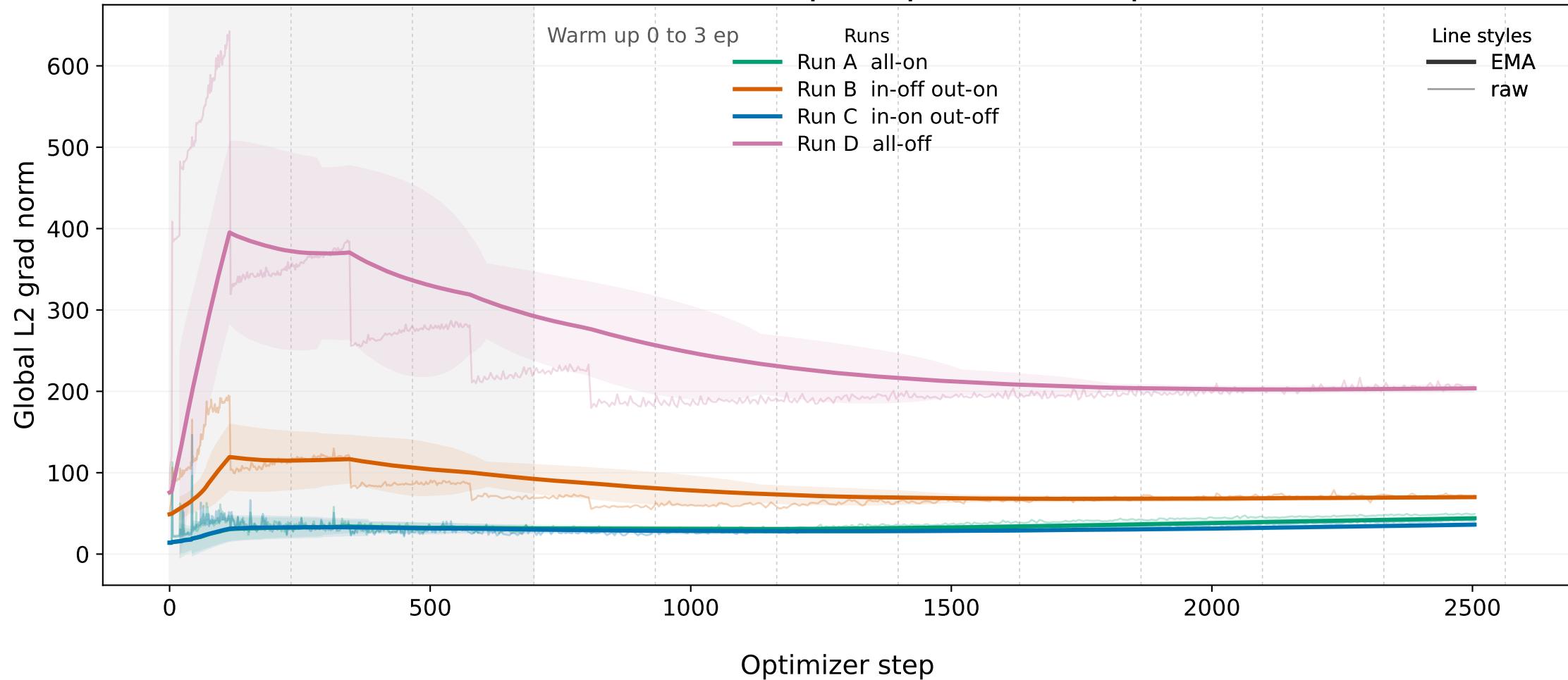


# Gradient norm per optimizer step



Lower peak and EMA at warm up end with lower plateau coefficient of variation indicate higher stability.

Metric	Run A	Run B	Run C	Run D
Peak warm up	113.1	194.9	146.7	642.6
EMA at warm up end	31.6	92.6	30.2	293.0
Plateau mean $\pm$ std last 1000	$36.4 \pm 8.4$	$88.3 \pm 32.9$	$32.6 \pm 8.3$	$284.2 \pm 124.7$
Plateau coefficient of variation	0.232	0.373	0.256	0.439