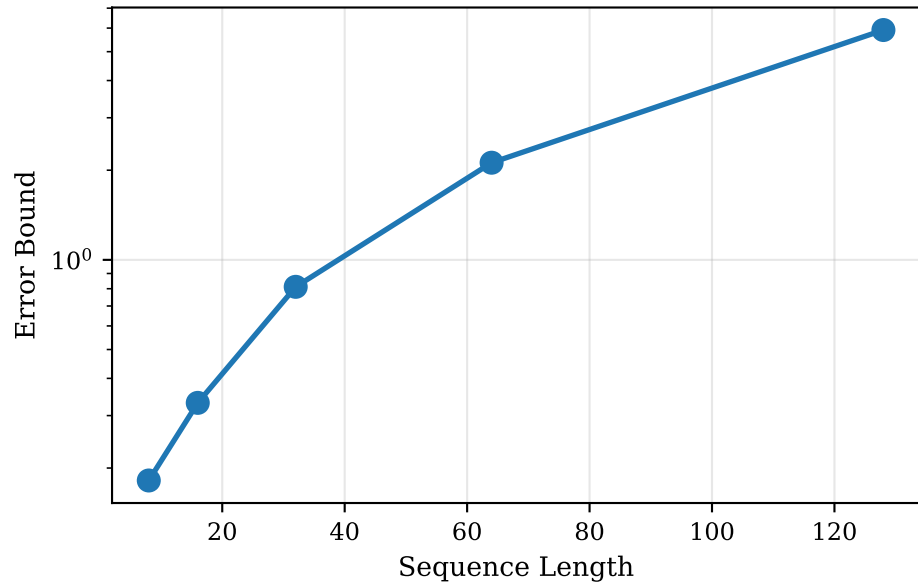


(a) Attention Error vs Sequence Length



(b) Attention Component Lipschitz Constants

