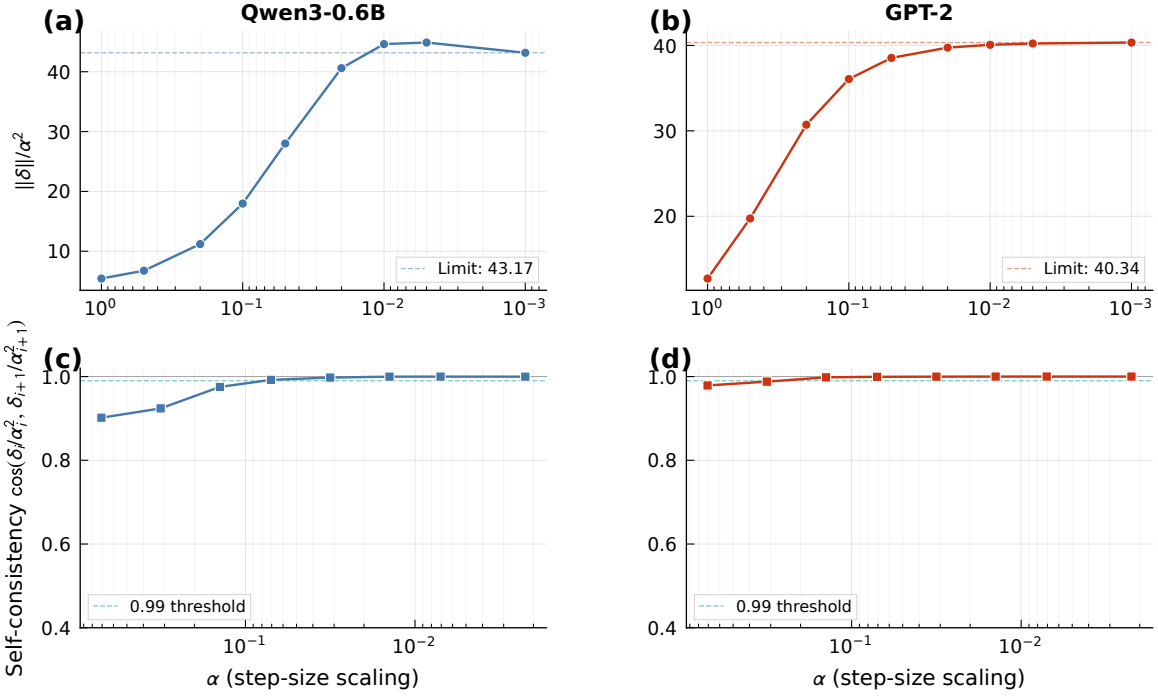
(a)
Qwen3-0.6B
(b)
GPT-2
$\|\delta\|/\alpha^2$
40
30
20
10
Limit: 43.17
Limit: 40.34
10^0
10^{-1}
10^{-2}
10^{-3}
(c)
(d)
Self-consistency $\cos(\delta_i/\alpha_i^2,\ \delta_{i+1}/\alpha_{i+1}^2)$
1.0
0.8
0.6
0.4
0.99 threshold
α (step-size scaling)