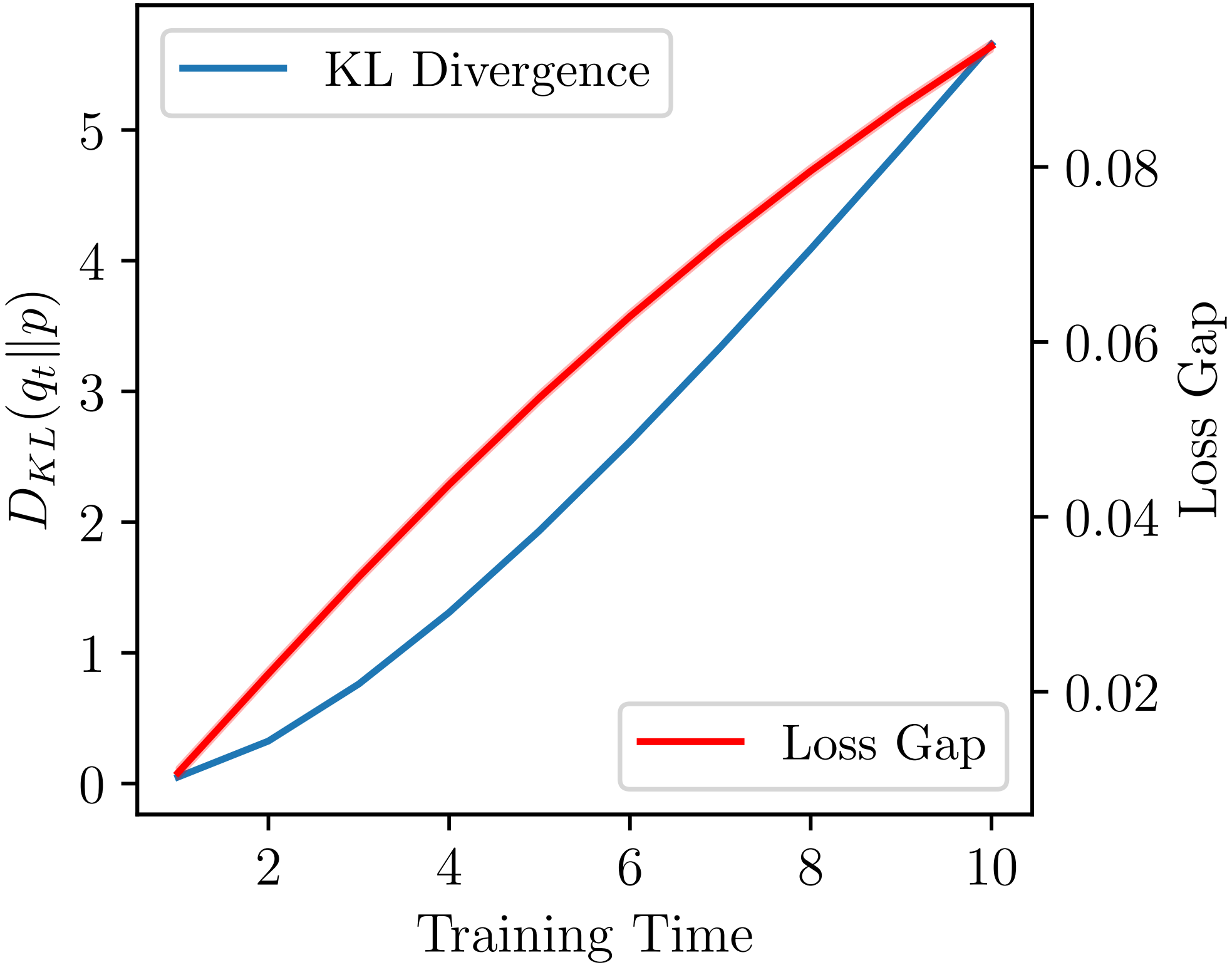
KL Divergence
Loss Gap
$D_{KL}(q_t||p)$
Loss Gap
Training Time
0
1
2
3
4
5
0.02
0.04
0.06
0.08
2
4
6
8
10