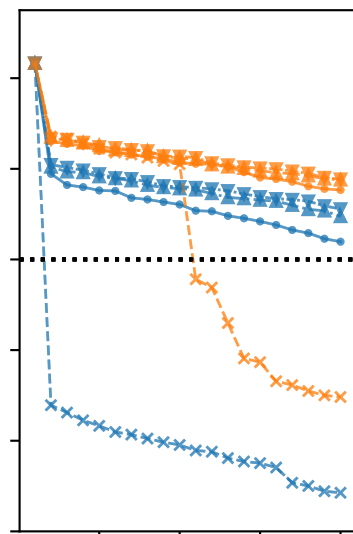
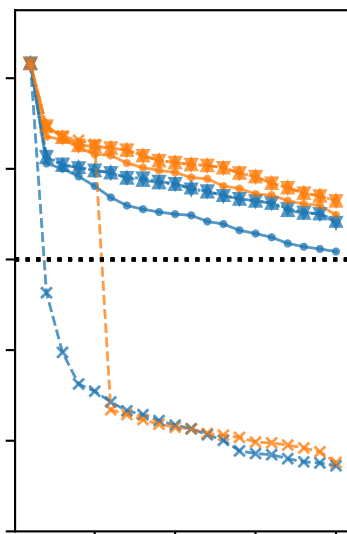
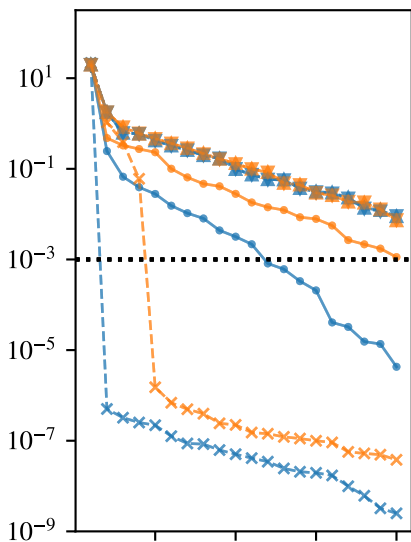
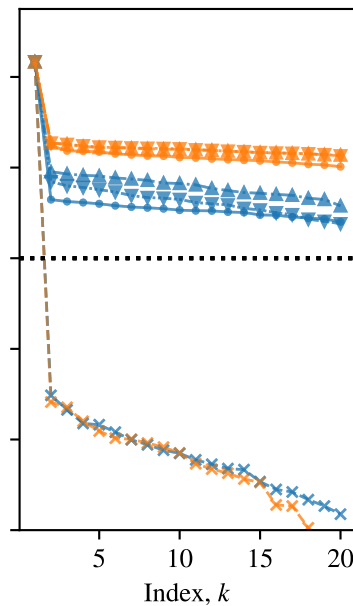
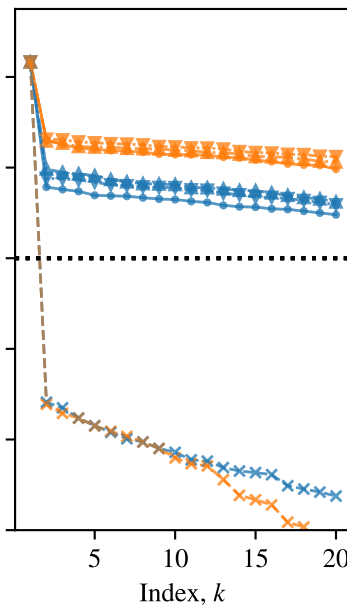
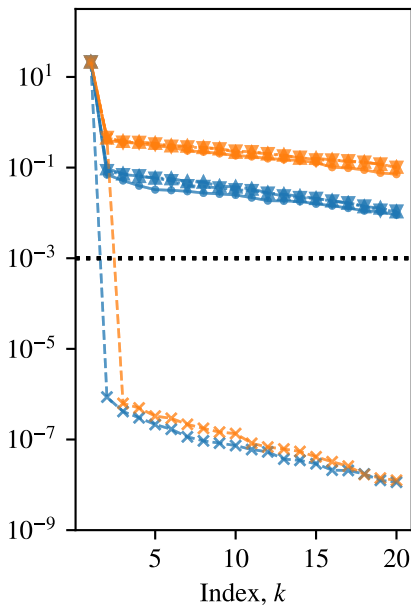


Singular Values of Trained Networks, $\sigma_k(\hat{f})$ $n = 64$ $n = 128$ $n = 256$  $n = 512$ $n = 1024$ $n = 2048$  $\sigma = 0$ $\sigma = 0.25$

Shallow

Linear Layers then ReLU

Deep ReLU

ReLU then Linear Layers then ReLU

effective rank tolerance, $\varepsilon = 10^{-3}$