MLP\_MNIST Ablation: GGD\_LW\_ablate\_scale\_aware: MLP\_MNIST Ablation GGD\_LW\_ablate\_scale\_aware: Loss Optimizer GGD\_LW\_baseline\_scale\_aware GGD LW scale aware False 1.6 1.4 1.2 Average Loss
8.0 0.6 0.4 0.2 2 6 10 12 14

**Epoch** *Error bands represent standard error of the mean across runs*