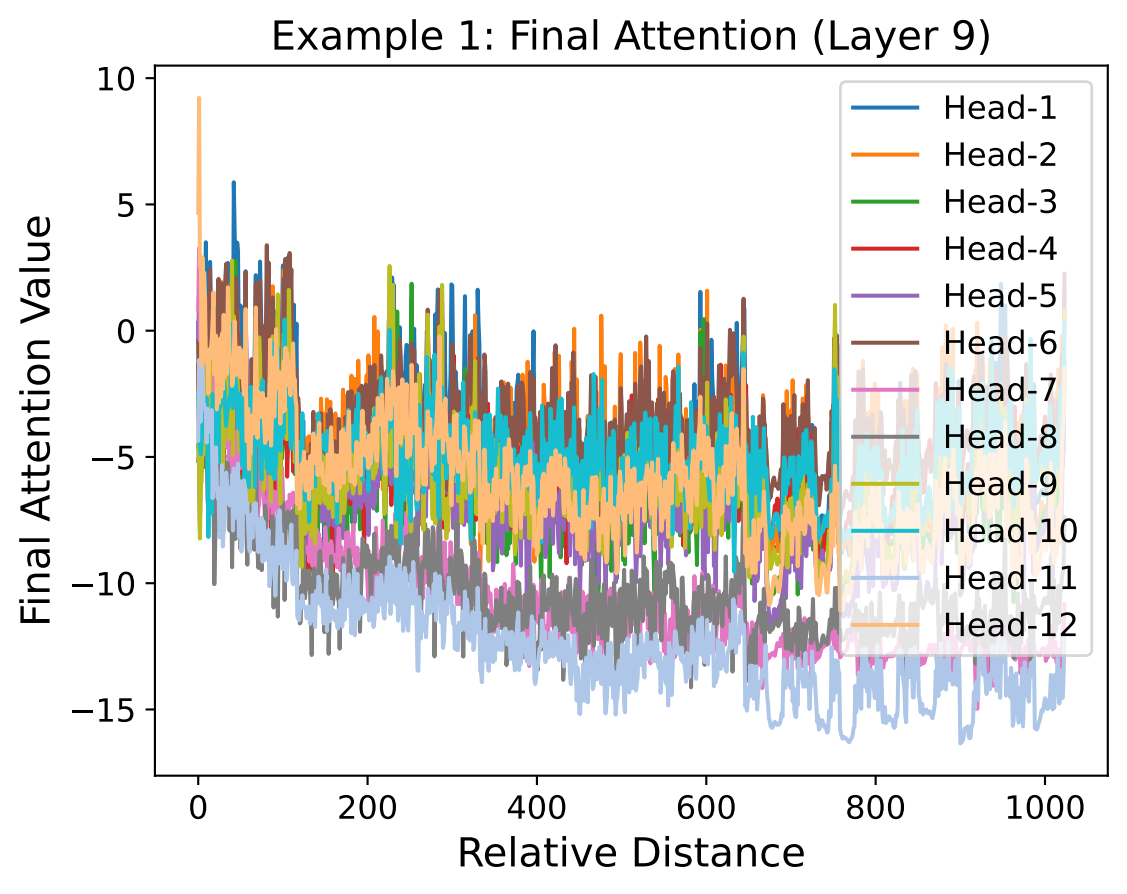
Example 1: Final Attention (Layer 9)
Final Attention Value
Relative Distance
10
5
0
−5
−10
−15
0
200
400
600
800
1000
Head-1
Head-2
Head-3
Head-4
Head-5
Head-6
Head-7
Head-8
Head-9
Head-10
Head-11
Head-12