

---

## Supplementary Material

### Active Probabilistic Inference on Matrices for Pre-Conditioning in Stochastic Optimization

---

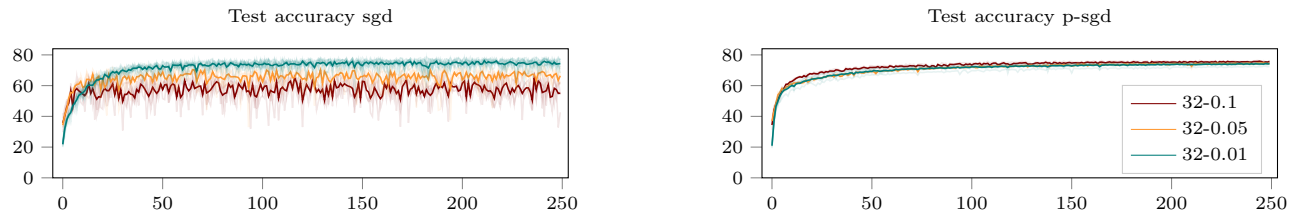


Figure 6: Test accuracy for sgd (left) and pre-conditioned sgd (right) on the CIFAR-10 data-set for different learning rates and batch-size of 32 over 250 epochs. Both graphs share y-axis and colors to facilitate comparison between the optimizers. These graphs were collected at the same runs as the results in Fig. 3.

## A Deep Learning

All deep learning experiments used a neural network consisting of 3 convolutional layers with 64, 96 and 128 output channels of size  $5 \times 5$ ,  $3 \times 3$  and  $3 \times 3$  followed by 3 fully connected layers of size 512, 256, 10 with cross entropy loss function on the output and  $L_2$  regularization with magnitude 0.01. All layers used the ReLU nonlinearity and the convolutional layers had additional max-pooling.