Gradient norm vs iteration for various learning rates s