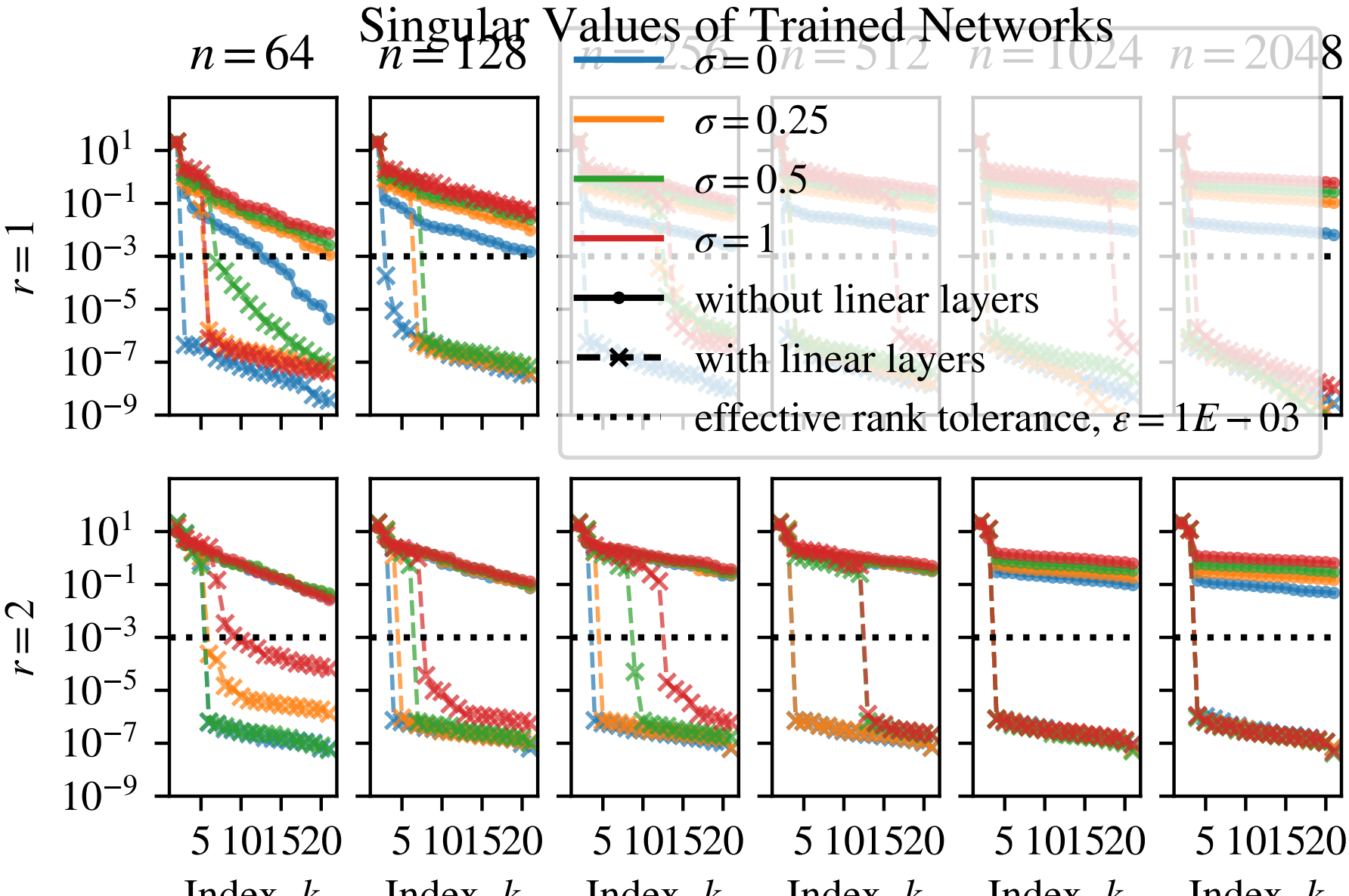
Singular Values of Trained Networks
$n = 64$
$n = 128$
$n = 256$
$n = 512$
$n = 1024$
$n = 2048$
$r = 1$
$r = 2$
10^{1}
10^{-1}
10^{-3}
10^{-5}
10^{-7}
10^{-9}
5 10 15 20
Index, k
$\sigma = 0$
$\sigma = 0.25$
$\sigma = 0.5$
$\sigma = 1$
without linear layers
with linear layers
effective rank tolerance, $\varepsilon = 1E-03$