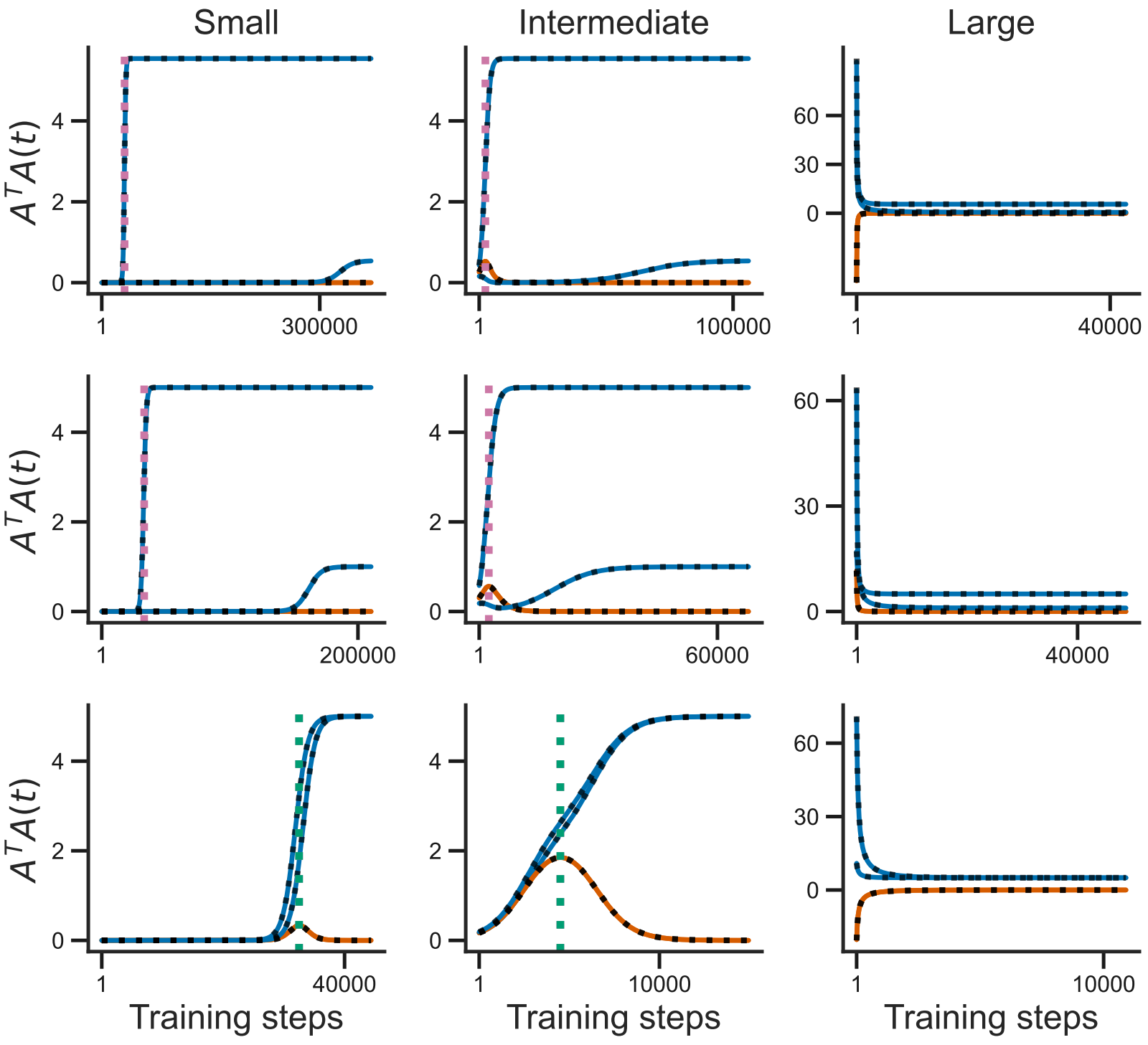
Small
Intermediate
Large
$A^TA(t)$
0
2
4
30
60
1
300000
100000
40000
200000
60000
10000
Training steps