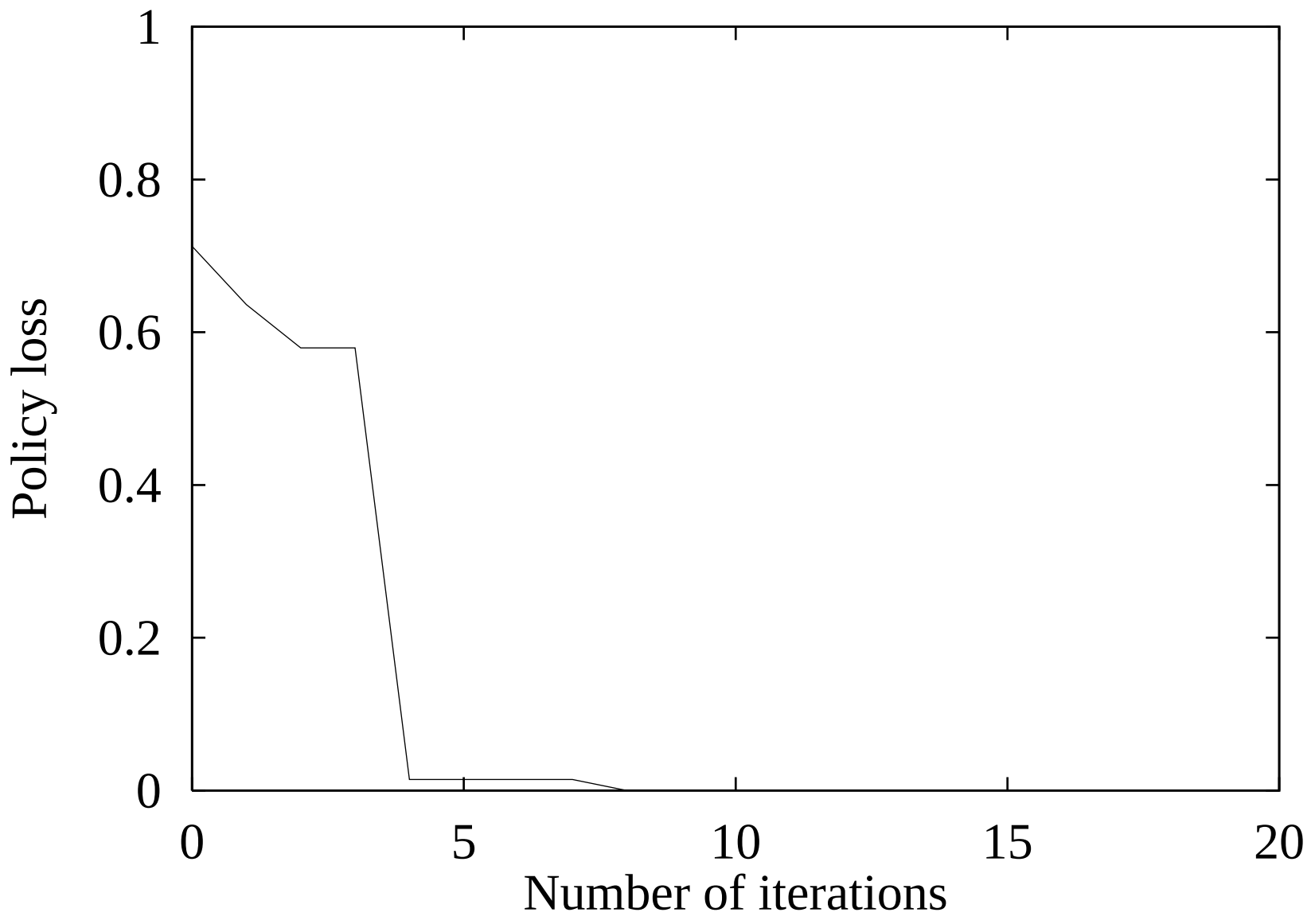
1
0.8
0.6
0.4
0.2
0
Policy loss
0
5
10
15
20
Number of iterations