dim=1024hidden dim=64data dim=512sample size=64training step=128r parallel initialization 8.0 gaussian initialization orthogonal initialization 0.6 SSO 0.4 0.2 0.0 20 40 60 80 100 120

step