

Evolution of loss during training: unrenormalized, ada-hessian

accuracy=0.25

$\text{lr}=0.001$, $\text{betas}=(0.9, 0.999)$, $\text{weight_decay}=0.0005$, $\text{hessian_power}=1$

