Vanilla RNN Gradient Flow

×0.9   ×0.9   ×0.9

t=0   t=1   t=2   t=3

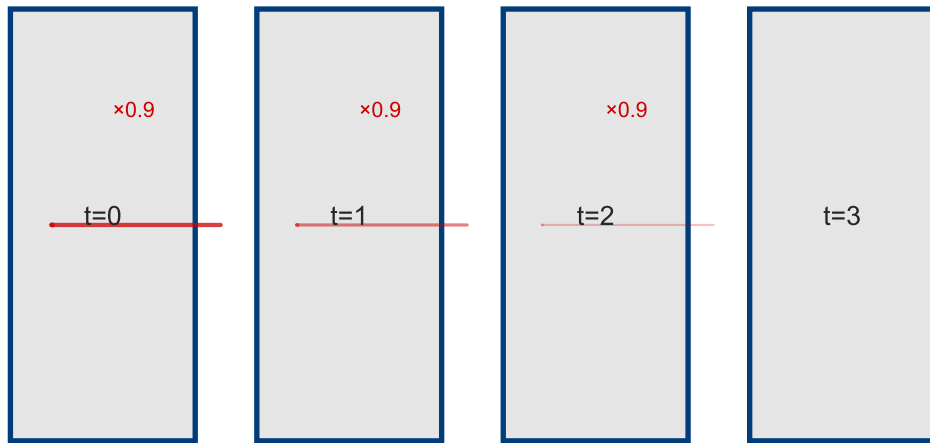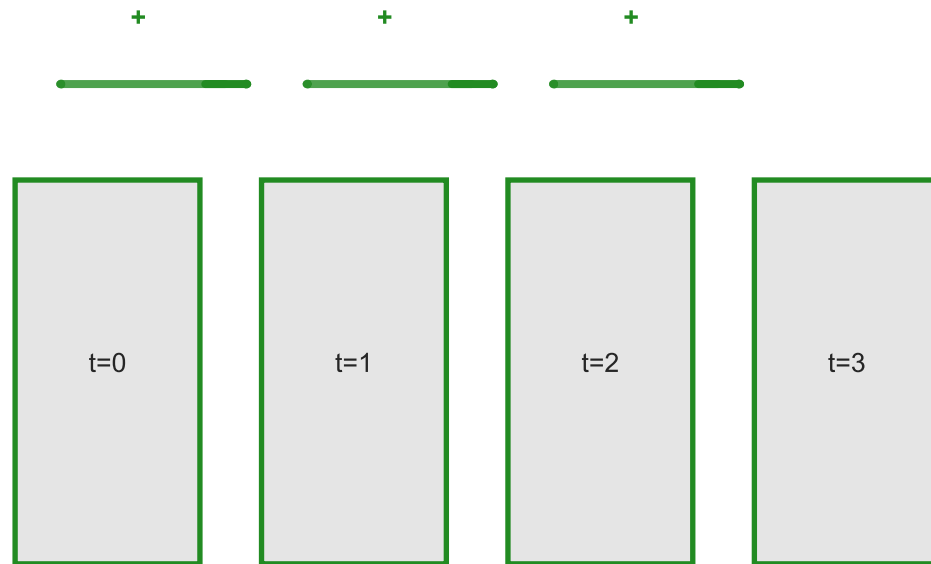*Gradient vanishes through multiplications*

LSTM Gradient Flow

+   +   +

t=0   t=1   t=2   t=3

Gradient flows through cell state highway