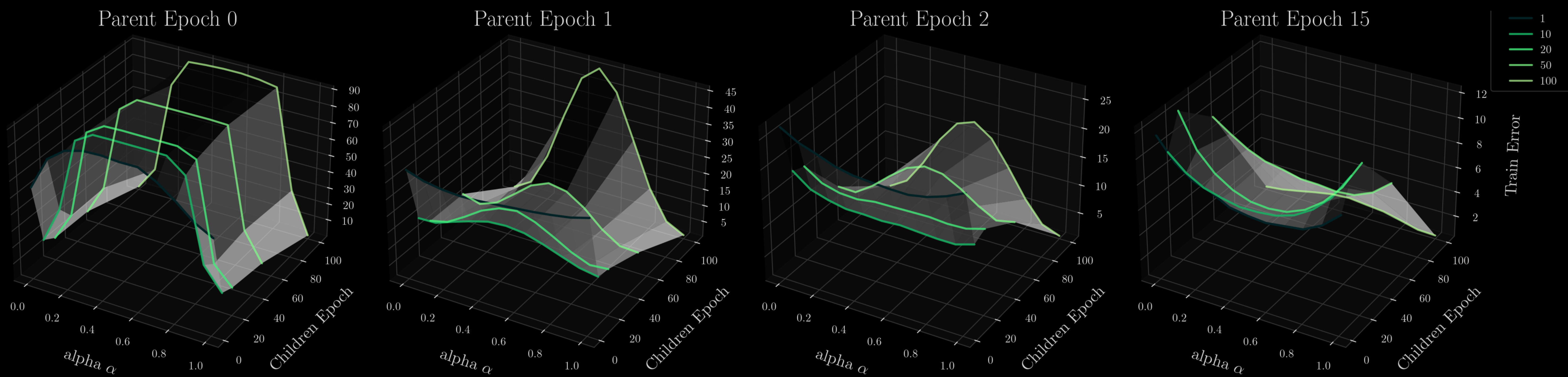


Landscaping Linear Mode Connectivity

Sidak Pal Singh, Linara Adilova, Michael Kamp,
Asja Fischer, Bernhard Schölkopf, Thomas Hofmann



How should the landscape be to show LMC (Frankle et al., 2019) ?



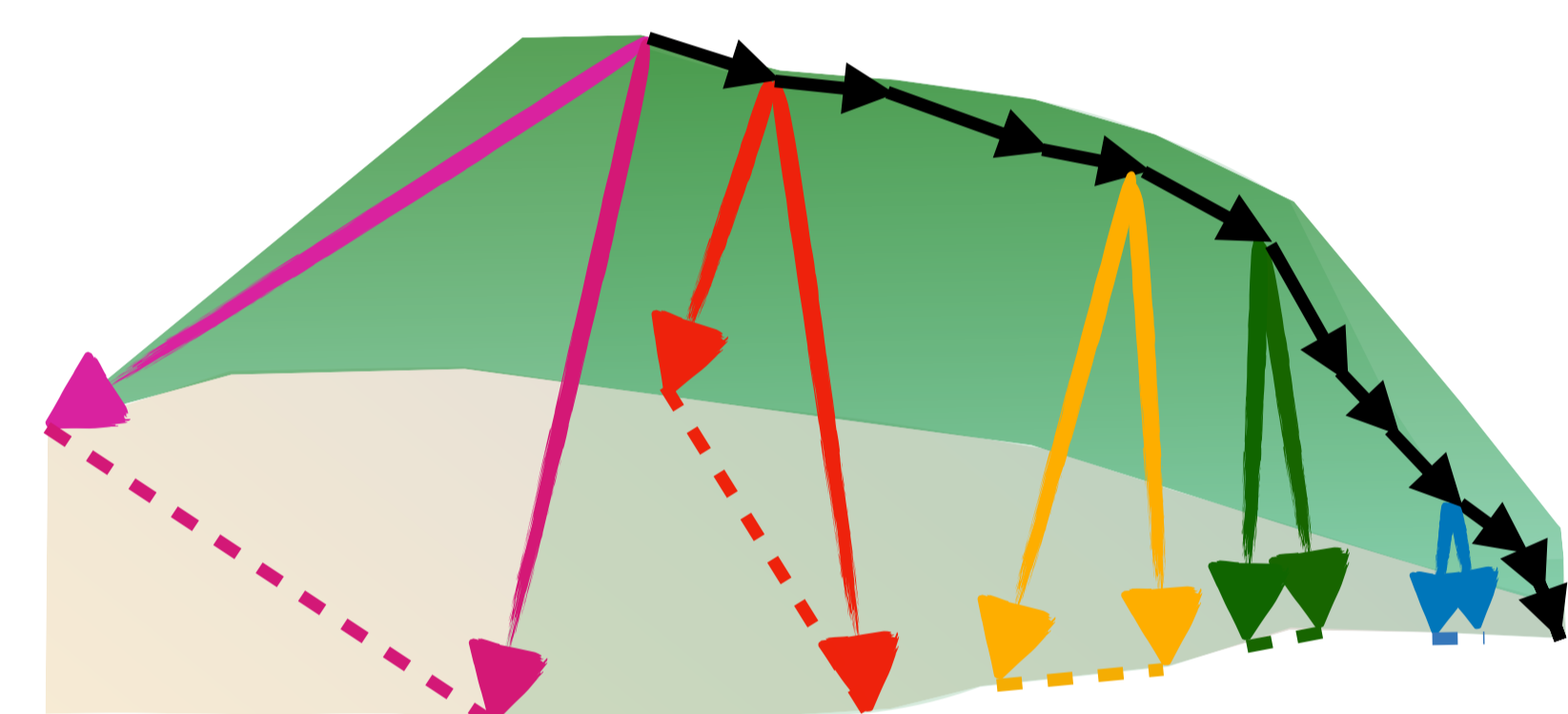
Cross-sectional Sibling Landscape: Evolution of training error curves when forked at different epochs (0,1,2,15)

Background

- Create two copies a network after some point in training and train them in an exactly similar manner, except for the random seed
- Frankle et, al., 2019 showed that: when forked in the early phase the child networks are separated by a barrier; but otherwise are connected by a low-loss linear path

A Metaphor for LMC

“Imagine going down a mountain top with a friend in foggy weather. If you lose them towards the end, they are gonna be a short walk at the base. But, if near the start, they might actually end up on the other side of the mountain”



Hypothesis: Mountain Top with Ridges

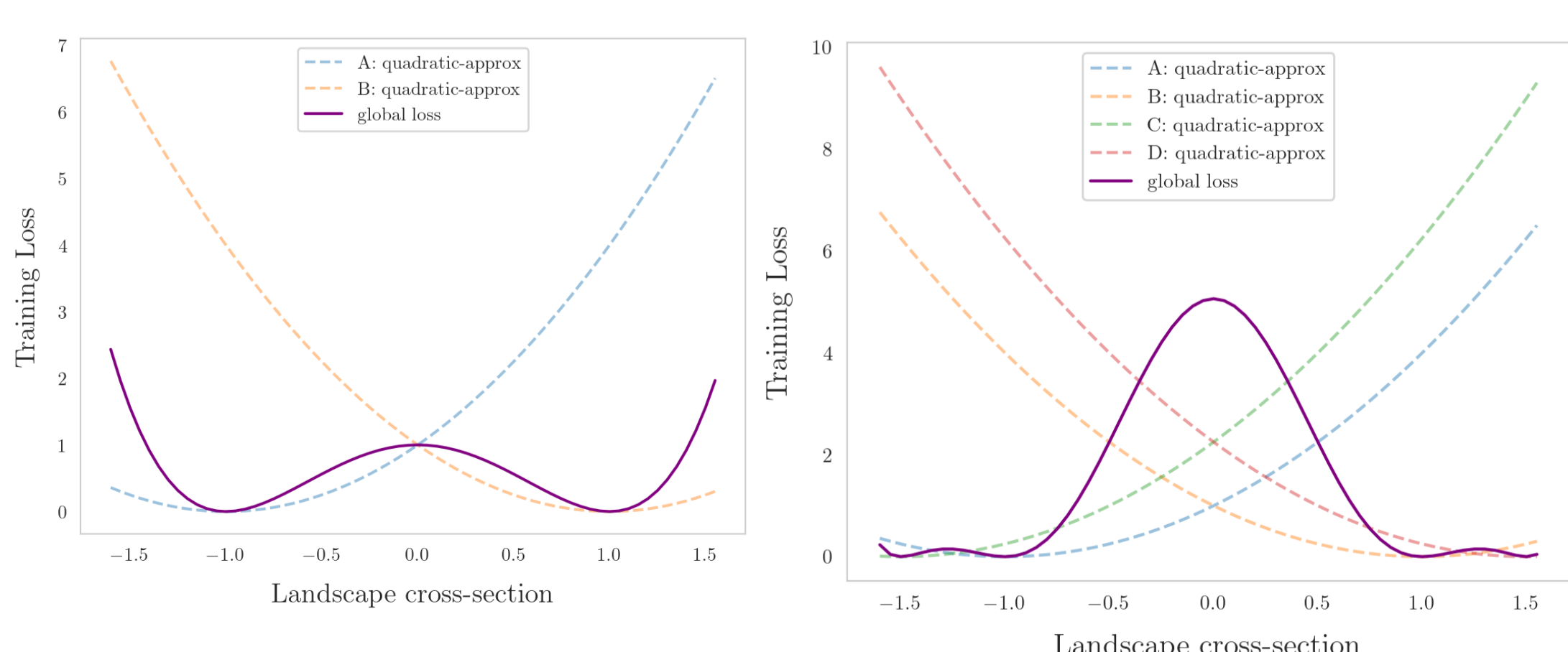
Measuring Loss Barrier

The loss barrier: $\mathcal{B}(\alpha) = \mathcal{L}((1-\alpha)\theta_1^* + \alpha\theta_2^*) - (1-\alpha)\mathcal{L}(\theta_1^*) - \alpha\mathcal{L}(\theta_2^*)$ when linearly interpolating θ_1^* and θ_2^* with weights $1-\alpha$ and α is

$$\mathcal{B}(\alpha) \approx \frac{\alpha(1-\alpha)}{2} (\Delta\theta^*)^\top \left(\alpha \nabla_{\theta}^2 \mathcal{L}(\theta_1^*) + (1-\alpha) \nabla_{\theta}^2 \mathcal{L}(\theta_2^*) \right) \Delta\theta^*$$

Qualitatively captures the empirically trend of LMC and layer wise LMC

Retrospective on how barriers arise

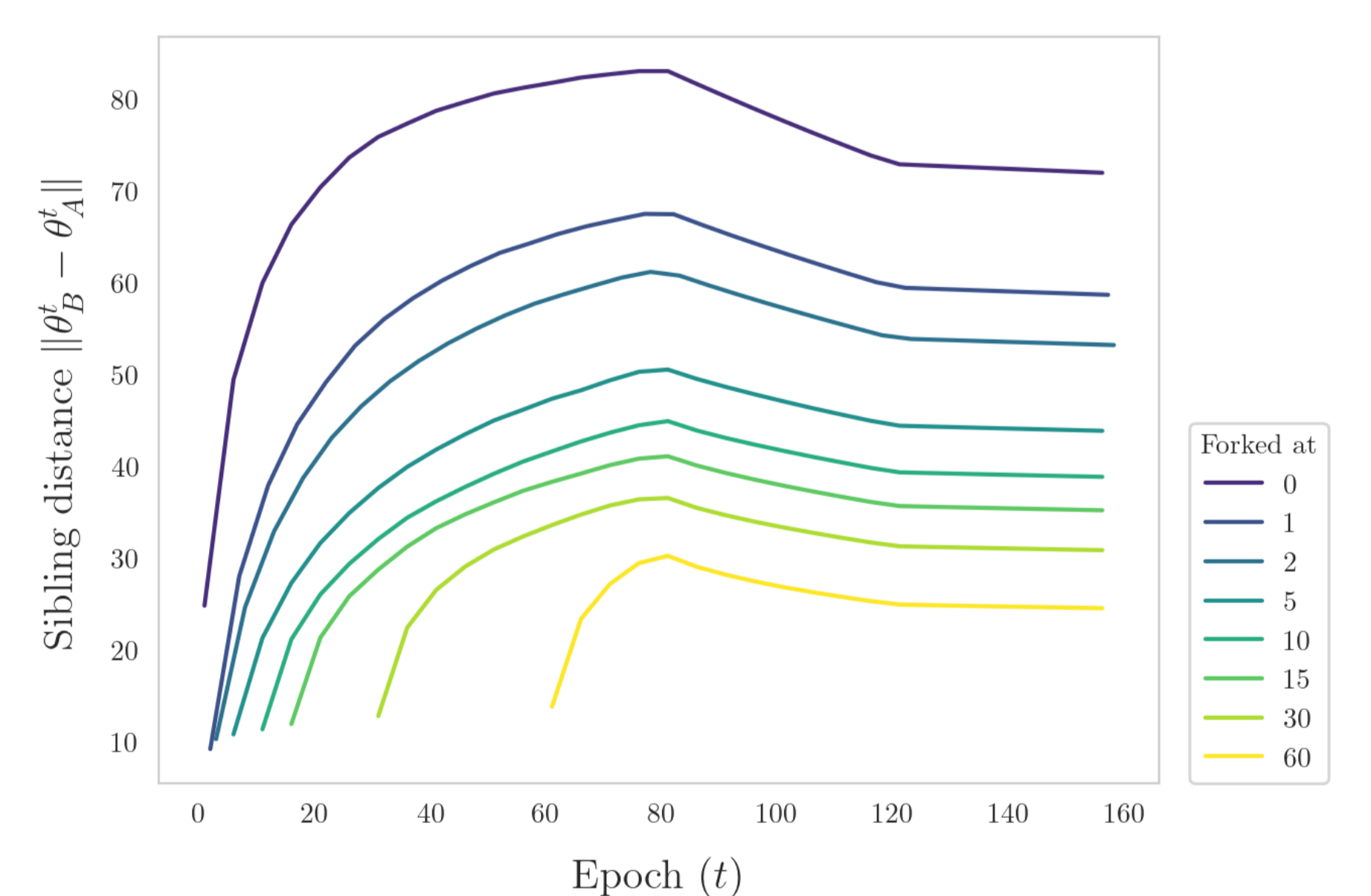


Product of local Landscapes $\prod_{i=1}^m (\theta - \theta_i^*)^2$

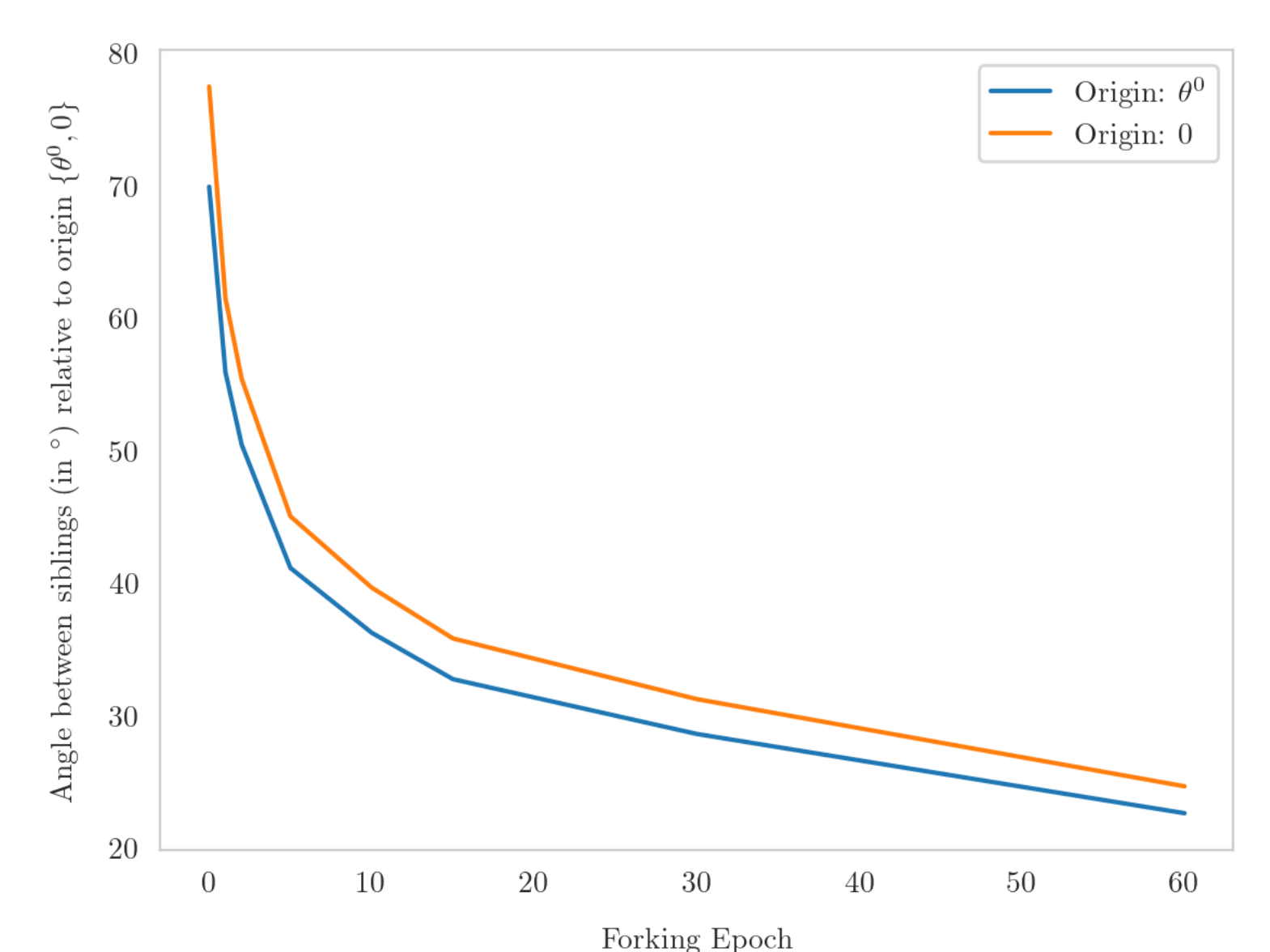
Insight:

Interplay of numerous nearby solutions and degenerate landscape geometry

Empirical Observations



Distance between model copies (siblings) during the course of training



Final angle traced between model copies