Best/Worst Grad Flow of Const. LR Optimizers Every 10 Epochs

Best Grad (batch_size, learning_rate, momentum): (1, 0.01, 0.1)

Worst Grad (batch_size, learning_rate, momentum): (900, 0.001, 0.1)

Best Grad (batch_size, learning_rate, momentum): (1, 0.001, 0.9)

Worst Grad (batch_size, learning_rate, momentum): (900, 0.001, 0.1)

Best Grad (batch_size, learning_rate): (1, 0.01)

Worst Grad (batch_size, learning_rate): (900, 0.001)

Legend:
- Weight Gradients
- Bias Gradients

Y-axis labels: momentum, nesterov, sgd

X-axis label: Epoch