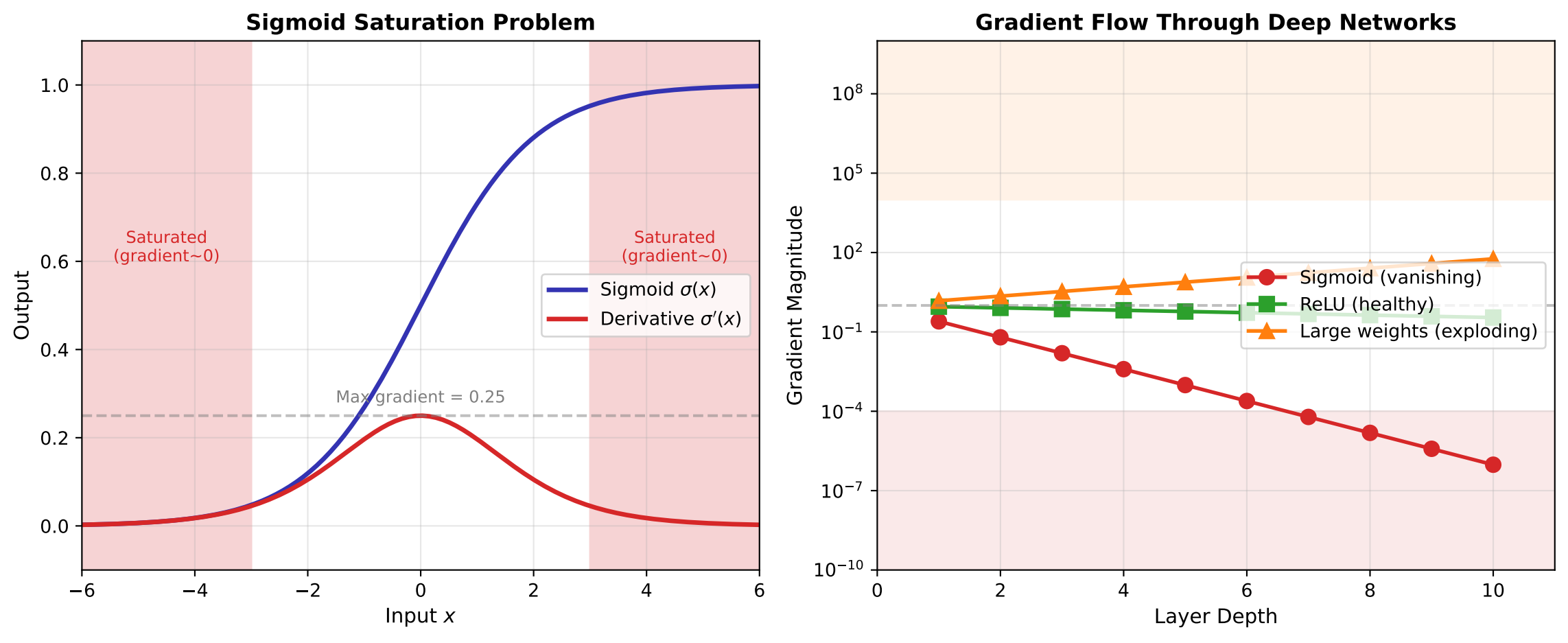
Sigmoid Saturation Problem
Saturated
(gradient~0)
Saturated
(gradient~0)
Max gradient = 0.25
Sigmoid $\sigma(x)$
Derivative $\sigma'(x)$
Output
Input x
1.0
0.8
0.6
0.4
0.2
0.0
−6
−4
−2
0
2
4
6
Gradient Flow Through Deep Networks
Sigmoid (vanishing)
ReLU (healthy)
Large weights (exploding)
Gradient Magnitude
Layer Depth
10^{8}
10^{5}
10^{2}
10^{-1}
10^{-4}
10^{-7}
10^{-10}
0
2
4
6
8
10