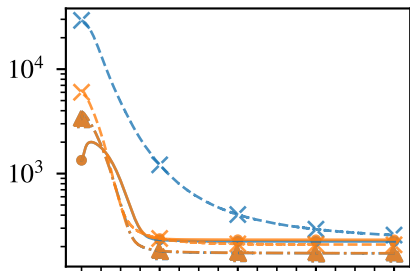
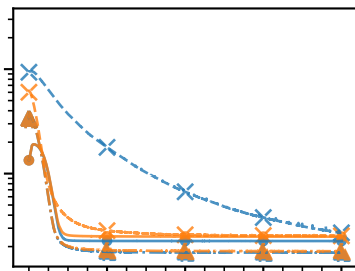


Weight Decay During Training across Architectures

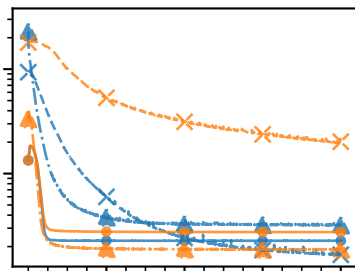
$n = 64$



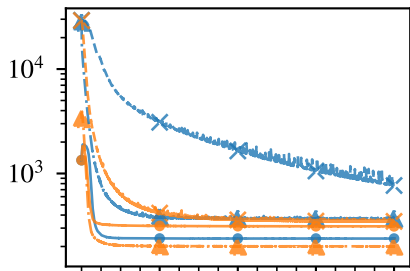
$n = 128$



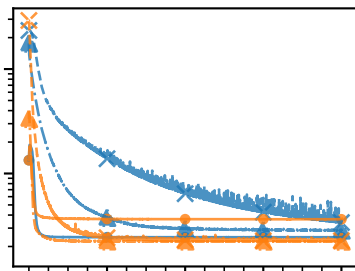
$n = 256$



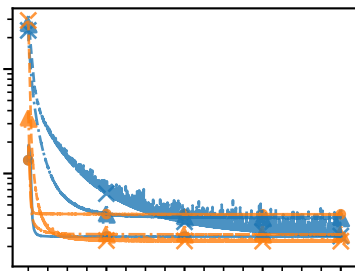
$n = 512$



$n = 1024$



$n = 2048$



15k 30k 45k 60k

Epochs

15k 30k 45k 60k

Epochs

15k 30k 45k 60k

Epochs

