic Analysis*, 62:194–236, 2023. doi: 10.1016/j.acha.2022.08.006. URL <https://www.sciencedirect.com/science/article/pii/S1063520322000768>.
- A. Bietti and J. Mairal. On the inductive bias of neural tangent kernels. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2019. URL <https://papers.nips.cc/paper/9449-on-the-inductive-bias-of-neural-tangent-kernels>.
- Y. Cao, Z. Fang, Y. Wu, D.-X. Zhou, and Q. Gu. Towards understanding the spectral bias of deep learning. *arXiv preprint arXiv:1912.01198*, 2020. URL <https://arxiv.org/abs/1912.01198>.
- A. Jacot, F. Gabriel, and C. Hongler. Neural tangent kernel: Convergence and generalization in neural networks. In *Advances in Neural Information Processing Systems (NeurIPS)*, pages 8580–8589, 2018. URL <https://proceedings.neurips.cc/paper/8076-neural-tangent-kernel-convergence-and-generalization-in-neural-networks.pdf>.
- J. Lee, L. Xiao, S. S. Schoenholz, Y. Bahri, R. Novak, J. Sohl-Dickstein, and J. Pennington. Wide neural networks of any depth evolve as linear models under gradient descent. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2019. URL <https://proceedings.neurips.cc/paper/2019/file/0d1a9651497a38d8b1c3871c84528bd4-Paper.pdf>.

- N. Rahaman, A. Baratin, D. Arpit, F. Draxler, M. Lin, F. A. Hamprecht, Y. Bengio, and A. Courville. On the spectral bias of neural networks. In *Proceedings of the 36th International Conference on Machine Learning (ICML)*, volume 97 of *Proceedings of Machine Learning Research*, pages 5301–5310, 2019. URL <https://proceedings.mlr.press/v97/rahaman19a.html>.
- H. Zhang, J. Lai, Y. Li, Q. Lin, and J. S. Liu. Towards a statistical understanding of neural networks: Beyond the neural tangent kernel theories. *arXiv preprint arXiv:2412.18756*, 2024. URL <https://arxiv.org/abs/2412.18756>.