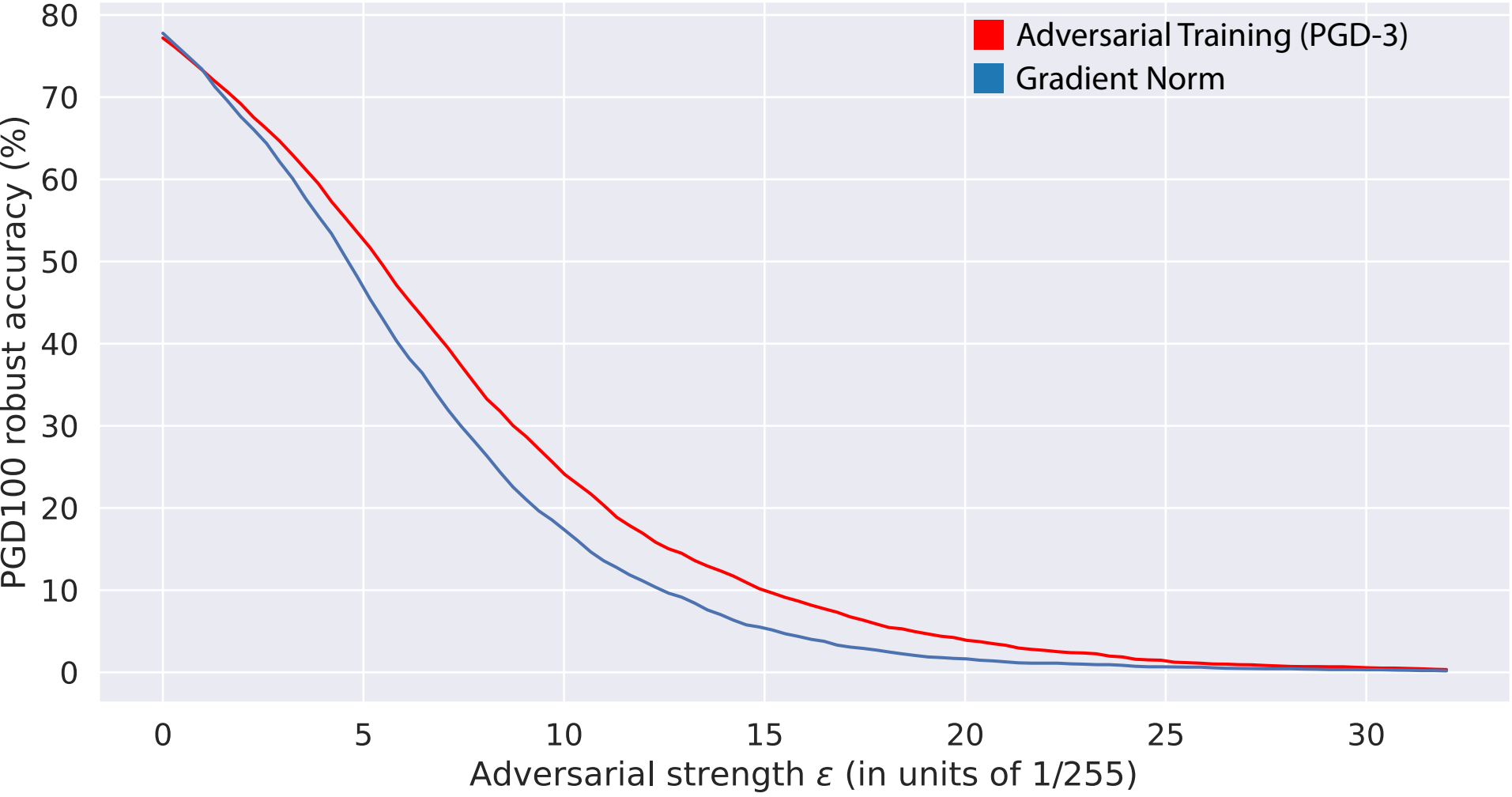
Adversarial Training (PGD-3)
Gradient Norm
PGD100 robust accuracy (%)
80
70
60
50
40
30
20
10
0
0
5
10
15
20
25
30
Adversarial strength ε (in units of 1/255)