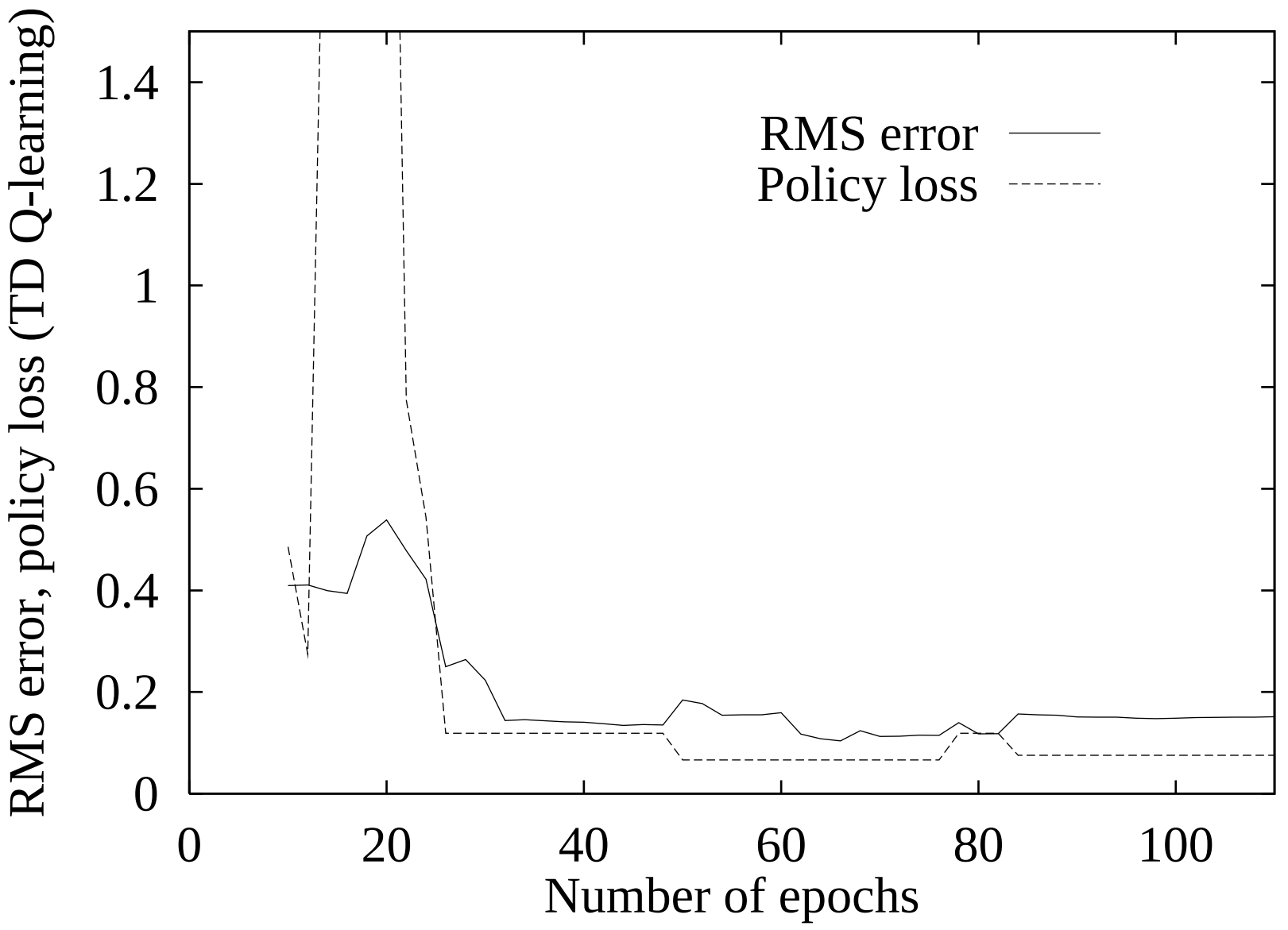
RMS error, policy loss (TD Q-learning)
0
0.2
0.4
0.6
0.8
1
1.2
1.4
0
20
40
60
80
100
Number of epochs
RMS error
Policy loss