EXP003: Mode-wise Decay of Training Residuals

NTK Eigenmodes and Fourier Mixture Components

Shreyas Kalvankar

December 2025

Purpose of This Analysis

This brief note presents two diagnostics that track how learning progresses along different function-space directions:

1. **NTK eigenmode decay:** projection of the training residual onto the top 5 eigenvectors of the empirical NTK at each training step.
2. **Fourier mixture component decay:** projection of the residual onto the seven components of the target Fourier mixture:

$$f^*(\gamma) = \sum_{j=1}^{7} a_j \sin(K_j\gamma + \phi_j), \qquad K_j \in \{2, 4, 7, 11, 16, 23, 32\}.$$

These plots reveal:

- which eigenmodes are reduced earliest during training;
- which Fourier harmonics of the target are learned early vs. late;
- how these behaviours depend on network width.

1 Residual Projections onto NTK Eigenmodes

For each snapshot step t, the empirical NTK on the training set is decomposed as $K_{tt}(t) = U(t)\Lambda(t)U(t)^\top$. We track the magnitude of the projection of the residual r_t onto the top five eigenvectors.

A dashed line indicates the NTK freeze time for each width.

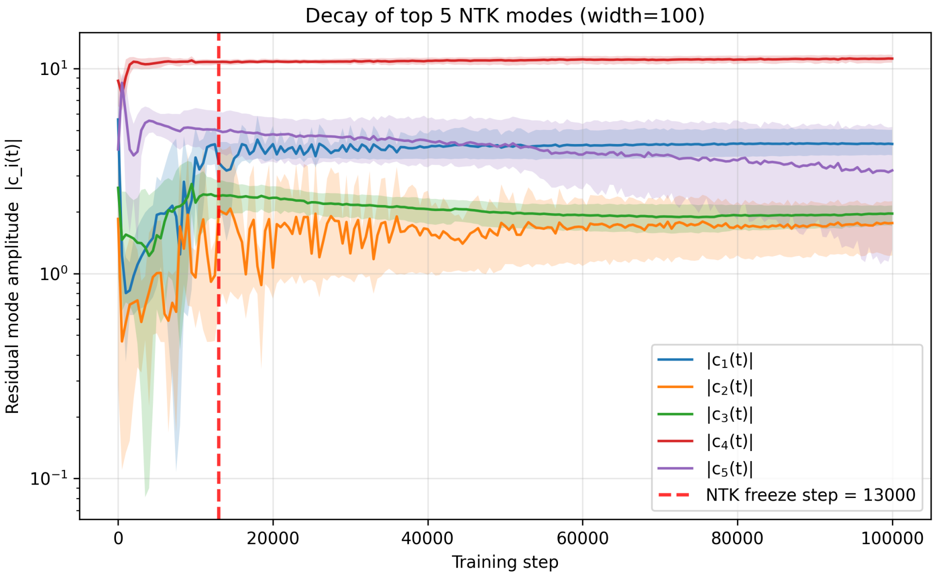

Figure 1: NTK eigenmode decay, width 100.

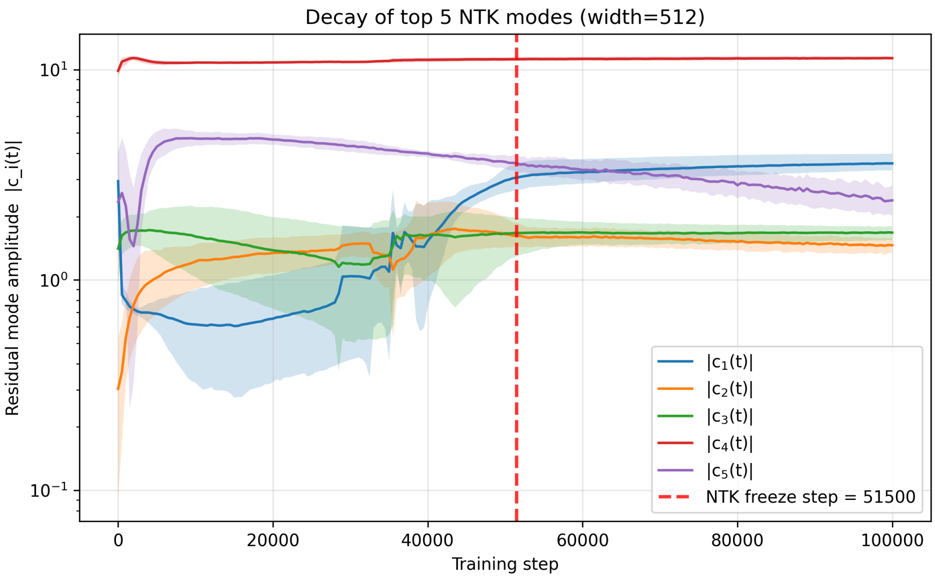

Figure 2: NTK eigenmode decay, width 512.

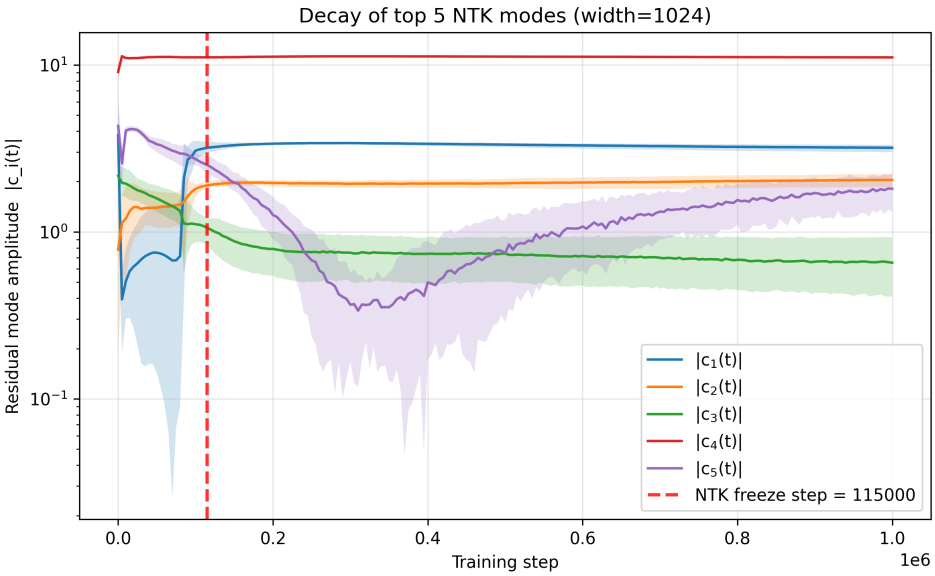

Figure 3: NTK eigenmode decay, width 1024.

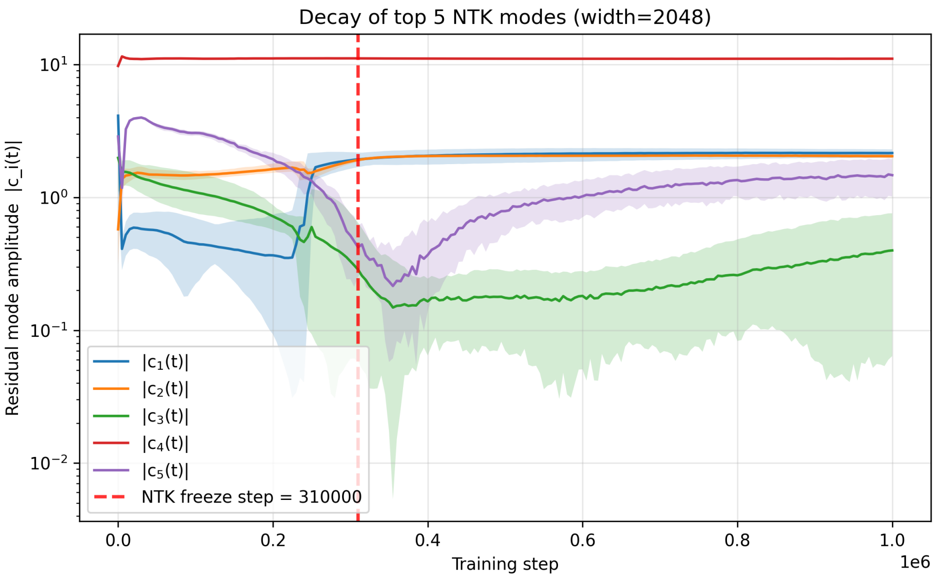

Figure 4: NTK eigenmode decay, width 2048.

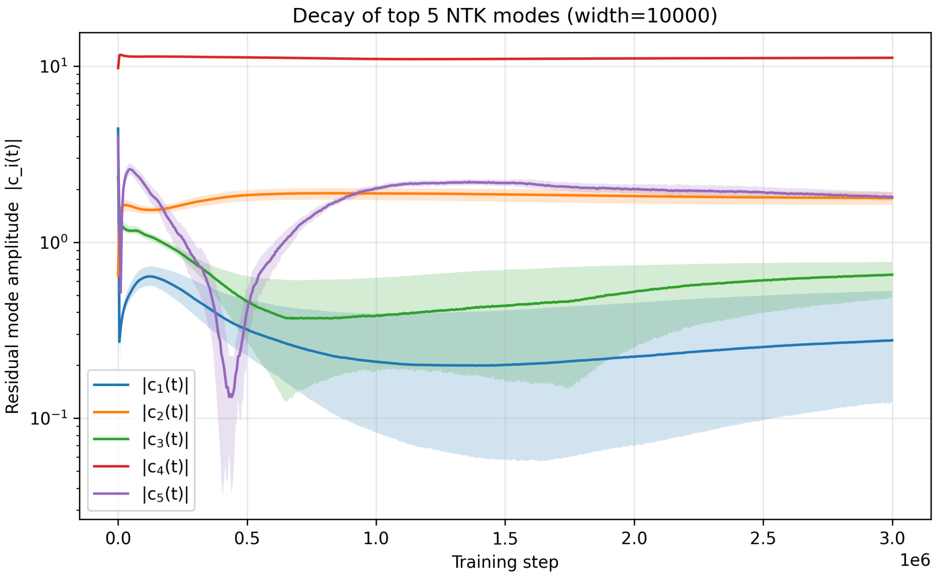

Figure 5: NTK eigenmode decay, width 10000.

2 Residual Projections onto Fourier Mixture Components

Let the seven basis vectors be:

$$b_j(\gamma) = \sin(K_j \gamma + \phi_j), \qquad K_j \in \{2, 4, 7, 11, 16, 23, 32\}.$$

For each step, we track the magnitude of the residual projected onto these components:

$$c_j(t) = \frac{\langle r_t, b_j \rangle}{\langle b_j, b_j \rangle}.$$

This shows how rapidly each harmonic of the target is learned.

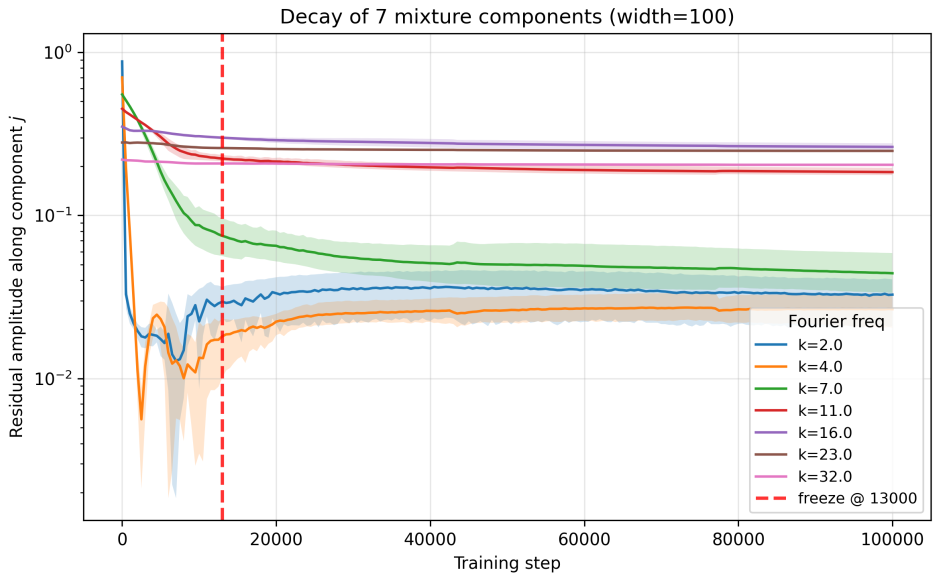

Figure 6: Fourier mixture component decay, width 100.

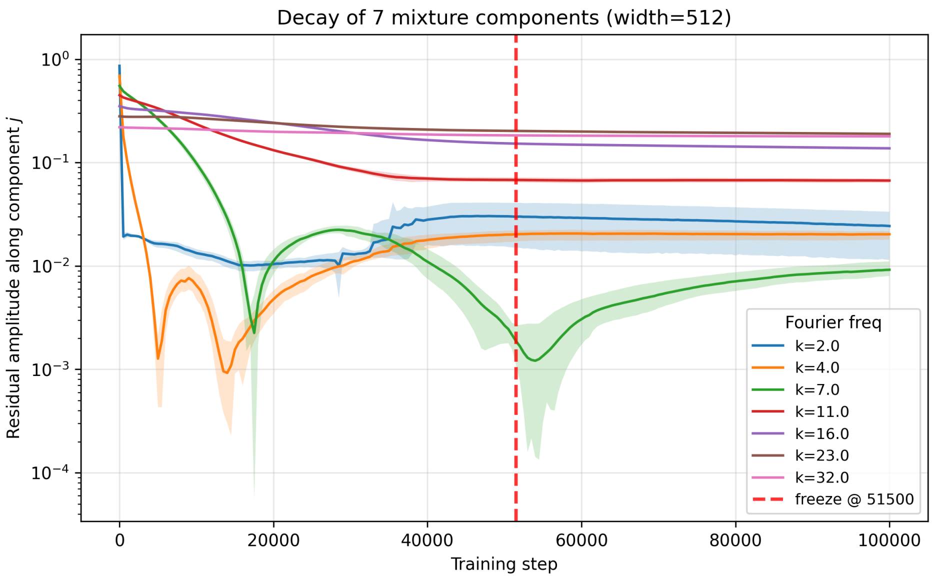

Figure 7: Fourier mixture component decay, width 512.

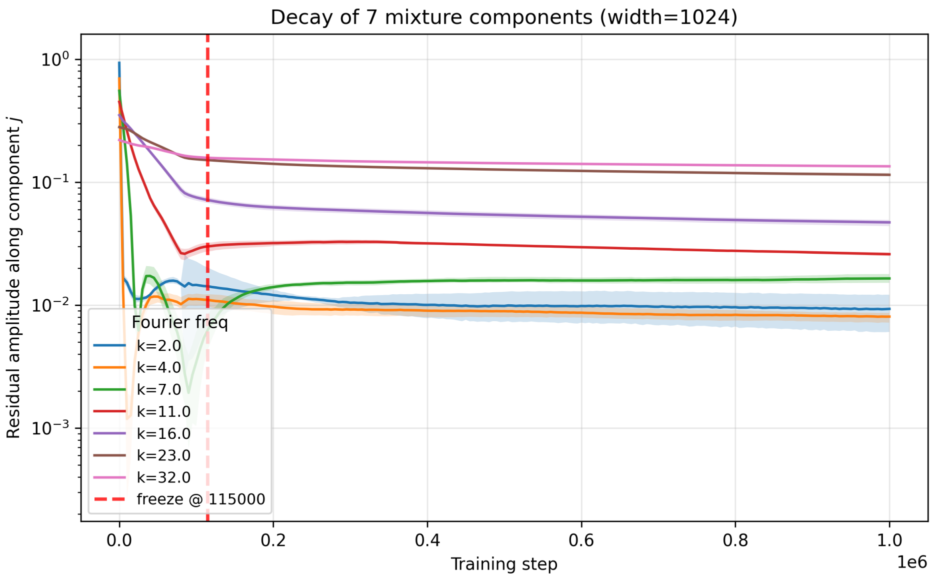

Figure 8: Fourier mixture component decay, width 1024.

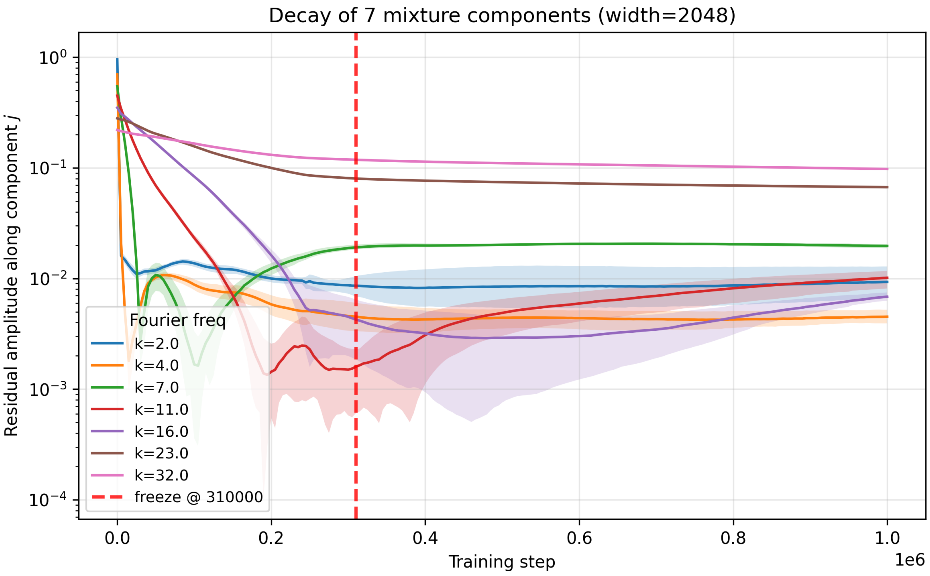

Figure 9: Fourier mixture component decay, width 2048.

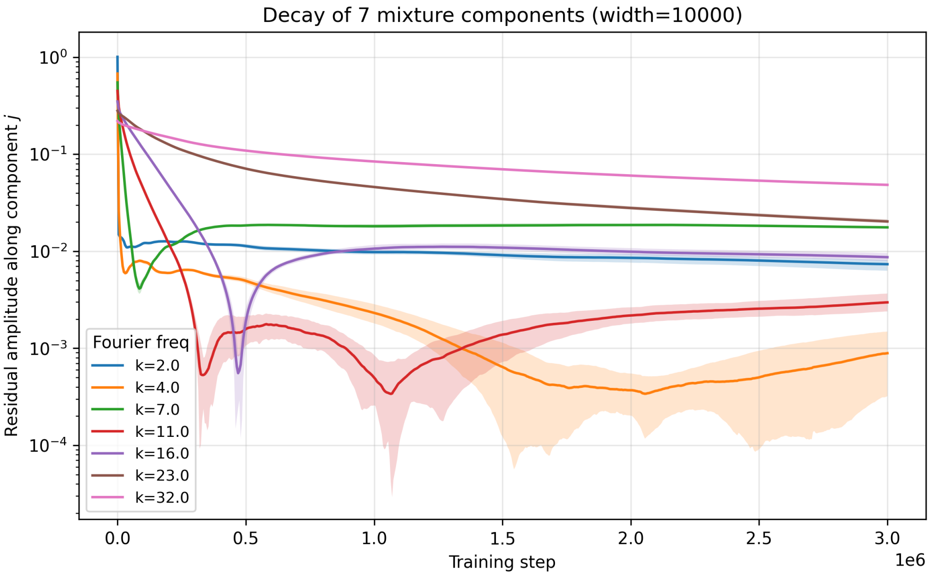

Figure 10: Fourier mixture component decay, width 10000.

3 Key Observations

- **Eigenmodes:** Across all widths, the top-5 mode amplitudes show non-monotonic behaviour (dips, rebounds, long drifts), and do not collapse after the NTK freeze point. This indicates that feature-learning–type dynamics continue even when the empirical NTK appears nearly frozen in Frobenius norm.
- Increasing width smooths the trajectories but does not eliminate this behaviour: wider networks show less noise and slower drift. This suggests that NTK freezing (as measured by drift) is not sufficient to imply lazy-training behaviour at finite width.
- **Fourier harmonics:** Low frequencies ($k = 2, 4$) decay early, while higher frequencies

($k \geq 16$) decay much more slowly, confirming a strong spectral bias.