

CS5260 Assignment 6

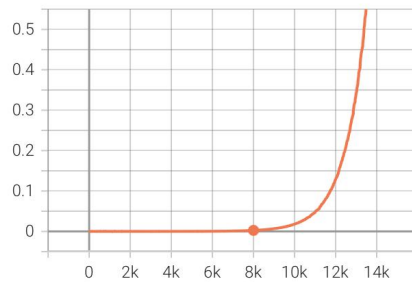
-- Colossalai and LR Range Test

In assignment 6, I use 4 setting to complete the experiment:

1. SGD optimizer + LambdaLR:

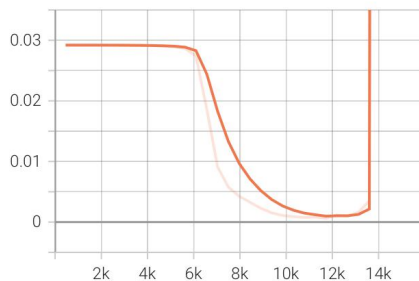
The loss starts to drop at 8k steps, with the learning rate $2.58\text{e-}3$

LR/train
tag: LR/train

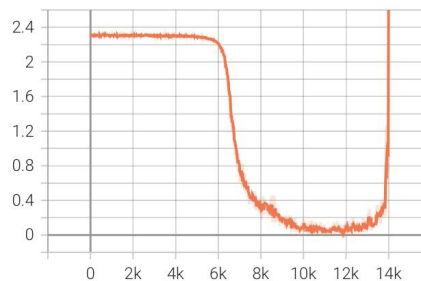


	Name	Smoothed	Value	Step	Time	Relative
Loss	ParallelMode.GLOBAL_rank_0	2.5116e-3	2.5844e-3	8.001k	Mon Mar 28, 16:38:20	5m 24s

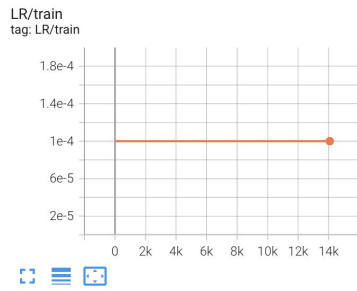
Loss/test
tag: Loss/test



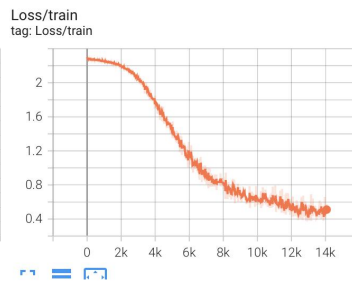
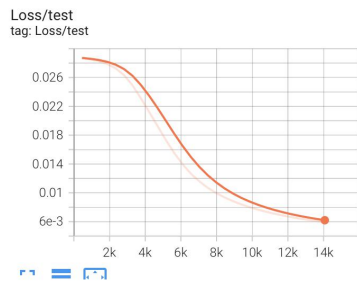
Loss/train
tag: Loss/train



2. SGD optimizer + MultiStepLR:

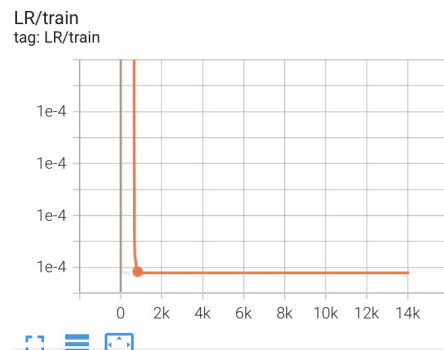


Loss

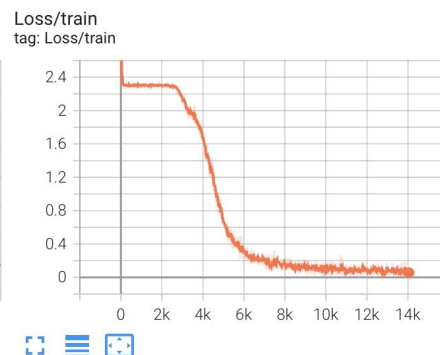
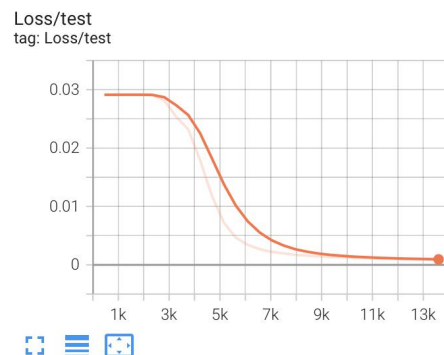


3. Adam optimizer + MultiStepLR:

The loss starts to drop at 0.8k steps, with the learning rate 1e-4. It drops greatly.

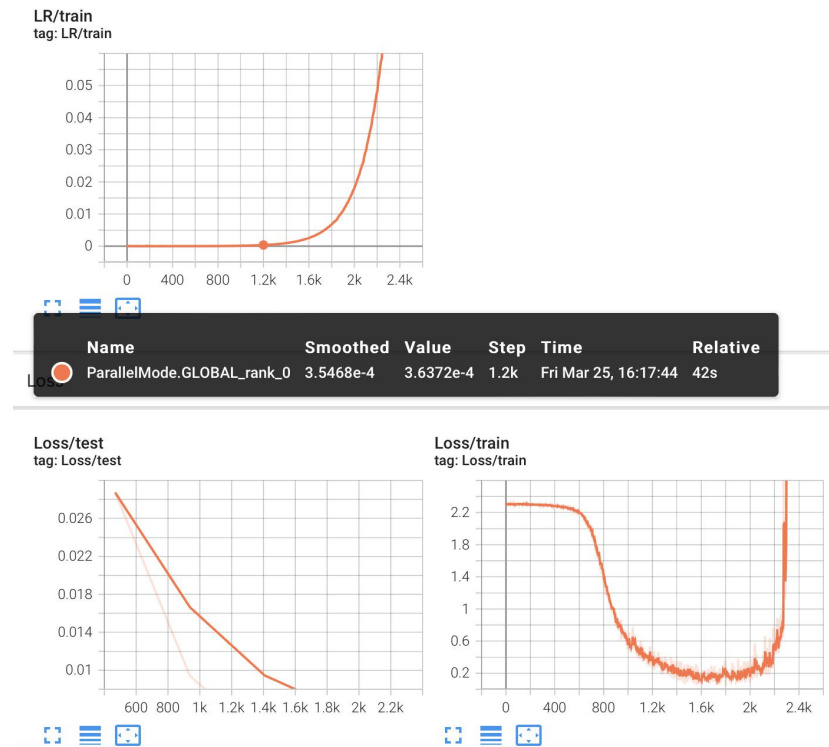


Name	Smoothed	Value	Step	Time	Relative
ParallelMode.GLOBAL_rank_0	1e-4	1e-4	828	Mon Mar 28, 17:02:07	32s



4. Adam optimizer + LambdaLR:

The loss starts to drop at 1.2k steps, with the learning rate 3.64e-3



From the above experimental results, it can be seen that compared with SGD optimizer, Adam has a larger gradient and faster convergence speed. MultiStepLR is better than LambdaLR on MNIST dataset training LeNet5.

Github link: <https://github.com/xiaoheng-byte/Colossalai>