dim=1024hidden dim=32data dim=256sample size=32training step=128r parallel initialization 0.7 gaussian initialization orthogonal initialization 0.6 0.5 0.4 0.3 0.2 0.1

60

step

80

100

120

0.0

0

20

40