Language Model(350M) loss curve GPT-method2 3.4 **GPT-method3** GPT-method3-scale GPT-method5-scale GPT-sincos-method4.2 3.2 GPT-sincos-method4.2-scale **Training Loss** LLaMA-method2 LLaMA-method4.2 2.8 2.6 0 20 40 60 80 100 Number of iterations(k)