Best/Worst Grad Flow of Adaptive LR Optimizers Every 10 Epochs
Best Grad (batch\_size, learning\_rate): (32, 1)
Worst Grad (batch\_size, learning\_rate): (900, 0.001) Weight Gradients Weight Gradients Bias Gradients Bias Gradients adadelta -Best Grad (batch size, learning rate): (32, 1) Worst Grad (batch\_size, learning\_rate): (900, 0.001) Weight Gradients Weight Gradients Bias Gradients Bias Gradients adagrad --Best Grad (batch\_size, learning\_rate): (32, 0.01) Worst Grad (batch\_size, learning\_rate): (900, 0.001) Weight Gradients Weight Gradients Bias Gradients Bias Gradients adam Best Grad (batch\_size, learning\_rate): (32, 0.001) Worst Grad (batch size, learning rate): (900, 0.001) Weight Gradients Weight Gradients Bias Gradients Bias Gradients rmsprop **Epoch** Epoch