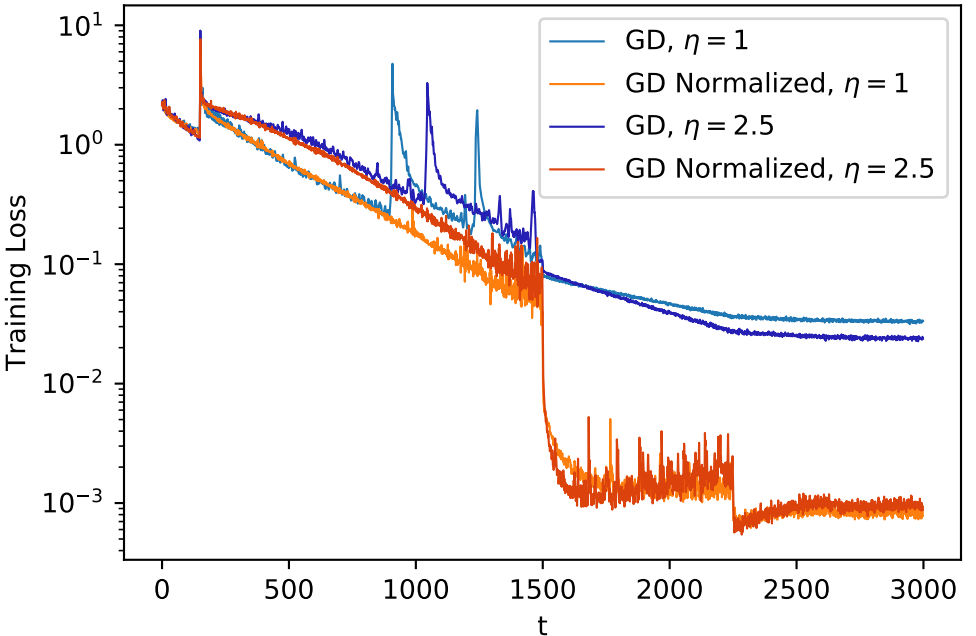

GD, $\eta = 1$
GD Normalized, $\eta = 1$
GD, $\eta = 2.5$
GD Normalized, $\eta = 2.5$
Training Loss
10^{1}
10^{0}
10^{-1}
10^{-2}
10^{-3}
0
500
1000
1500
2000
2500
3000
t