Convergence of the training loss as a function of the number of steps B = 8B = 16B = 32B = 64Training loss Thousand steps