Runtime for differentially private deep learning

Timings for Ghost Clipping

I timed the training time spent on a single batch by different implementations of the differentially private sgd algorithm.

Below are the results for a simple conv net that consists of 2 fully connected layers and 2 convolutional layers on the MNIST dataset. The batch size used is 120. The timing is the average over 500 batches.

- functorch_dp, 0.008851111078984104
- opacus_dp, 0.007145165540161542
- mixed ghost clipping, 0.010015537165221758
- ghost clipping, 0.010277516969712451
- not mixed not ghost, 0.010565361546818168
- public training, 0.0040898429183289405

4 convolutional layers on Cifar10, batch size = 100, seconds:

- functorch_dp, 0.014373262761160731
- opacus_dp, 0.008994201084831729
- mixed ghost clipping, 0.014257113210158422
- ghost clipping, 0.01808279032376595
- not mixed not ghost, use papers authors' privacy engine: 0.014138890799833461
- public, 0.00498583750706166

VGG11 on Cifar 10, batch size = 100

- functorch_dp, 0.12854671721719205
- opacus_dp, 0.08520836335443892
- ghost clipping, 0.045737753581255675
- mixed ghost clipping, 0.040571901844581586

- not mixed not ghost, use papers authors' privacy engine, 0.07100804852251895
- public, 0.007439900805708021

VGG11 on CIFAR10, batch size = 100, 20 epochs.

- functorch dp: total training time = 1430.7988447240787s
- public: total training time = 205.92813059699256s
- opacus: total training time = 1386.7571989760036s

breakdown (functorch, GPU):

- total training time = 1423.7334604880307s
- total time for per sample gradient during training: 197.47676673321985
- total time for clipping during training: 23.502324198838323
- total time for noising during training: 11.241958117345348
- total time for getting data batch: 1022.9345700764097
- total time for optimizer step: 22.030423796037212

functorch, CPU: for every epoch:
per sample gradient time: 2160.2697375514545s
clipping time: 79.99212361080572s
noising time: 41.36028096359223s
data batch time: 0.0011884046252816916s
optimizer step time: 7.1352214510552585s
total time for per sample gradient during training: 40573.62095386046
total time for clipping during training: 1569.095354638528
total time for noising during training: 823.5452335844748
total time for getting data batch: 0.023356984835118055
total time for optimizer step: 140.10640525864437
total training time = 43255.40618532291s

Multivariate Guassian Noise in JAX vs Pytorch