Training Loss vs. Wall-clock Time for MLP\_MNIST Ablation: MILO\_LW\_ablate\_scale\_factor  $6 \times 10^{0}$ **Optimizer** MILO LW scale factor 0.2 MILO\_LW\_scale\_factor\_0.1 **-x-** MILO\_LW\_scale\_factor\_0.5  $4 \times 10^{0}$  $3 \times 10^{0}$ Loss (Training)  $5\times 7\times 10^{0}$  $10^{0}$ 10 0 2 6 12 14 16 Time (s)

Error bands represent standard error of the mean across runs. Markers indicate run end on walltime plots.