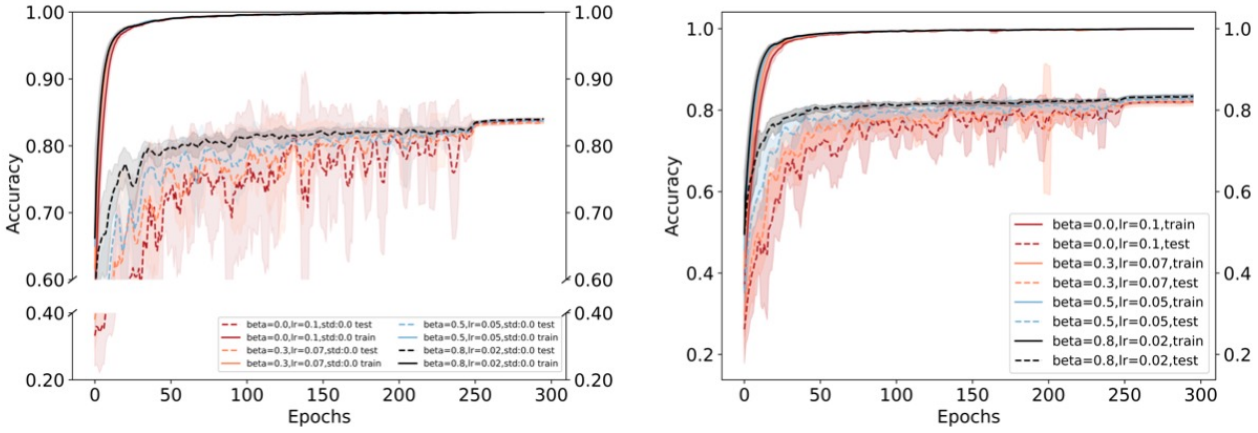
5.2 Effect of learning-rate scheduler

Learning rate schedulers are a common practice in training classification networks hence exploring the effect of IGR and IGR-M in schedulers is important. In the experiment, we train a Resnet-18 and a Resnet-50 network to classify CIFAR-10 dataset to compare the performance of (SGD) and (SGD+M) under the effect of learning rate scheduler.

We observe that just like our previous experiments comparing (SGD) and (SGD+M), the test accuracy is higher with increasing β. We attributed this effect due to the stronger implicit regularization for momentum than plain SGD. However, after the effect of scheduler, the learning rate is decreased by a factor of 10. This diminishes the effect of the implicit regularizer for both SGD and SGD+M as $IGR \propto h$. However, from empirical observations (Fig-2) the difference in test accuracy of (SGD) and (SGD+M) (near convergence) still exists but may not be in a pronounced way as the initial iterations. We believe this is because during the earlier iterations, the significantly stronger IGR for (SGD+M) guides it's trajectory through flatter sub-manifolds than that of (SGD). The effect is prominent enough that even after scheduler is activated (also



(a) Resnet-18 (b) Resnet-50

Figure 2: Classification result of CIFAR-10 with step scheduler ($\frac{1}{10}$) activated at epoch =250 with various β but the same effective learning rate $\frac{h}{(1-\beta)}$.

near convergence), (SGD+M) still has a slightly higher test accuracy than (SGD).