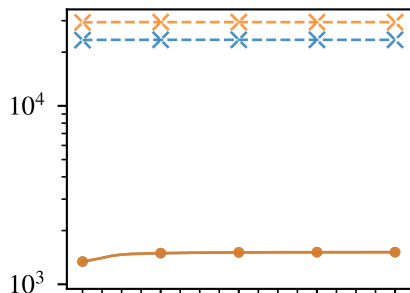
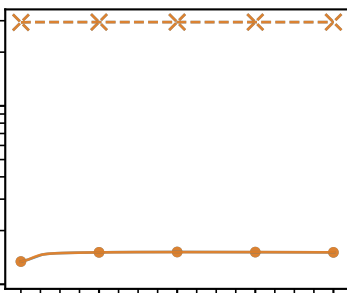


Weight Decay During SGD Training

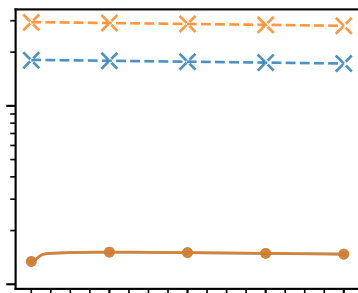
$n = 64$



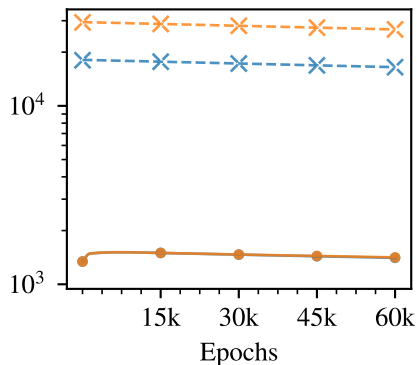
$n = 128$



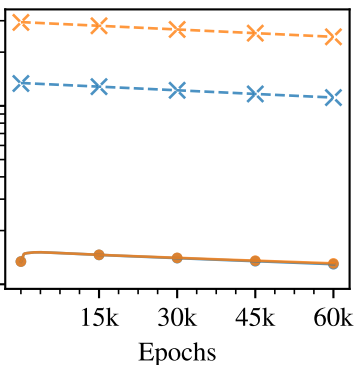
$n = 256$



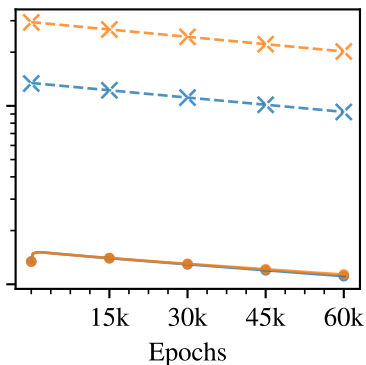
$n = 512$



$n = 1024$



$n = 2048$



$\sigma=0$

$\sigma=0.25$

without extra linear layers

with extra linear layers