

Convergence of the training loss as a function of the number of steps

