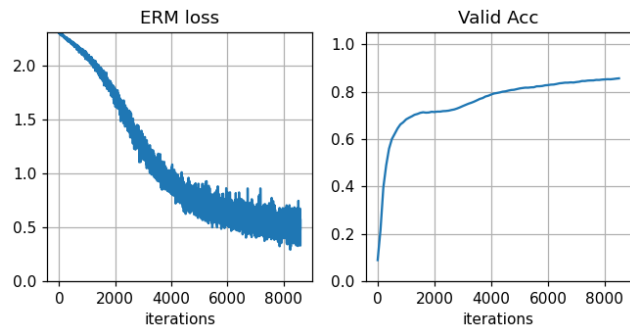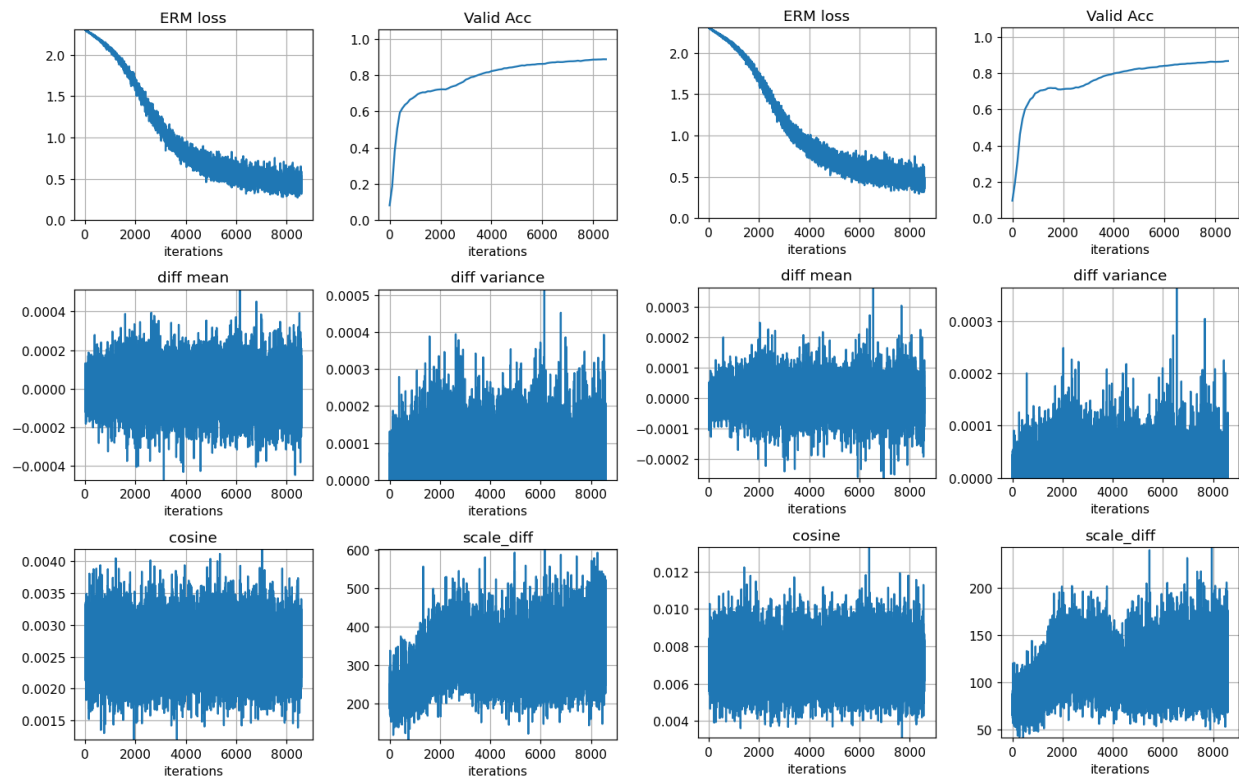# The projection method:

Setting: MLP (3M parameters); MNIST; 1e-3 SGD; Num_dirs = 20

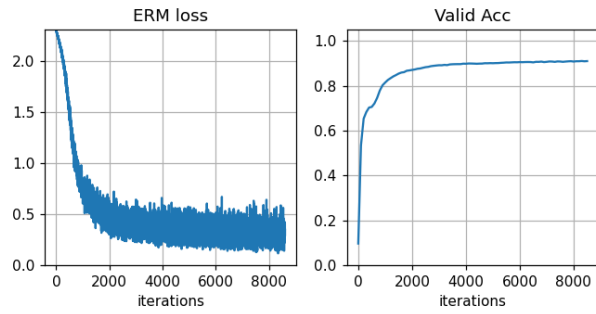Backpropagation baseline:

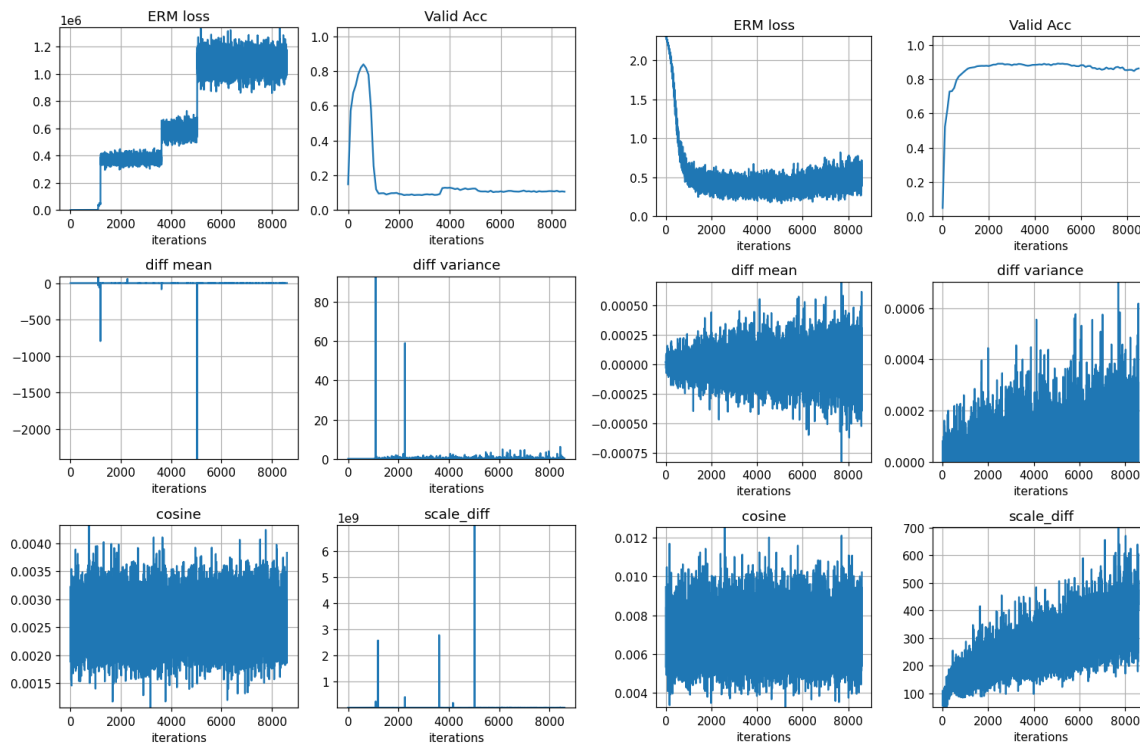Naïve forward-mode/Projection forward-mode:

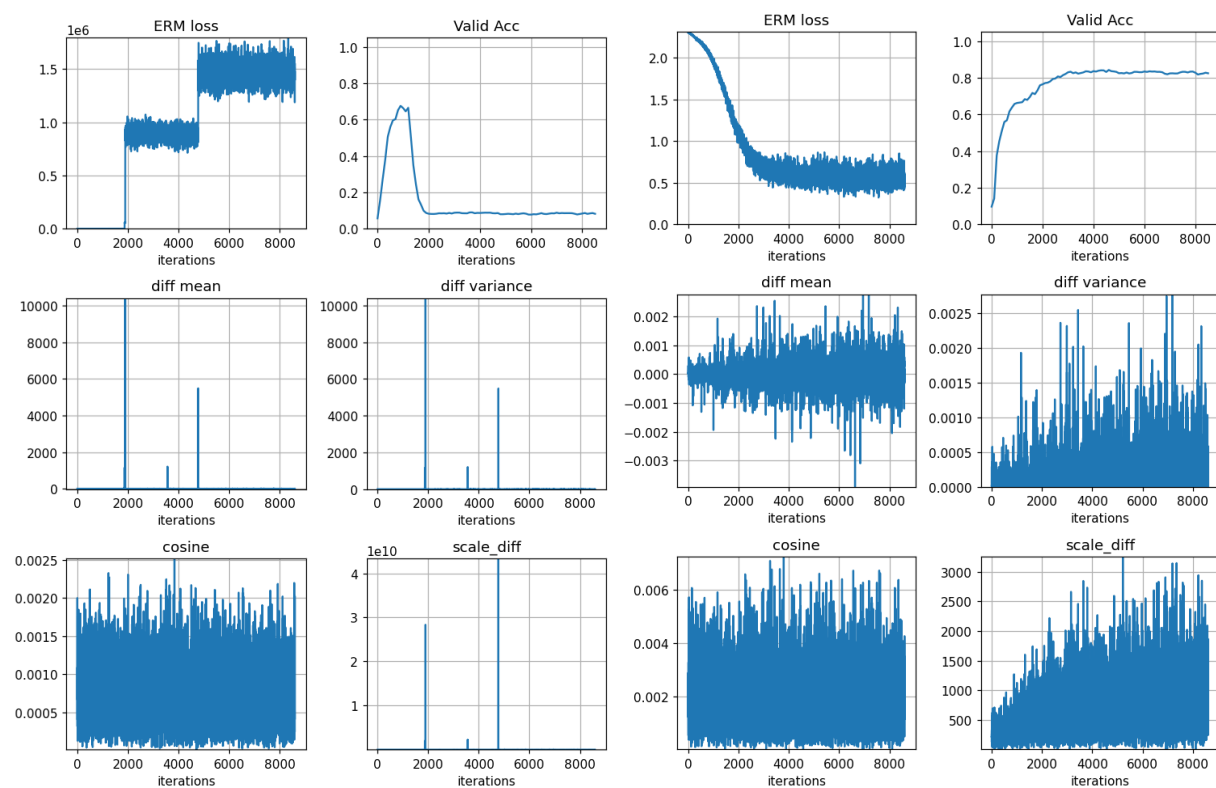## Compare to setting 1, enlarge the lr to 5e-3:
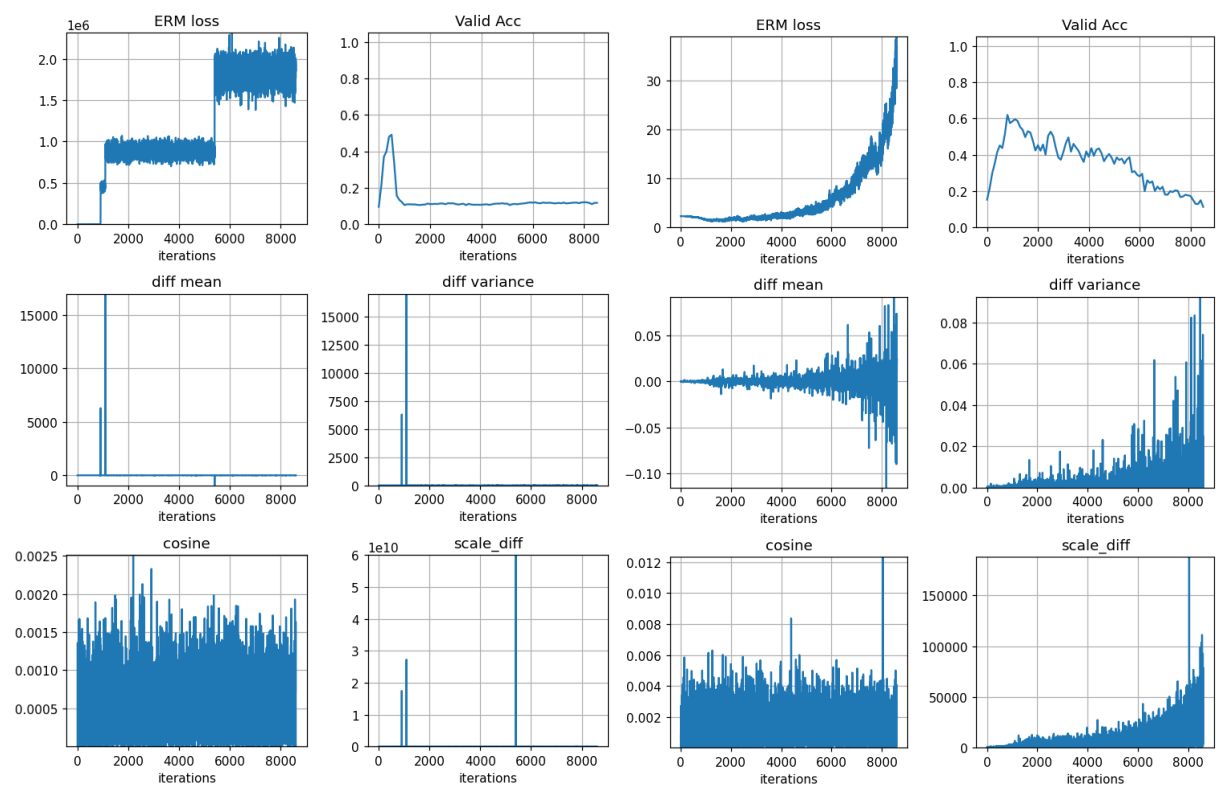
Backpropagation baseline:
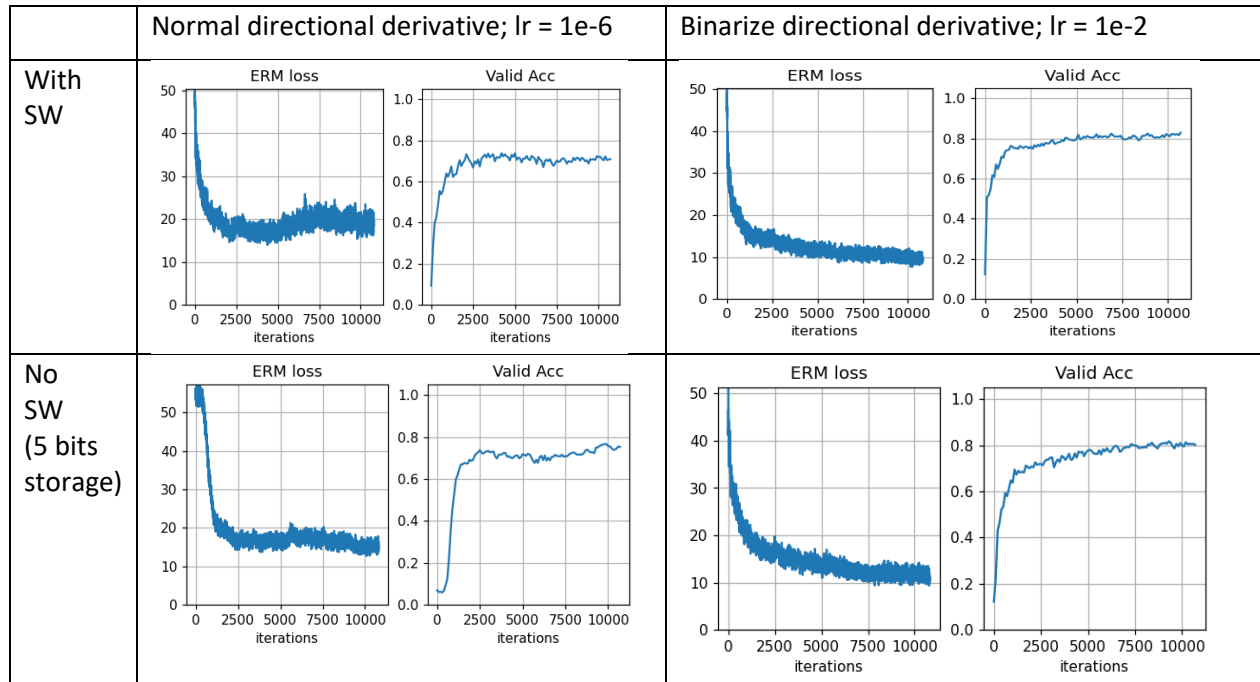

Naïve forward-mode/Projection forward-mode:
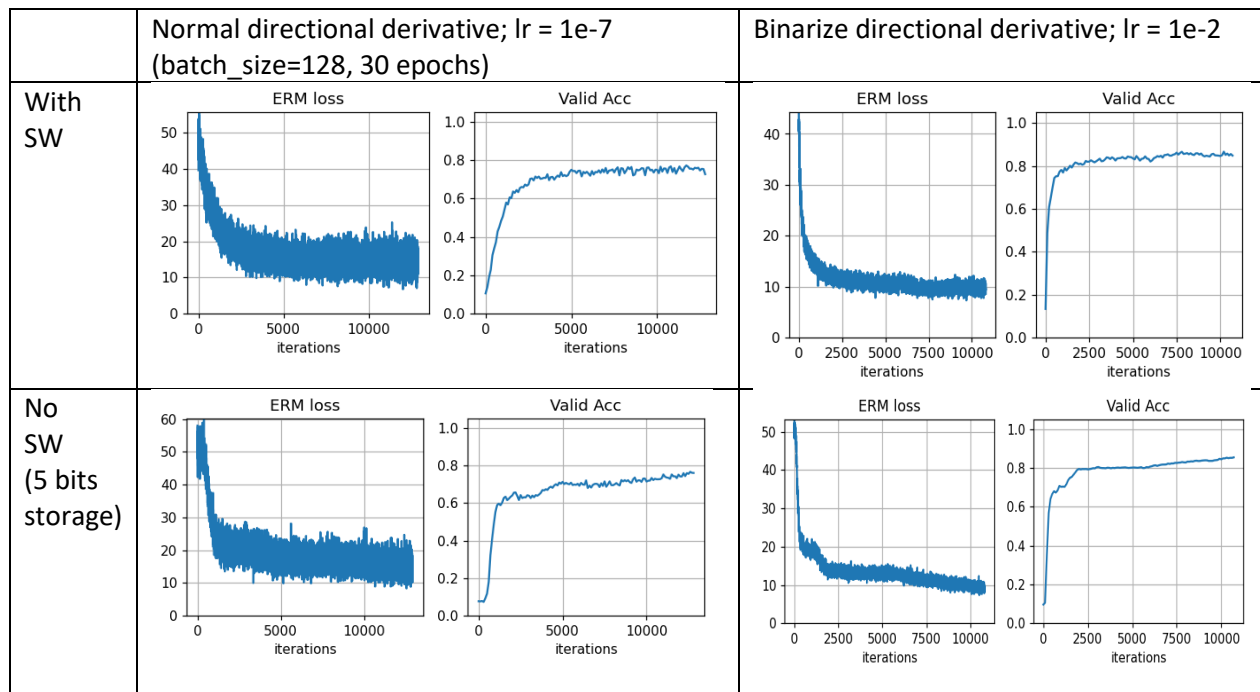
Compare to setting 1, change num_dir = 2:

Compare to setting 1, change num_dir = 1:

Try on BCNN (Where FC layer is still dominant):

Chart from last meeting.

	Normal directional derivative; lr = 1e-6	Binarize directional derivative; lr = 1e-2
With SW		
No SW (5 bits storage)		

Extra experiment that using projection to generate random vectors:

	Normal directional derivative; lr = 1e-7 (batch_size=128, 30 epochs)	Binarize directional derivative; lr = 1e-2
With SW		
No SW (5 bits storage)		

Training VGG16 ~ 19 million parameters (Not binary version) on Cifar-10:（Different structure from last meeting; 60% of the parameters are from linear layer's weight）

Backprop baseline:
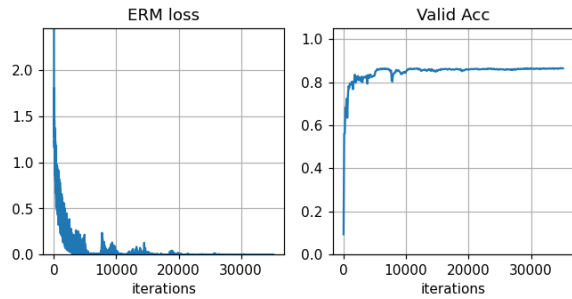
Naïve Forward/Forward using projection, num_dir = 25