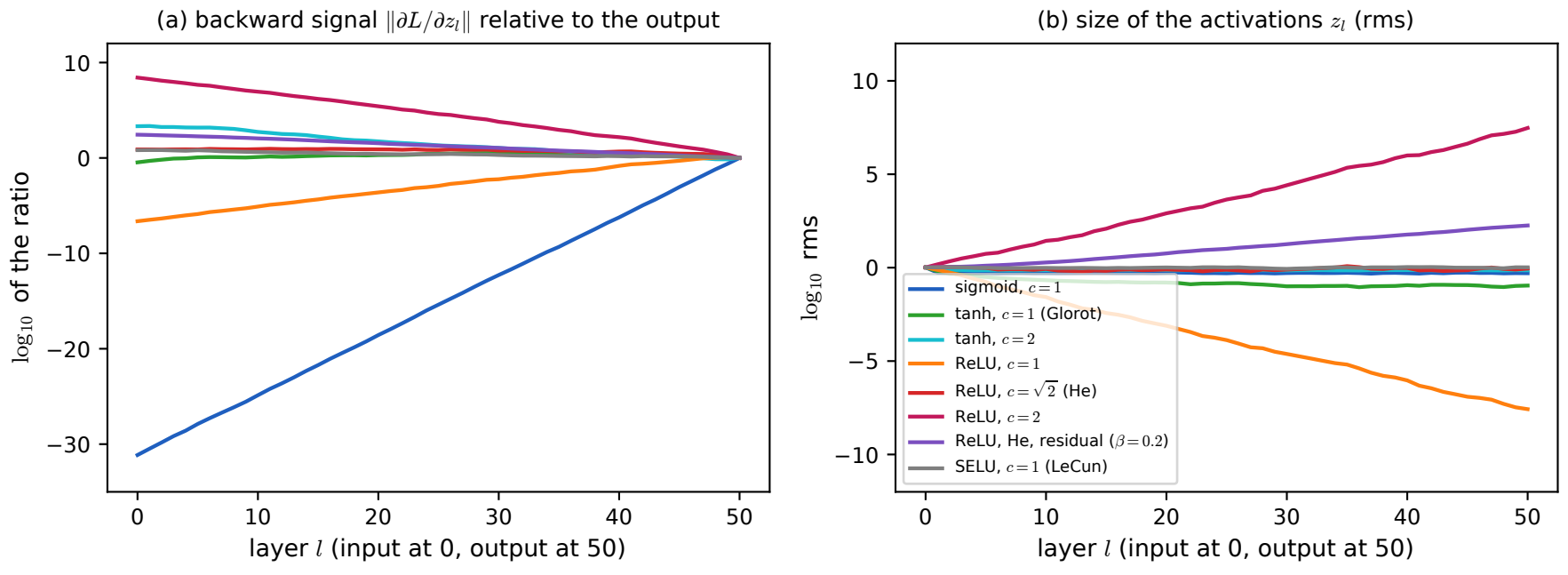

(a) backward signal $\|\partial L/\partial z_l\|$ relative to the output
$\log_{10}$ of the ratio
layer l (input at 0, output at 50)
(b) size of the activations z_l (rms)
$\log_{10}$ rms
layer l (input at 0, output at 50)
sigmoid, $c=1$
tanh, $c=1$ (Glorot)
tanh, $c=2$
ReLU, $c=1$
ReLU, $c=\sqrt{2}$ (He)
ReLU, $c=2$
ReLU, He, residual ($\beta=0.2$)
SELU, $c=1$ (LeCun)