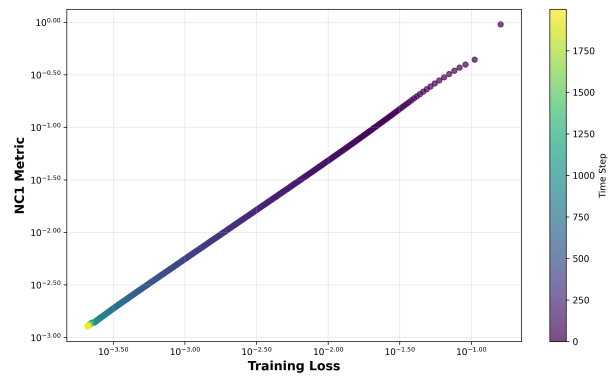
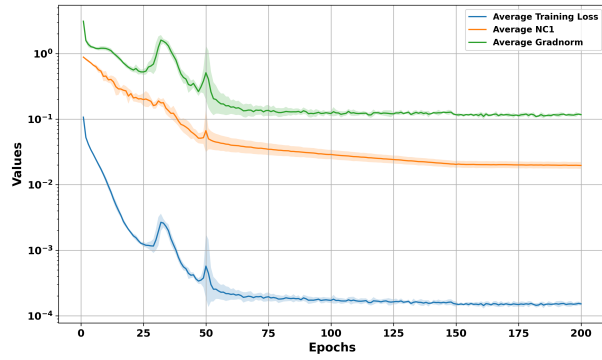


(a) Grad norm v.s. NC1

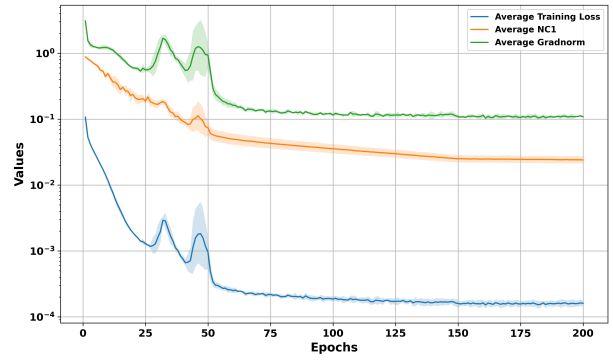


(b) Training loss v.s. NC

Figure 3: Scatterplot of average gradient norm/training loss and average NC1 for training a three-layers network on MNIST. The color bar indicates the training time steps.

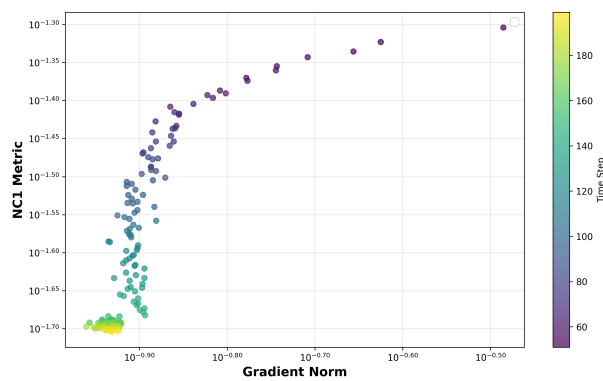


(a) ResNet-18

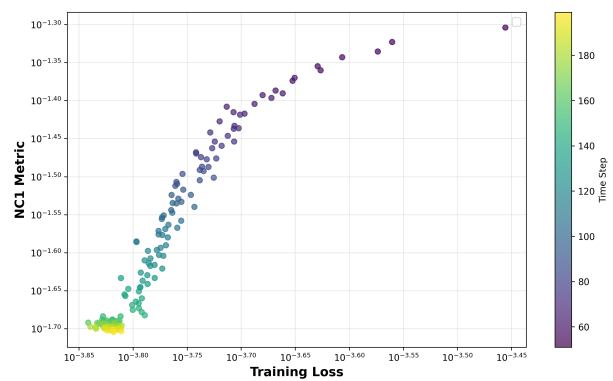


(b) VGG-18

Figure 4: Average training loss (blue), NC1 (orange) and gradient norm (green) during SGD training for ResNet-18 and VGG-11 on CIFAR-10. We report the average for 4 independent experiments, as well as the confidence interval at 1 standard deviation.

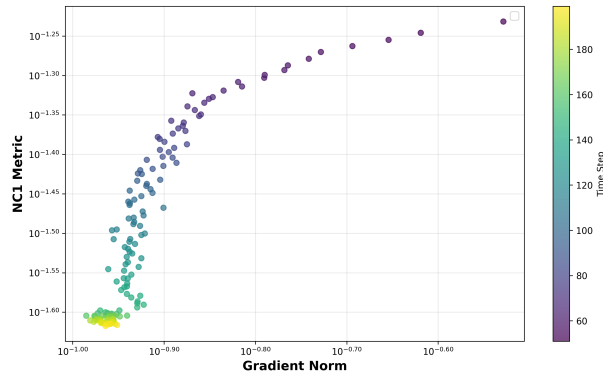


(a) Grad norm v.s. NC1

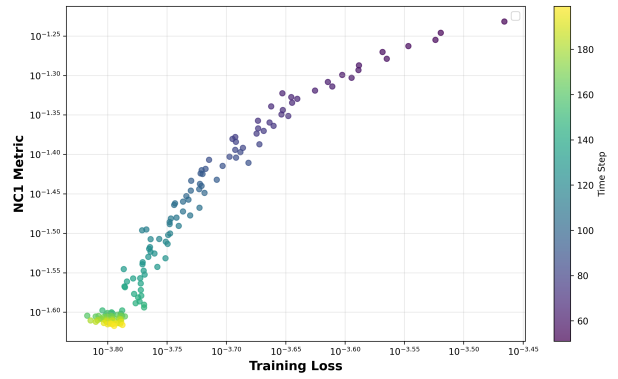


(b) Training loss v.s. NC

Figure 5: Scatterplot of gradient norm/training loss and NC1 for training ResNet-18 on CIFAR-10. The plot start from the 51 epoch right after the gradient is scheduled with factor 0.1. The color bar indicate the training time steps.



(a) Grad norm v.s. NC1



(b) Training loss v.s. NC

Figure 6: Scatterplot of average gradient norm/training loss and average NC1 for training VGG-11 on CIFAR-10. The plot start from the 51 epoch right after the gradient is scheduled with factor 0.1. The color bar indicate the training time steps.