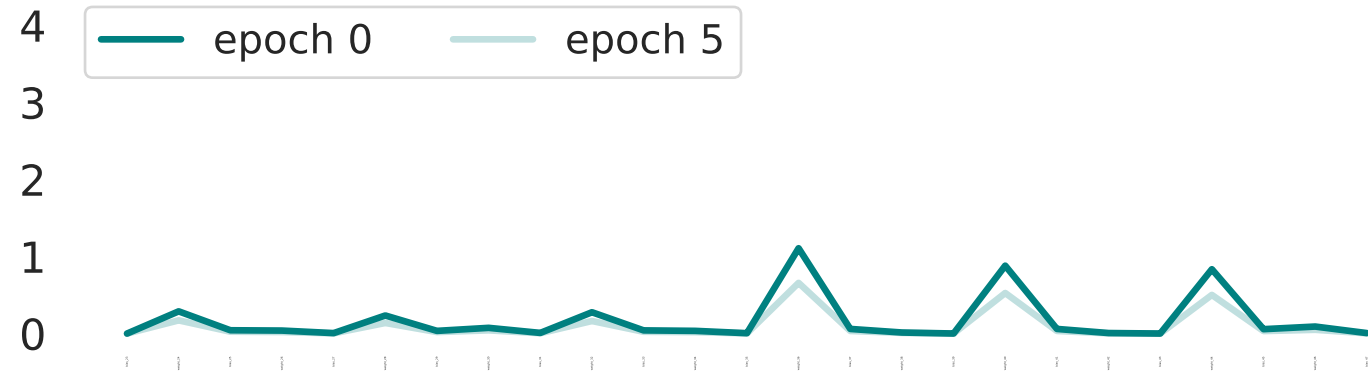
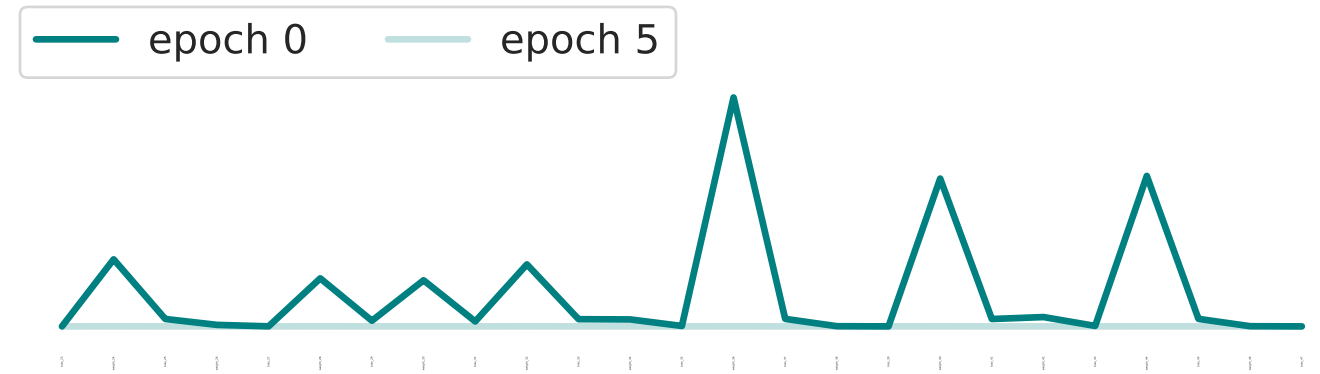


Gradient flow when increasing pre-training epochs

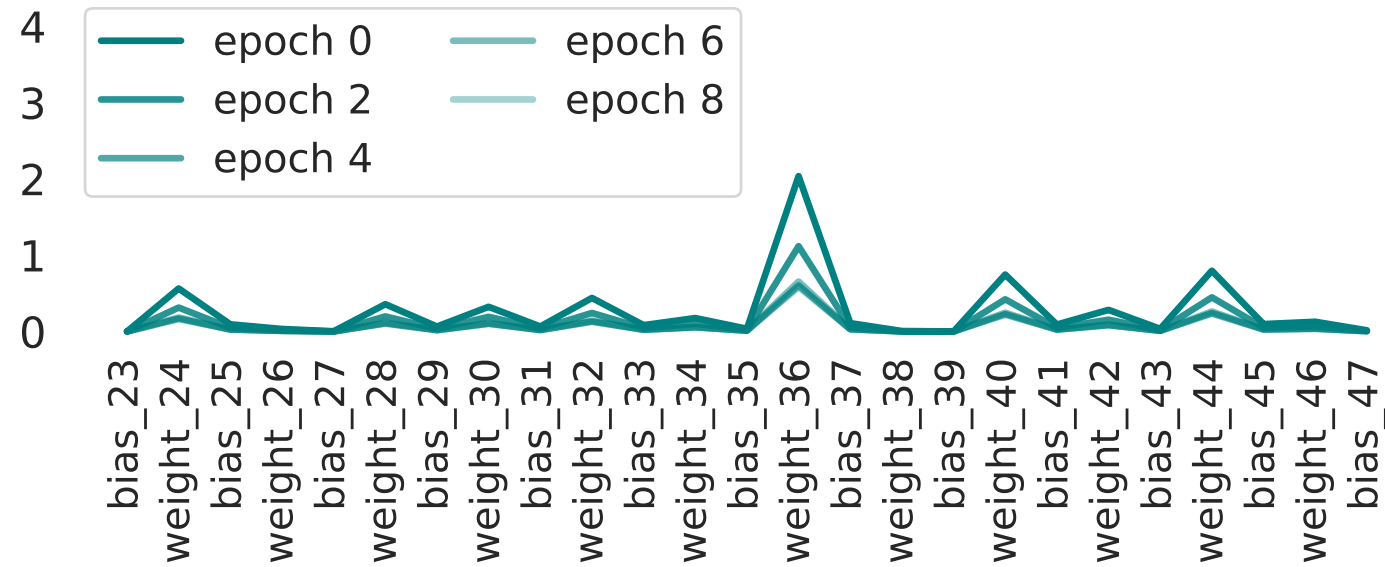
5 pretraining epochs



10 pretraining epochs



5 pretraining epochs + lr decay



10 pretraining epochs + lr decay

