Evolution of loss during training: unrenormalized, adahessian accuracy=0.25 Ir=0.001, betas=(0.9, 0.999), weight decay=0.0005, hessian power=1 train 2.50 test 2.25 2.00 1.75 1.50 -1.25 1.00 0.75 0.0 2.5 5.0 7.5 10.0 12.5 15.0 17.5 epoch