What can Linear Interpolation of Neural Network Loss Landscapes Tell Us?

Tiffany Vlaar (University of Edinburgh) and Jonathan Frankle (MosaicML)

Get in touch at: Tiffany.Vlaar@ed.ac.uk and jonathan@mosaicml.com

ICML 2022

Linear Interpolation of Neural Network Loss Landscapes

$$\theta_\alpha = (1-\alpha)\theta_i + \alpha\theta_f \quad \text{for} \quad \alpha \in [0,1]$$

Goodfellow et al., 2015

Initial Final Model Parameters

Linear interpolation is seen as "a simple and lightweight method to probe neural network loss landscapes" (Lucas et al., 2021).

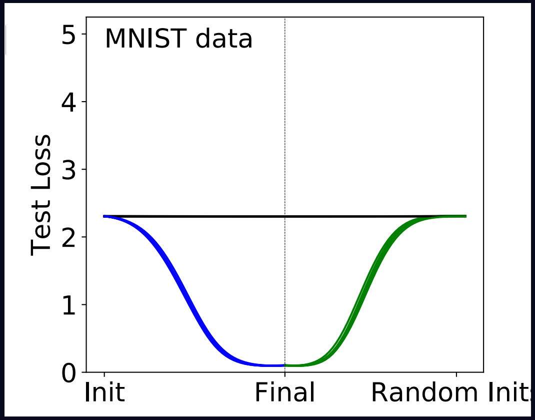
Linear Interpolation of Neural Network Loss Landscapes

$$\theta_\alpha = (1-\alpha)\theta_i + \alpha\theta_f \;\; \text{for} \;\; \alpha \in [0, 1]$$

Goodfellow et al., 2015

Initial Final Model Parameters

Linear interpolation is seen as “a simple and lightweight method to probe neural network loss landscapes” (Lucas et al., 2021).



Goodfellow et al., 2015

An absence of barriers along the linear path

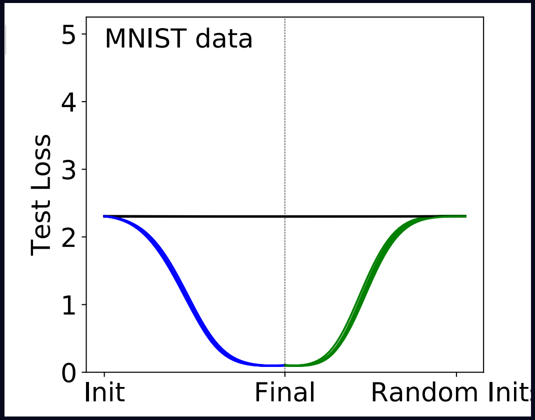
Linear Interpolation of Neural Network Loss Landscapes

$$\theta_\alpha = (1-\alpha)\theta_i + \alpha\theta_f \quad \text{for} \quad \alpha \in [0,1]$$

Goodfellow et al., 2015

Initial Final Model Parameters

Linear interpolation is seen as "a simple and lightweight method to probe neural network loss landscapes" (Lucas et al., 2021).



Goodfellow et al., 2015

An absence of barriers along the linear path

⇒ "tasks are relatively easy to optimize" (Goodfellow et al., 2015).

⇒"Though dimension is high, the space is in some sense simpler than we thought: [..] the walk could just as well have taken a straight line without encountering any obstacles" (Li et al., 2018).

Research Question

Does the shape of loss along the linear path relate to the "success" of optimization?

Research Question

Does the shape of loss along the linear path relate to the "success" of optimization?

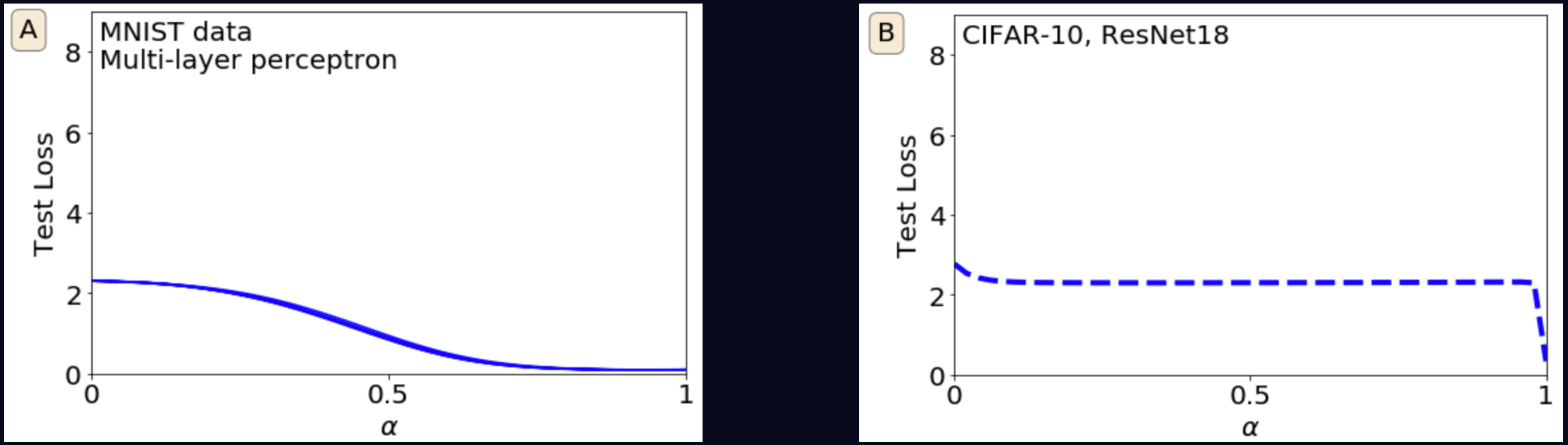
We study the influence of optimizer and architecture design choices:

- Role of Initialization
- Role of the data
- Role of the optimizer
- Role of the model

on the shape of the linear path AND the test accuracy of the final model

Base model: ResNet-18, CIFAR-10 data

Revisited: Linear Interpolation of Neural Network Loss Landscapes



In modern neural network architectures:
"Loss plateaus and error remains at the level of random chance …
… until near the optimum" (Frankle, 2020).

Similar observations by Lucas et al. (2021).

Layer-wise Linear Interpolation

Vary a single layer (or convolutional block) from initial to final state

$$\theta_\alpha^{(\ell)} = (1-\alpha)\theta_0^{(\ell)} + \alpha\theta_f^{(\ell)}, \quad \theta_\alpha^{(k)} = \theta_f^{(k)},\ k \neq \ell.$$

Chatterji et al., 2020

Keep all other parameters fixed at their final state.

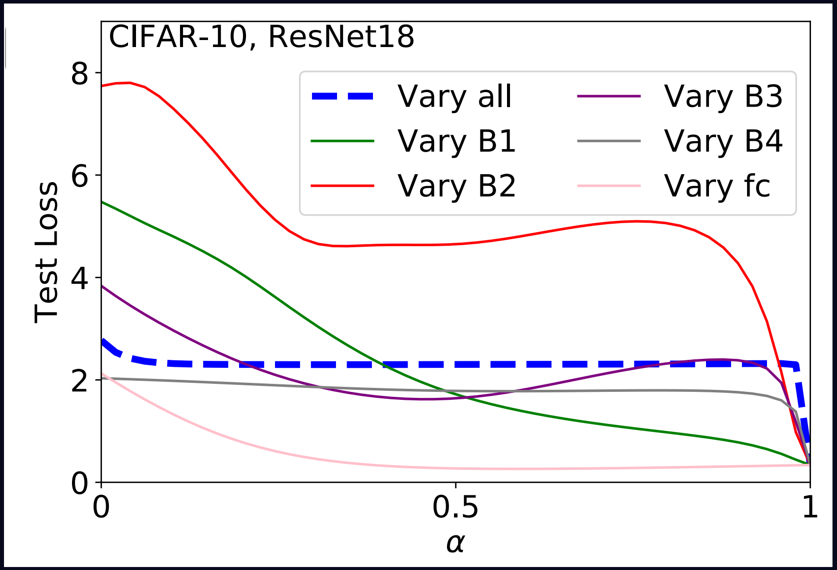
Layer-wise Linear Interpolation

Vary a single layer (or convolutional block) from initial to final state

$$\theta_\alpha^{(\ell)} = (1-\alpha)\theta_0^{(\ell)} + \alpha\theta_f^{(\ell)}, \quad \theta_\alpha^{(k)} = \theta_f^{(k)}, \; k \neq \ell.$$

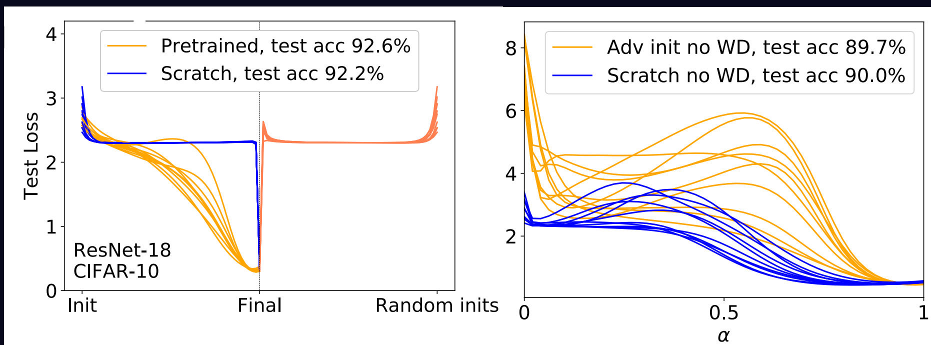
Chatterji et al., 2020

Keep all other parameters fixed at their final state.

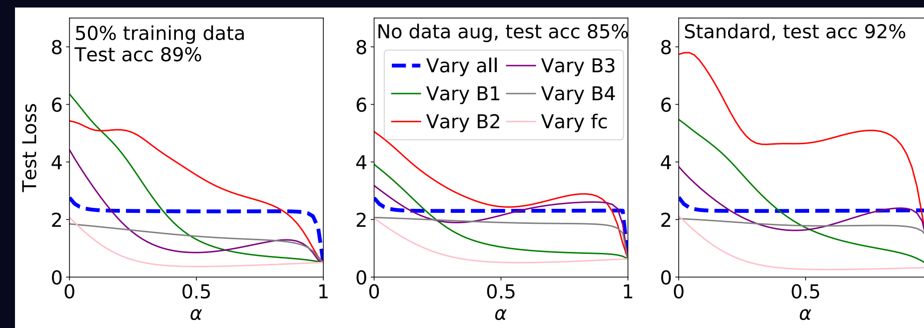


Interventions
Effect on: 1) Shape of Linear Path 2) Final Test Accuracy

Role of Initialization

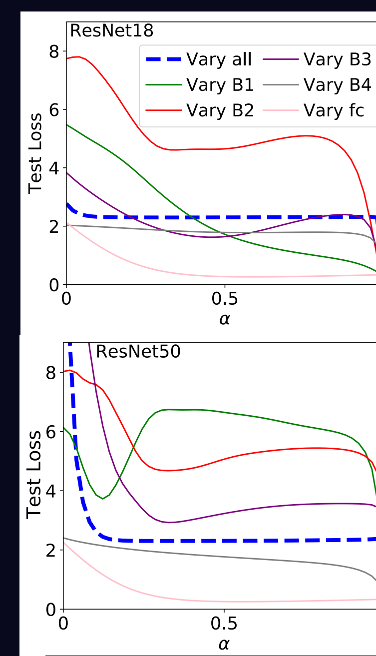


Role of Data

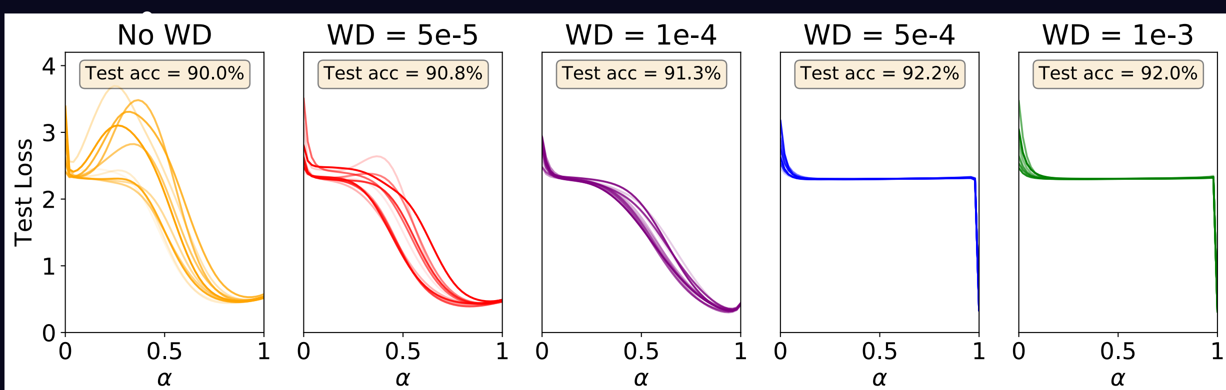


Role of Model

•



Role of Optimization



Findings

Question I: Does the shape of loss along the linear path relate to the "success" of optimization?

⇒ No!

An absence of barriers along the linear path.. "tasks are relatively easy to optimize" (Goodfellow et al., 2015).

Findings

Question I: Does the shape of loss along the linear path relate to the "success" of optimization?

$\Rightarrow$ No!

An absence of barriers along the linear path.. "tasks are relatively easy to optimize" (Goodfellow et al., 2015).

Question II: Does the shape of loss along the linear path relate to other aspects of optimization?

"pre-trained weights guide the optimization to a flat basin of the loss landscape." (Neyshabur et al., 2020)

"Large distances moved in weight space encourage non-monotonic interpolation" (Lucas et al., 2021).

Findings

Question I: Does the shape of loss along the linear path relate to the "success" of optimization?

⇒ No!

An absence of barriers along the linear path.. "tasks are relatively easy to optimize" (Goodfellow et al., 2015).

Question II: Does the shape of loss along the linear path relate to other aspects of optimization?

⇒ Pre-training on ImageNet consistently removes the presence of barriers for ResNet architectures, whereas adversarial initialization on random labels increases barriers.

"pre-trained weights guide the optimization to a flat basin of the loss landscape." (Neyshabur et al., 2020)

⇒ Distance between initial and final parameter state is **not** a reliable indicator of non-monotonic behaviour along linear path.

"Large distances moved in weight space encourage non-monotonic interpolation" (Lucas et al., 2021).

Layer-wise Findings

The adversarial effect of partial pre-training

Set model to trained (T) state
→ Re-set specific layer to random initialization (RI)
→ Re-train
→ Worse test accuracy

Layer-wise Findings

The adversarial effect of partial pre-training

Set model to trained (T) state
→ Re-set specific layer to random initialization (RI)
→ Re-train
→ Worse test accuracy

ResNet-18, CIFAR-10

Method	Test acc (%)
T-All but RI-1	91.79 ±0.23
T-All but RI-2	91.83 ±0.21
T-All but RI-3	92.35 ±0.20
T-All but RI-4	90.97 ±0.31
Train from scratch	92.2%

What can Linear Interpolation of Neural Network Loss Landscapes Tell Us?

Tiffany Vlaar (University of Edinburgh) and Jonathan Frankle (MosaicML)

Get in touch at: Tiffany.Vlaar@ed.ac.uk and jonathan@mosaicml.com