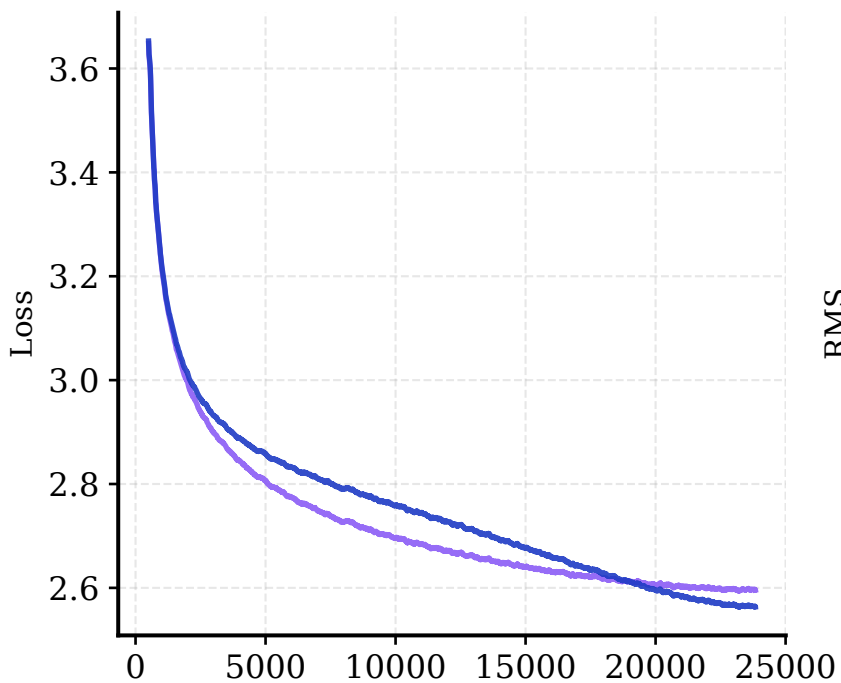
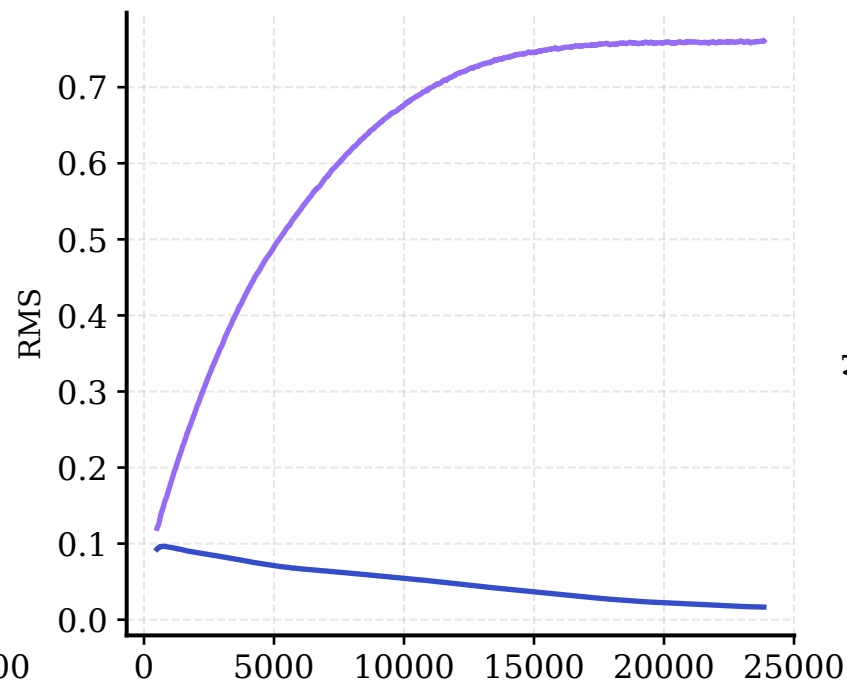


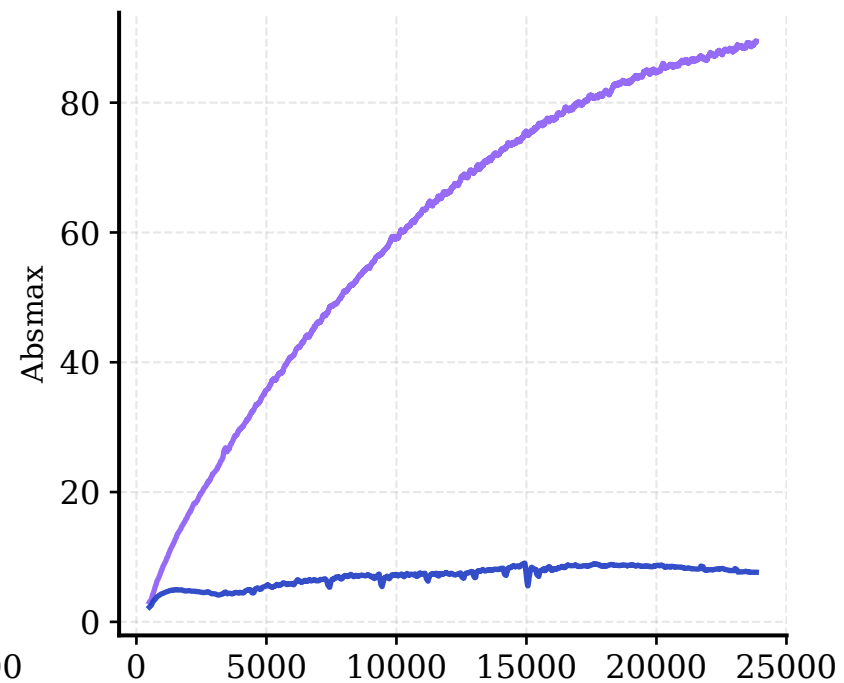
Train Loss



Attention Hidden State RMS



Attention Hidden State Absmax



— Muon without wd — Muon