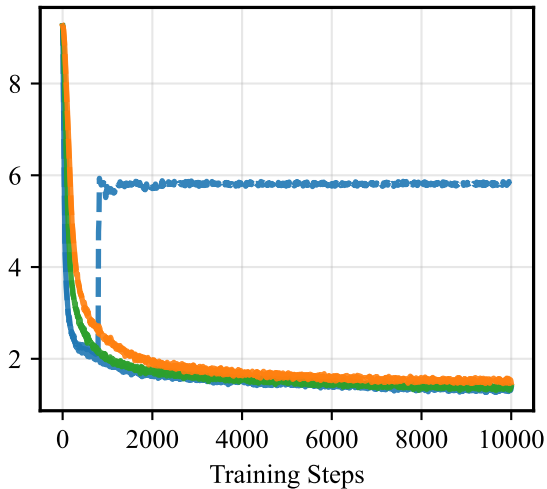
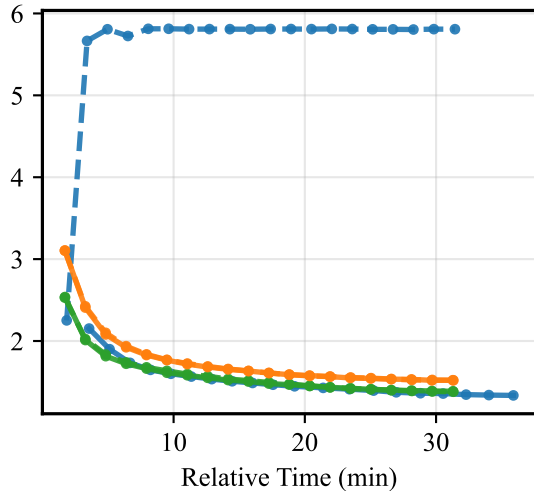


# Training Loss



# Validation Loss



$lr = 0.01$  Pre-Norm       $lr = 0.001$  Pre-Norm       $lr = 0.0003$  Pre-Norm  
 $lr = 0.01$  Post-Norm       $lr = 0.001$  Post-Norm       $lr = 0.0003$  Post-Norm