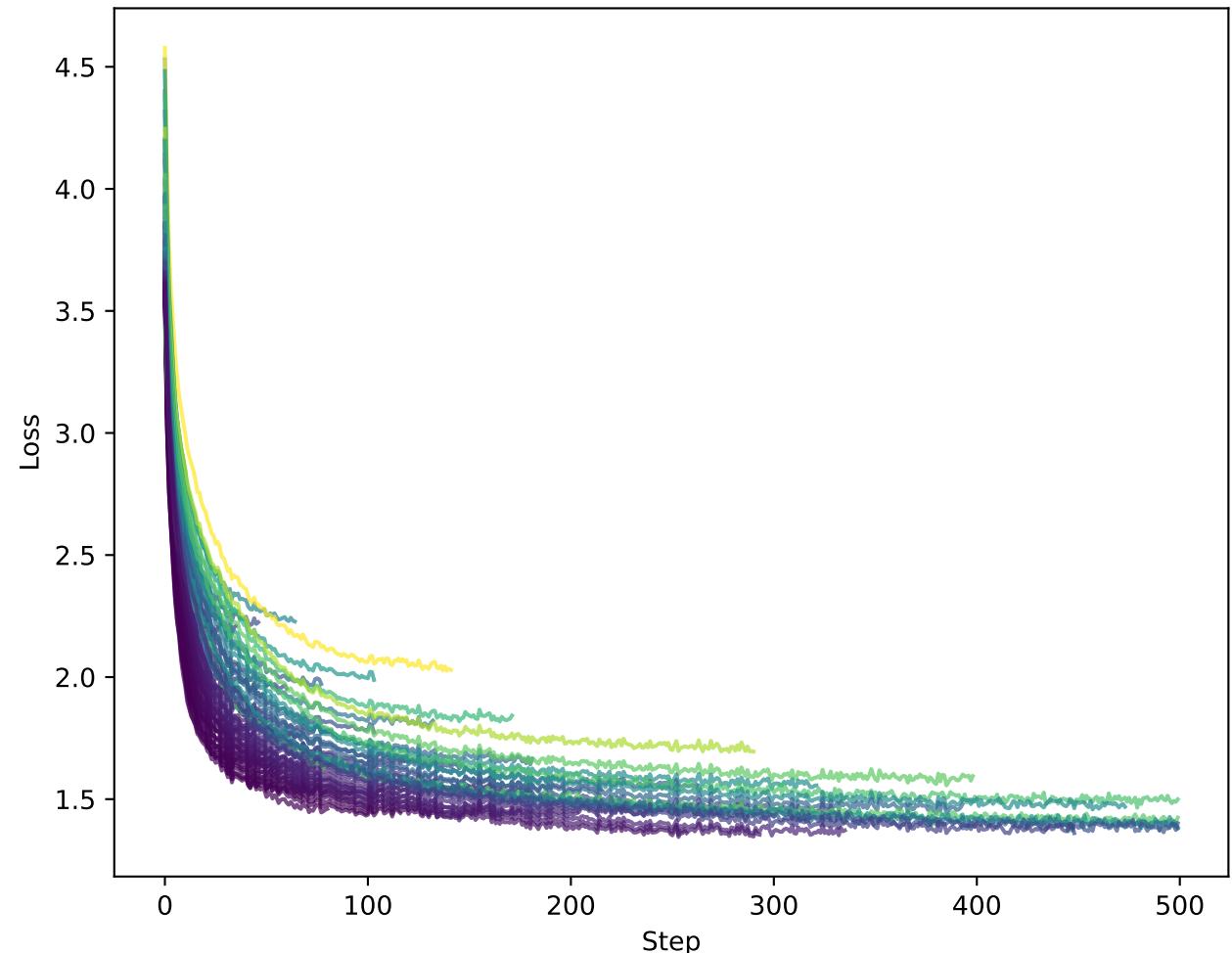


MXFP8 (E4M3 weights, E4M3 activations) Quantize Forward Only

Train CrossEntropyLoss



Optimizer Total Grad Norm

