$$h_t = \sigma(Rh_{t-1} + Ax_t)$$

$$\frac{dh_t}{dh_{t-1}} = \sigma'(Rh_{t-1} + Ax_t) \cdot \frac{d}{dh_{t-1}}(Rh_{t-1} + Ax_t)$$

$$\boxed{\frac{dh_t}{dh_{t-1}} = \sigma'(Rh_{t-1} + Ax_t) \cdot R} \quad — (1)$$

$$y_t = Bh_t$$

$$\frac{dy_t}{dx_1} = B \times \frac{dh_t}{dx_1}$$

chain rule

$$\frac{dy_t}{dx_1} = B \times \frac{dh_t}{dh_{t-1}} \times \frac{dh_{t-1}}{dh_{t-2}} \times \frac{dh_{t-2}}{dh_{t-3}} \cdots \frac{dh_1}{dx_1}$$

$$\frac{dy_t}{dx_1} = B \times \left(\prod_{K=1}^{K=t-1} \sigma'(h_K)\right) \times R^{t-1} \times \frac{d}{dx_1}(Rh_0 + Ax_1)$$

from → (1)

$$\frac{dy_t}{dx_1} = B \times \left(\prod_{K=1}^{K=t-1} \sigma'(h_K)\right) \times R^{t-1} \times A$$

$$\therefore \left\|\frac{dy_t}{dx_1}\right\| <= \|B\| \times \left(\prod_{K=1}^{t} |\sigma'(Rh_{K-1} + Ax_K)|_\infty\right) \cdot \|R^{t-1}\| \cdot \|A\|$$

here $|\sigma'(h)|_\infty = \max(|\sigma'(h_1)| \ldots |\sigma'(h_d)|)$,

where $\sigma'(h) = 1 - \sigma^2(h)$ lets us bound the value of LHS to be less than or equal to the RHS. In the RHS the only variable is the product of hidden state matrices with tanh and getting the product of max of $|\sigma'(h)|_\infty$ at each time step allows for us to estimate the maximum change in y_t for a change in x_1 to be less than or equal to R.H.S.