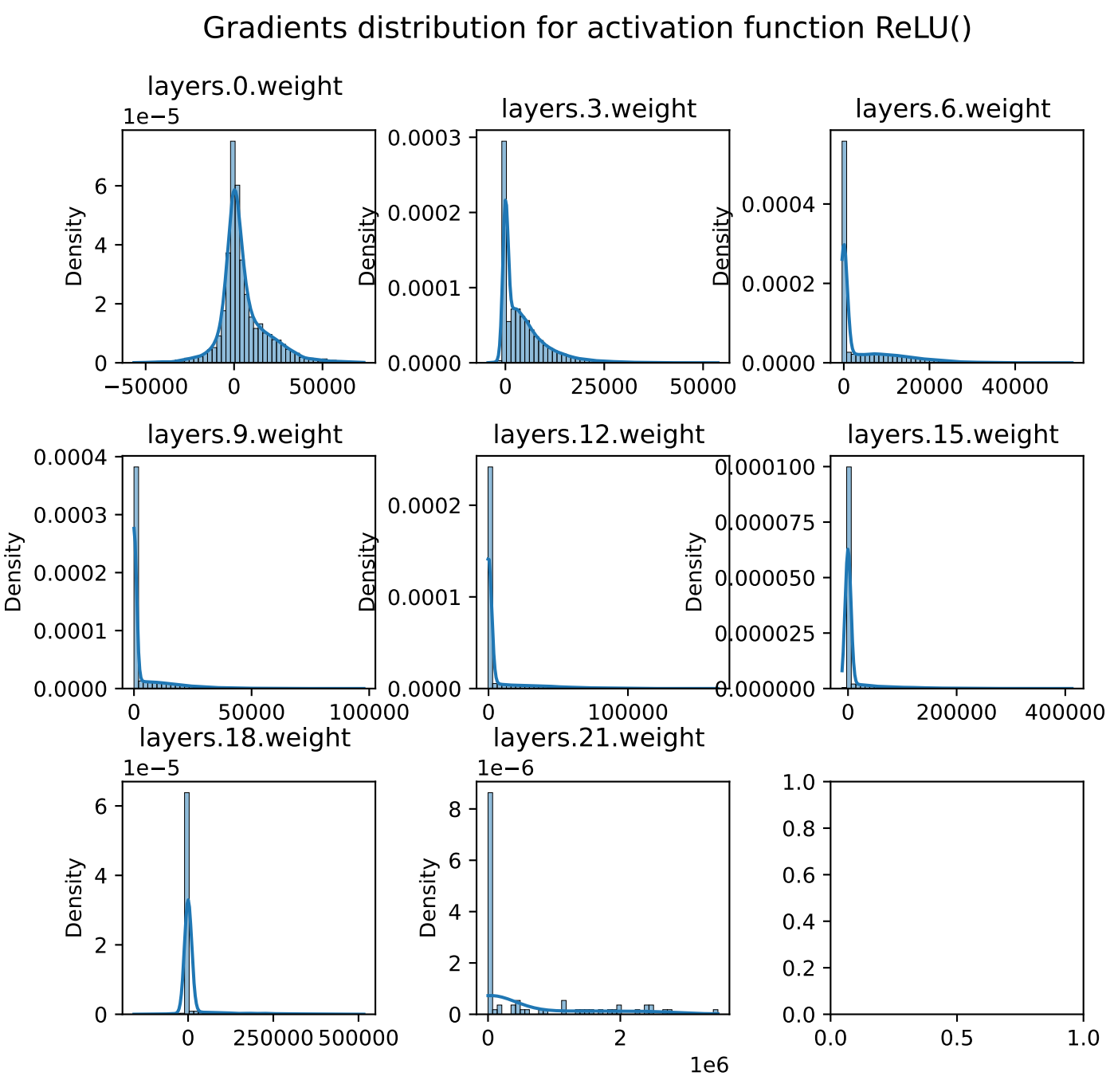
Gradients distribution for activation function ReLU()
layers.0.weight
1e−5
Density
layers.3.weight
layers.6.weight
layers.9.weight
layers.12.weight
layers.15.weight
layers.18.weight
1e−5
layers.21.weight
1e−6
1e6