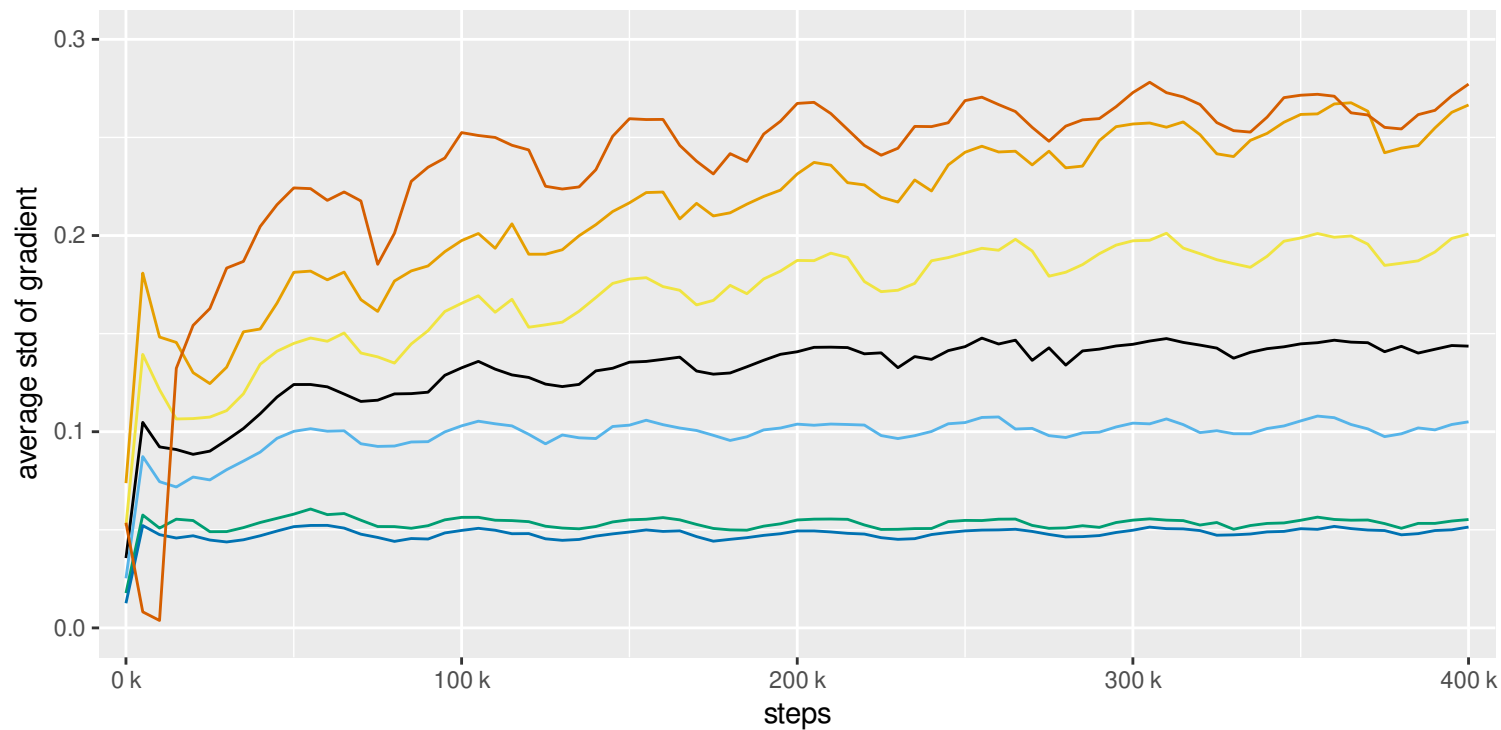


Average standard deviation of gradient of each parameter (n=10)



batch size 87.5 sec 150 sec 5 min 10 min 20 min 40 min 80 min