Training and Generalization of Overparametrized Neural Networks

Shreyas Kalvankar

October 9, 2025

Master's Thesis Proposal Meeting

Agenda

- Go over the thesis proposal, feedback on the approach and tentative plan
- Discuss goals and set expectations for the first stage review
- Discuss communication channels and schedule for weekly meetings

Motivation

- Overparametrized neural networks can perfectly fit data yet generalize well.
- The **Neural Tangent Kernel (NTK)** provides a framework for analyzing training dynamics at infinite width.
- However, the NTK remains fixed during training — no feature learning.
- Recent results (e.g. Hanin–Nica, 2019) show that when depth and width co-scale, finite-width networks exhibit evolving but *stable* kernels for small $\beta = d/n$, marking a weak feature-learning regime between lazy and fully non-linear training.
- **Goal:** Understand the “weak feature learning” regime where kernel evolution is small but significant.

Research Question

Main Question:

How can we characterize the evolution of the NTK in finite-width and finite-depth neural networks, and what does this imply for effective function spaces and generalization?

Sub-questions:

1. How can NTK evolution be written as $K_t = K_0 + \Delta K_t$?
2. How does the size of ΔK_t scale with $\beta = d/n$?
3. How can we interpret the evolving RKHSs $(\mathcal{H}_t)$ induced by K_t?

Background Overview

- **NTK framework:** training dynamics $\dot{f}_t = -K(f_t - y)$ with fixed kernel in infinite width.
- **Finite-width correction:** NTK is random and evolves; its variance scales as $\exp(c\beta)$.
- **RKHS perspective:** each kernel defines an RKHS $\mathcal{H}_t$; evolving kernels imply changing spaces.

Understanding kernel drift = understanding feature learning beyond lazy training.

Research Gap and Scope

Gaps:

- Lack of perturbative link between kernel drift and dynamics.
- Limited understanding of β thresholds for regime transitions.
- No formal interpretation of evolving RKHSs.

Scope:

- Analytical focus: perturbation expansions and scaling analysis.
- Empirical validation on small-scale JAX experiments (Hanin–Nica diagnostics).

Methodology Overview

1. **Perturbation Analysis:** Expand $K_t = K_0 + \Delta K_t$ and study its effect on $\dot{f}_t = -K_t(f_t - y)$.
2. **Scaling Regimes:** Analyze how $\beta = d/n$ separates lazy, weak, and unstable training regimes.
3. **Functional Analysis View:** Interpret training as motion through a family of RKHSs ($\mathcal{H}_t$) to study generalization.

Expected Outcomes

- Perturbation-theoretic model of NTK evolution.
- Characterization of training regimes via β scaling.
- Functional-analytic interpretation of kernel drift as evolving hypothesis spaces.
- Empirical validation of theoretical predictions on finite-width networks.

Tentative Timeline

Weeks 1–2: Study NTK theory; reproduce Hanin–Nica results in JAX.
Weeks 3–6: Develop perturbation expansion $K_t = K_0 + \Delta K_t$.
Weeks 7–10: Analyze dependence on $\beta = d/n$; identify regime boundaries.
Weeks 11–14: Study evolving RKHSs $(\mathcal{H}_t)$ and generalization effects.

Deliverables: Interim report + preliminary empirical findings for first-stage review.

Discussion Points

- Are the proposed analytical directions realistic for the first-stage timeline?
- How to balance theoretical and experimental parts?
- Feedback on feasibility of RKHS interpretation.

Thank you!

Looking forward to your feedback.