Revisiting the Goldilocks Zone in Inhomogeneous Networks

Zacharie Garnier Cuchet, Sarath Chandar, Ekaterina Lobacheva

Mila Université de Montréal · Polytechnique Montréal · utc Université de Technologie Compiègne

Goldilocks zone:

- Region in the weight space for **model initialization**.
- Location: narrow range of weight norms or softmax temperatures (equivalent for homogeneous models).
- Definition: loss exhibits an **excess of positive curvature**:

$$\frac{\mathrm{Tr}(H)}{||H||_F} = \frac{\sum_i \lambda_i}{\sqrt{\sum_i \lambda_i^2}} > 0$$

High if hessian eigenvalues are positive and similar to each other

- Importance: Believed to be connected to **good trainability**!

Inhomogeneous networks:

- Most of the practical networks are inhomogeneous!
- Definition:

$$f_{\alpha\theta}(x) \neq \alpha^L f_\theta(x)$$

- Reasons: components such as **biases**, **residuals**, and **normalization**.

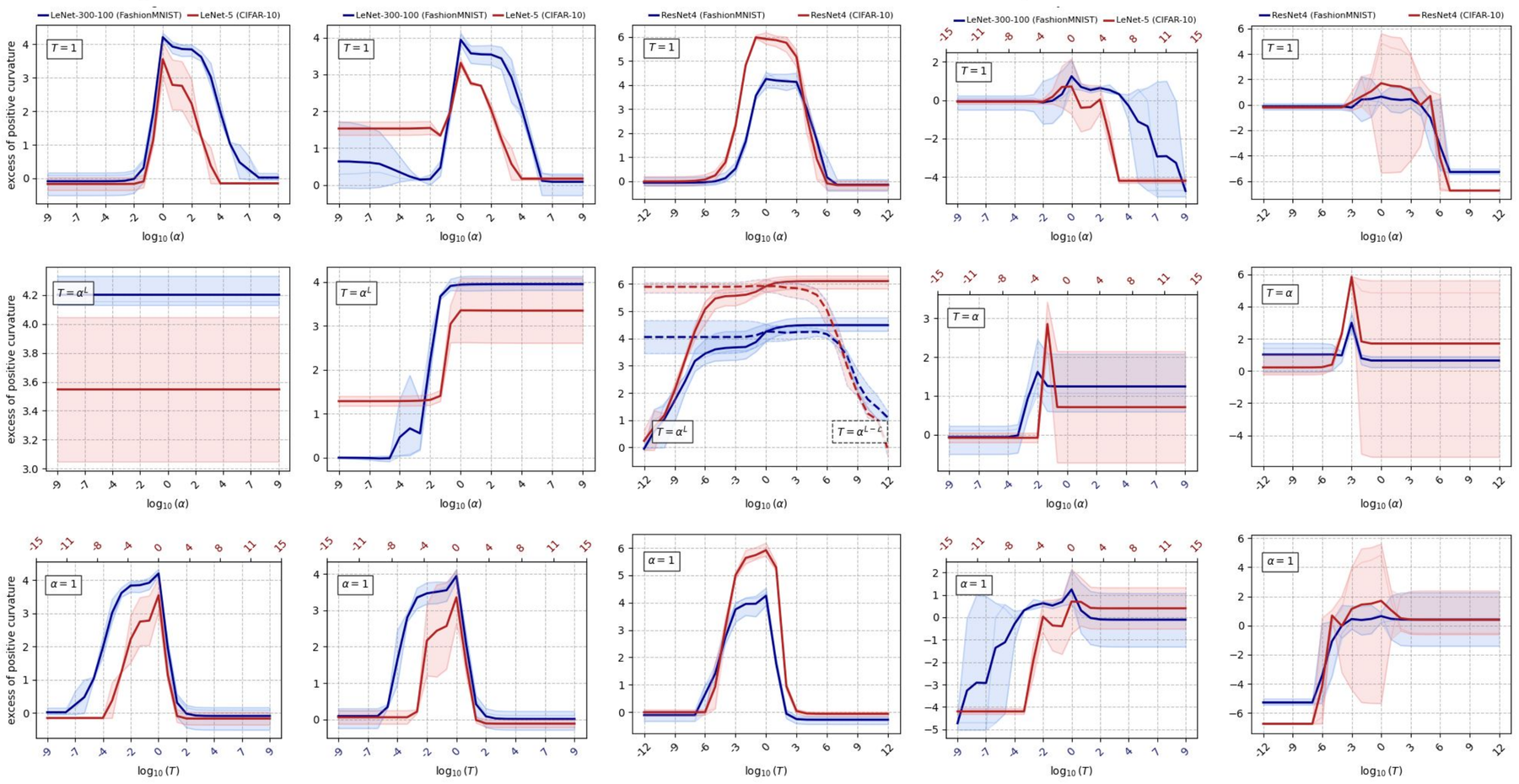
Goldilocks Zone in Inhomogeneous Networks

Localization

1. **Weight scaling to move in space**

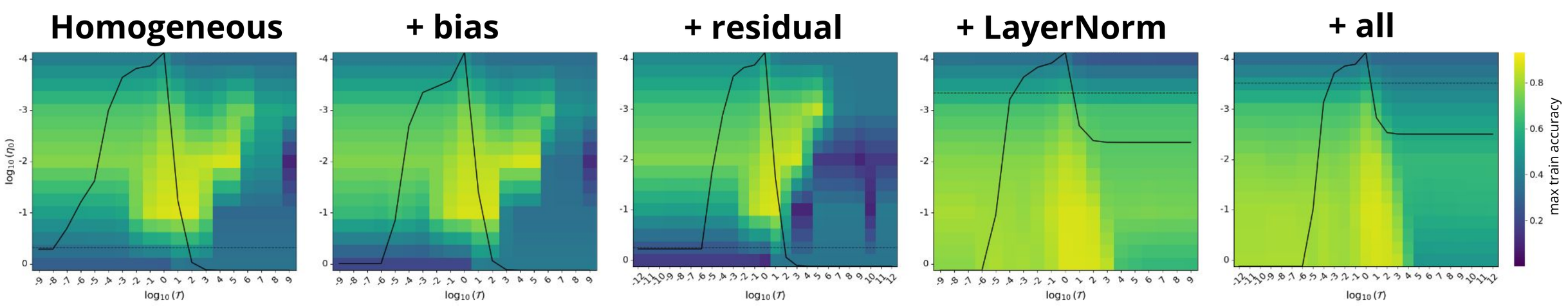
   Both scalings:
   - cancel out for homogeneous nets
   - different for inhomogeneous nets!

2. **Softmax temperature scaling to modulate confidence**



Goldilocks zone is present for all inhomogeneous networks, but LayerNorm leads to much lower and even negative curvature.

Goldilocks Zone vs Trainability



- **No strong alignment** between Goldilocks zone and trainability: both poor trainability inside the Goldilocks zone and good trainability outside of it are possible.
- LayerNorm improves trainability for a wide range of initializations despite weak or negative curvature.