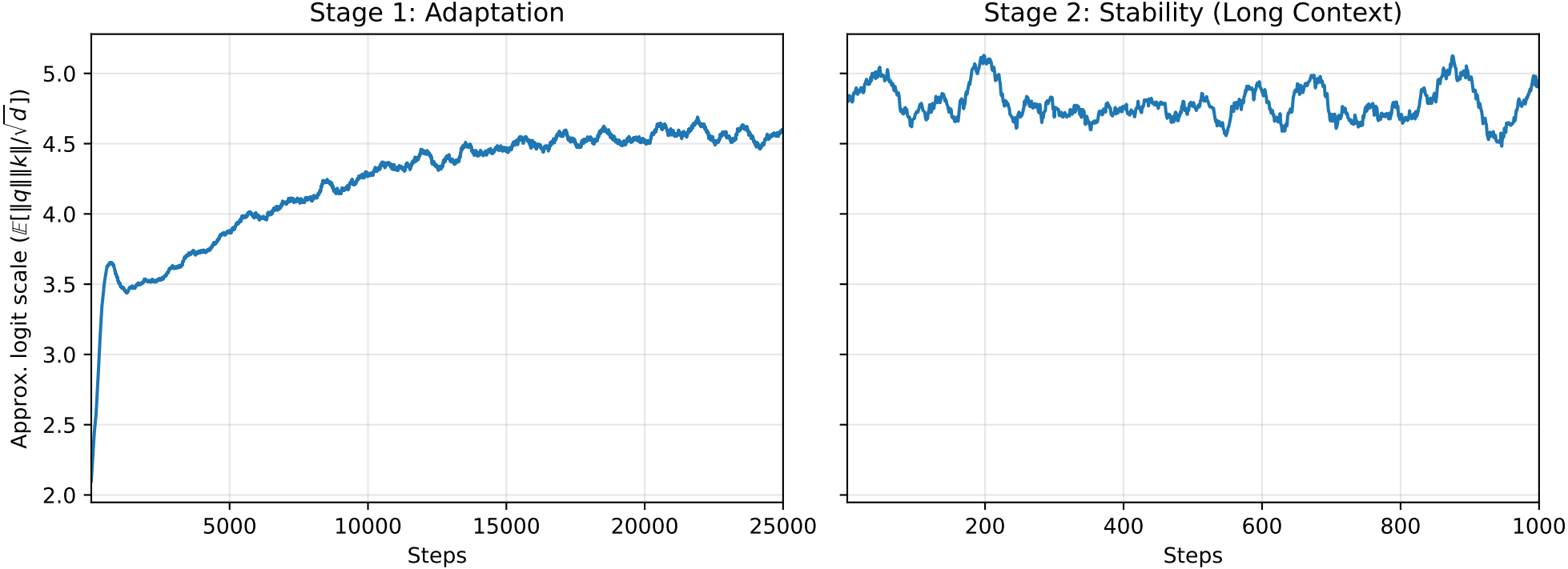
Stage 1: Adaptation
Stage 2: Stability (Long Context)
Approx. logit scale ($\mathbb{E}[\|q\|\|k\|/\sqrt{d}]$)
2.0
2.5
3.0
3.5
4.0
4.5
5.0
5000
10000
15000
20000
25000
200
400
600
800
1000
Steps
Steps