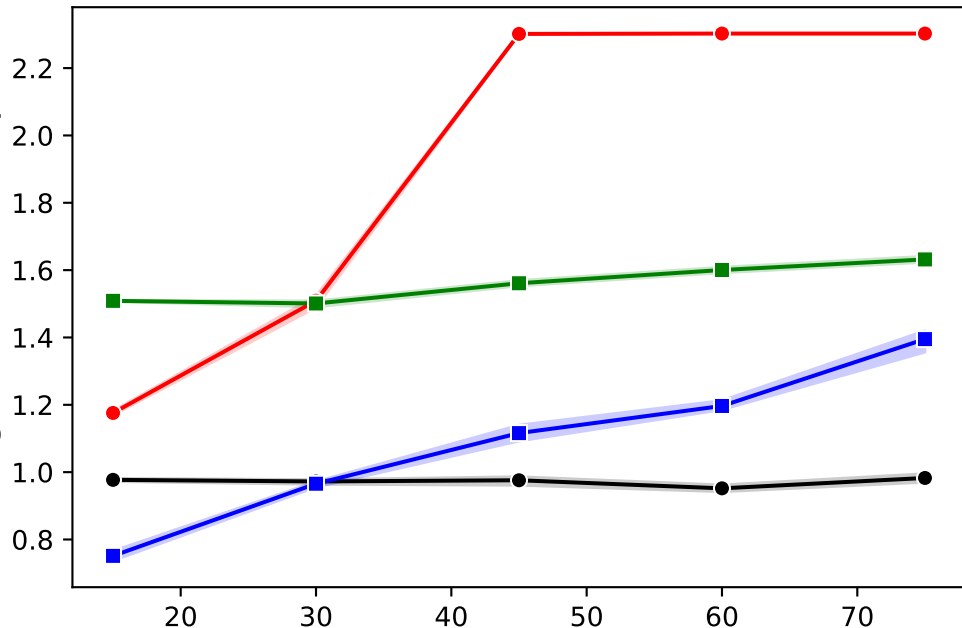


Training loss after 30 epochs



—●— iterative orthogonalization

—■— standard initialization

—■— adaptive gd clipping

—●— orthogonal weights