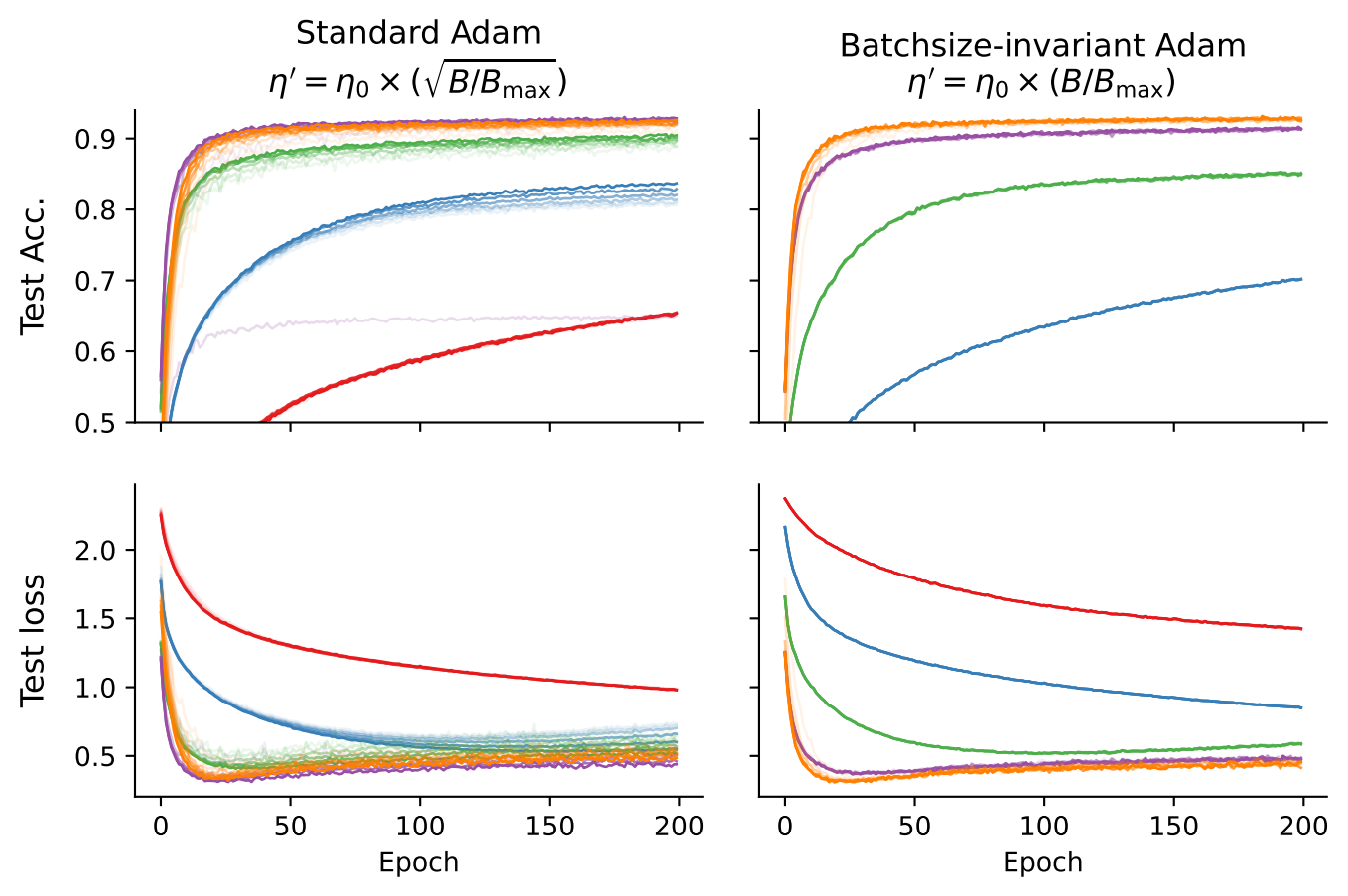
Standard Adam
$\eta' = \eta_0 \times (\sqrt{B/B_{\max}})$
Batchsize-invariant Adam
$\eta' = \eta_0 \times (B/B_{\max})$
Test Acc.
0.9
0.8
0.7
0.6
0.5
Test loss
2.0
1.5
1.0
0.5
0
50
100
150
200
Epoch
0
50
100
150
200
Epoch