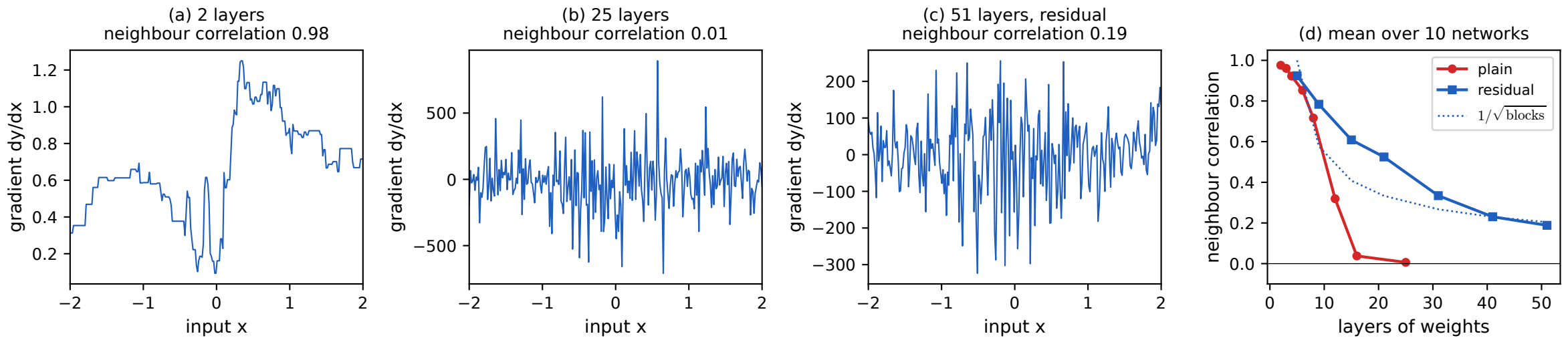

(a) 2 layers
neighbour correlation 0.98
gradient dy/dx
input x
(b) 25 layers
neighbour correlation 0.01
gradient dy/dx
input x
(c) 51 layers, residual
neighbour correlation 0.19
gradient dy/dx
input x
(d) mean over 10 networks
neighbour correlation
layers of weights
plain
residual
$1/\sqrt{\text{blocks}}$