Vanishing Gradient Problem in RNNs

Gradient Magnitude (log scale) vs Time Steps

- RNN (factor=0.9)
- RNN (factor=0.8)
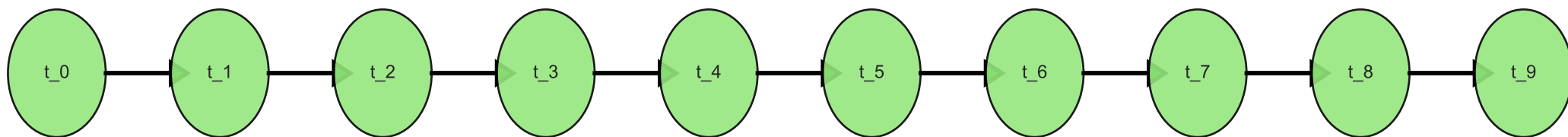- LSTM (factor≈1.0)
- Effective zero
- Vanishing zone
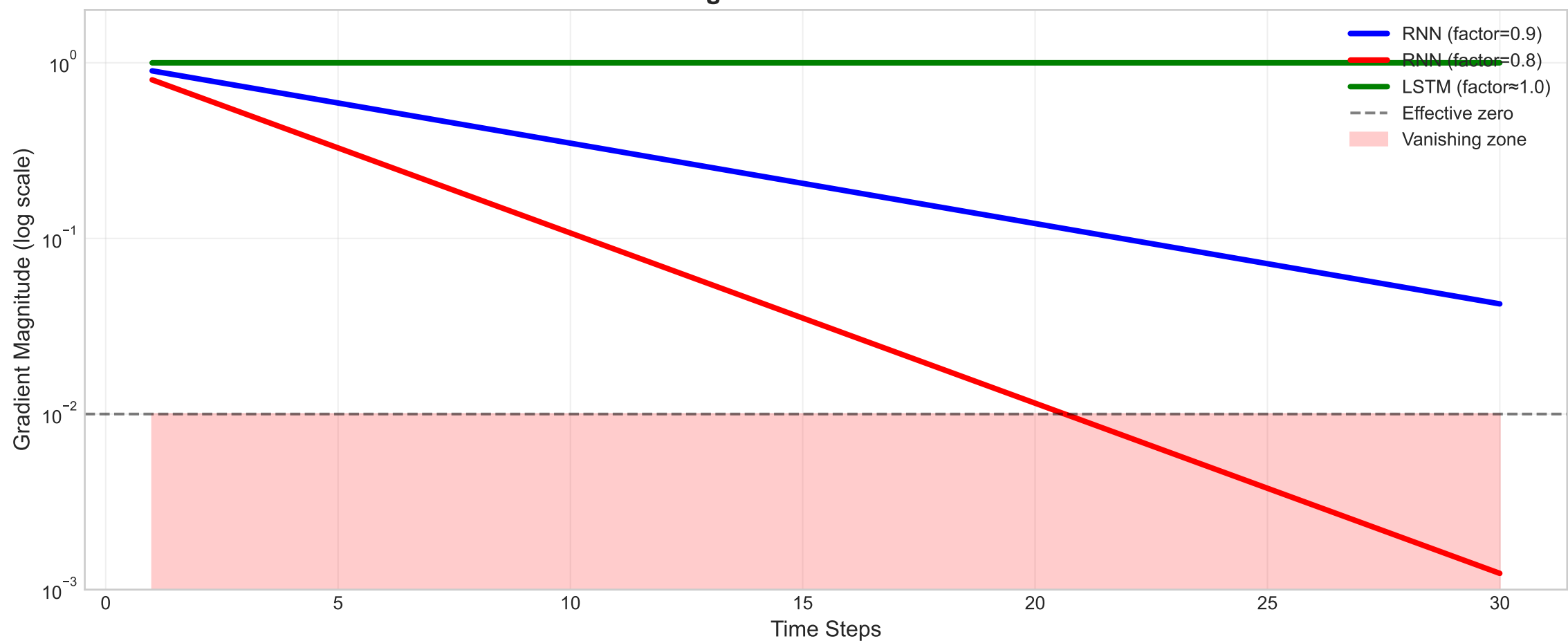
Gradient Flow Comparison: RNN vs LSTM

RNN: Gradient vanishes

t_0 → t_1 → t_2 → t_3 → t_4 → t_5 → t_6 → t_7 → t_8 → t_9
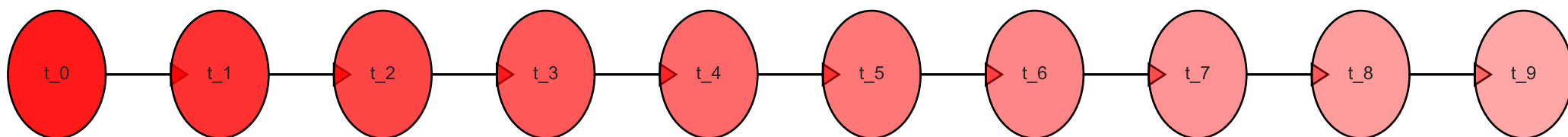
LSTM: Gradient preserved

t_0 → t_1 → t_2 → t_3 → t_4 → t_5 → t_6 → t_7 → t_8 → t_9

Gradient Highway