Language Model(774M) loss curve GPT-method2 3.4 GPT-method5-scale GPT-sincos-method4.2-scale LLaMA-method2 3.2 Training Loss 2.6 0 40 60 20 80 100 Number of iterations(k)