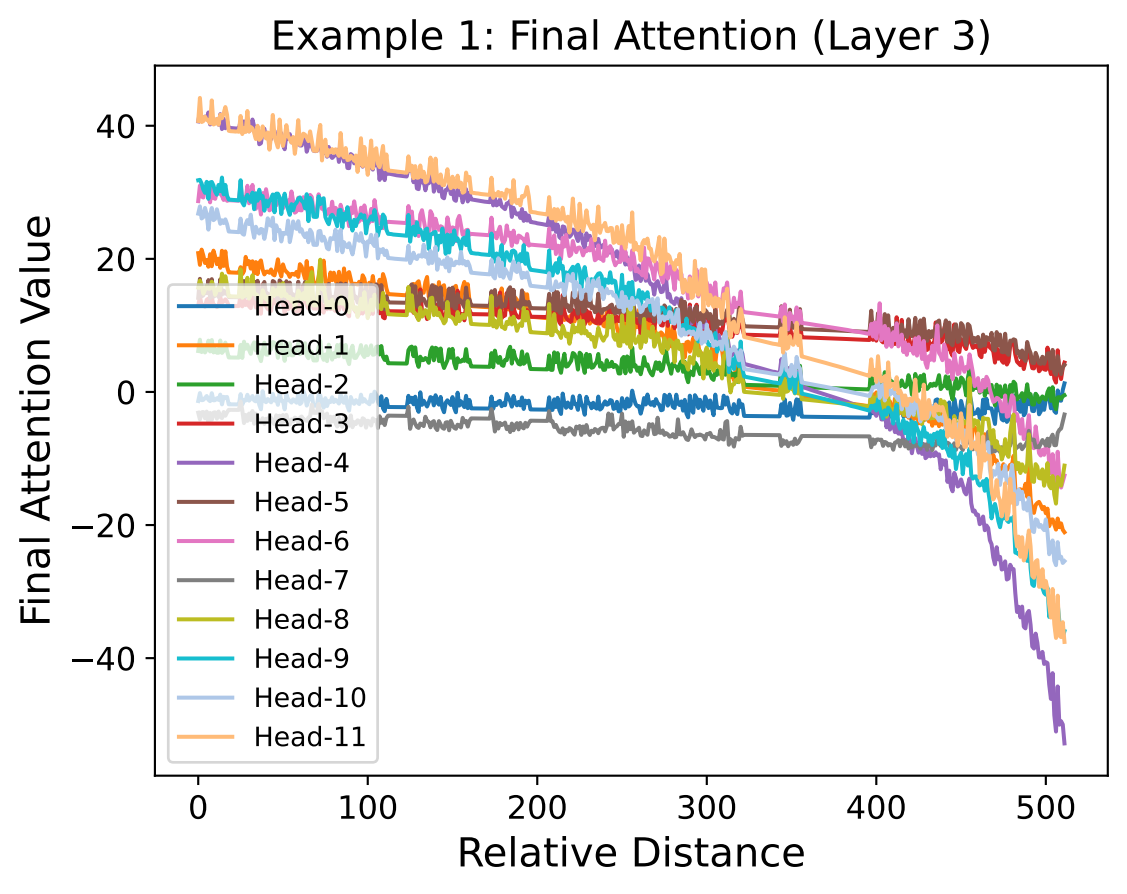
Example 1: Final Attention (Layer 3)
Final Attention Value
Relative Distance
40
20
0
−20
−40
0
100
200
300
400
500
Head-0
Head-1
Head-2
Head-3
Head-4
Head-5
Head-6
Head-7
Head-8
Head-9
Head-10
Head-11