Experiments with Logistic Regression: Fast vs Slow Communication

We now repeat the experiments from Section I of the paper but with logistic regression on *MNIST* dataset (LeCun et al., 2010) with 100 workers. We consider two regimes: fast and slow communication between workers. One can see that when the communication is fast, the gap between the methods is small, which is expected and compliant with the theory. However, Fragile SGD is much faster and has better test accuracy when the communication is slow.

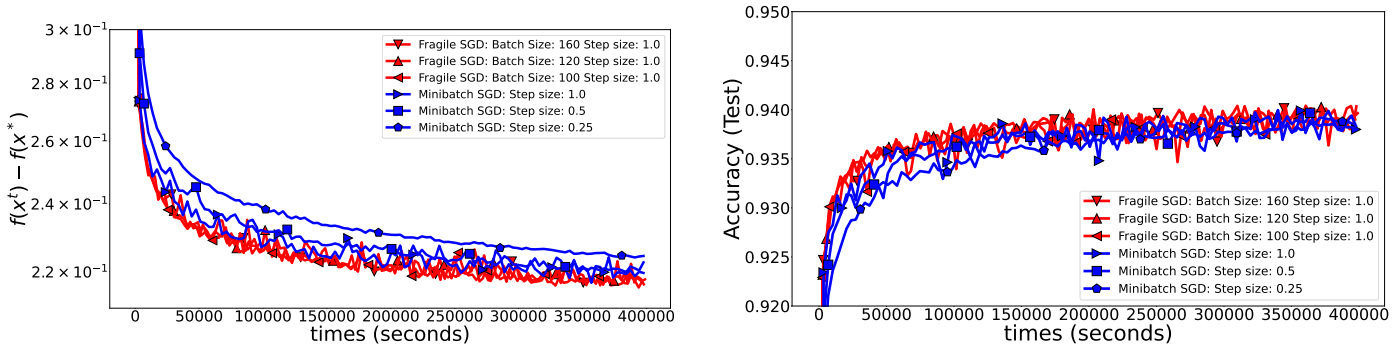

Figure 1: The communication time $\rho = 0.1$ seconds (Fast communication)

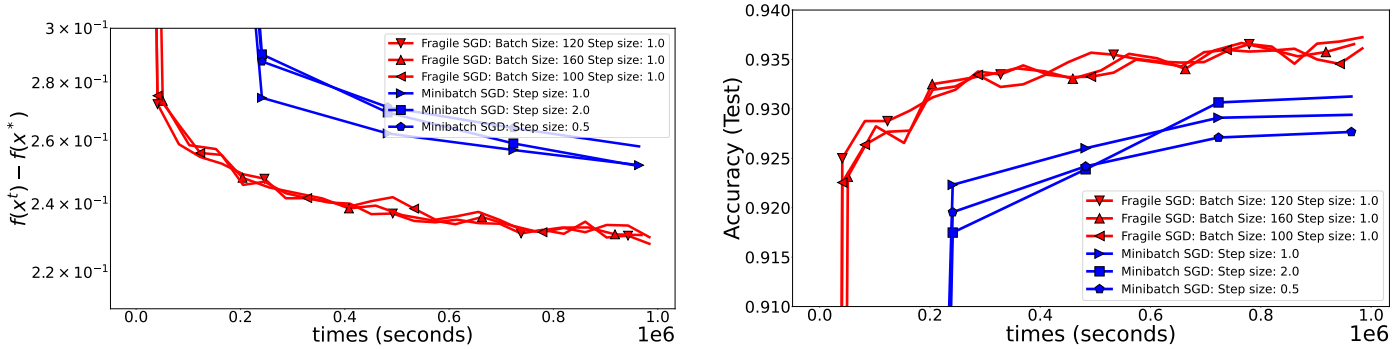

Figure 2: The communication time $\rho = 10$ seconds (Slow communication)

Experiments with ResNet-18

We test algorithms on an image recognition task, *CIFAR10* (Krizhevsky et al., 2009), with the *ResNet-18* (He et al., 2016) deep neural network (the number of parameters $d \approx 10^7$). We use the torus structure and 9 workers. We run all methods with the step sizes $\{0.025, 0.25, 2.5\}$. Our findings from the low-scale experiments are also evident in the large-scale experiments. Fragile SGD converges faster than Minibatch SGD in terms of function values. When we compare accuracies on the test split of *MNIST*, the superiority of Fragile SGD is even more transparent.

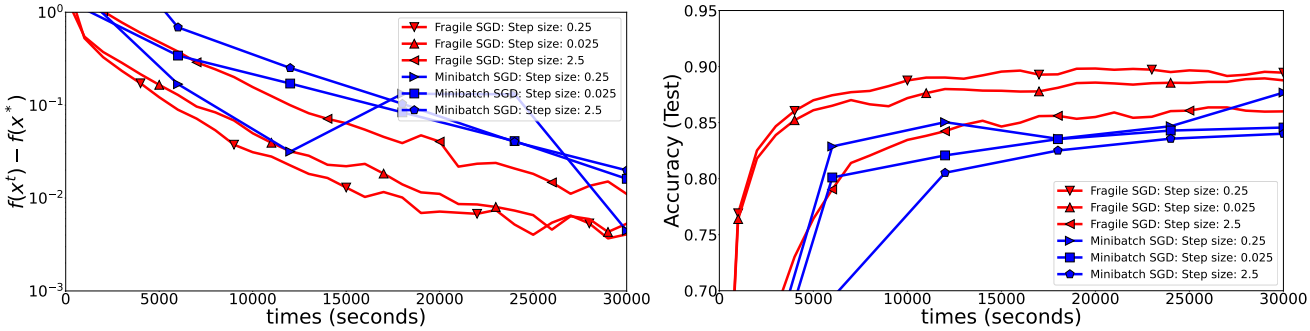

Figure 3: *ResNet-18* on *CIFAR10* dataset with 9 workers and the torus structure with the communication time $\rho = 1$ seconds (Medium communication)