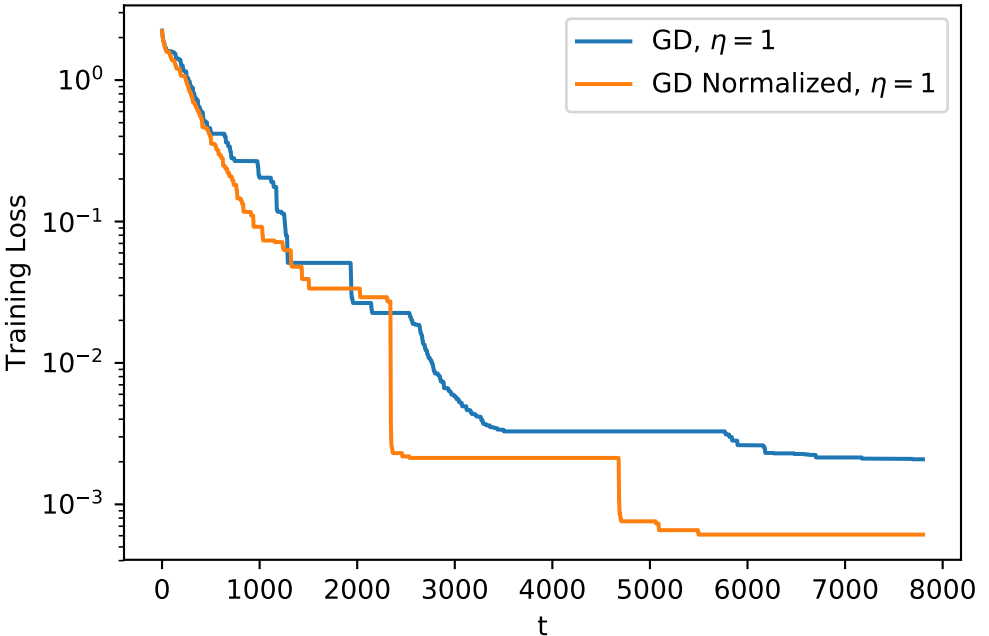

GD, $\eta = 1$
GD Normalized, $\eta = 1$
Training Loss
10^{0}
10^{-1}
10^{-2}
10^{-3}
0
1000
2000
3000
4000
5000
6000
7000
8000
t