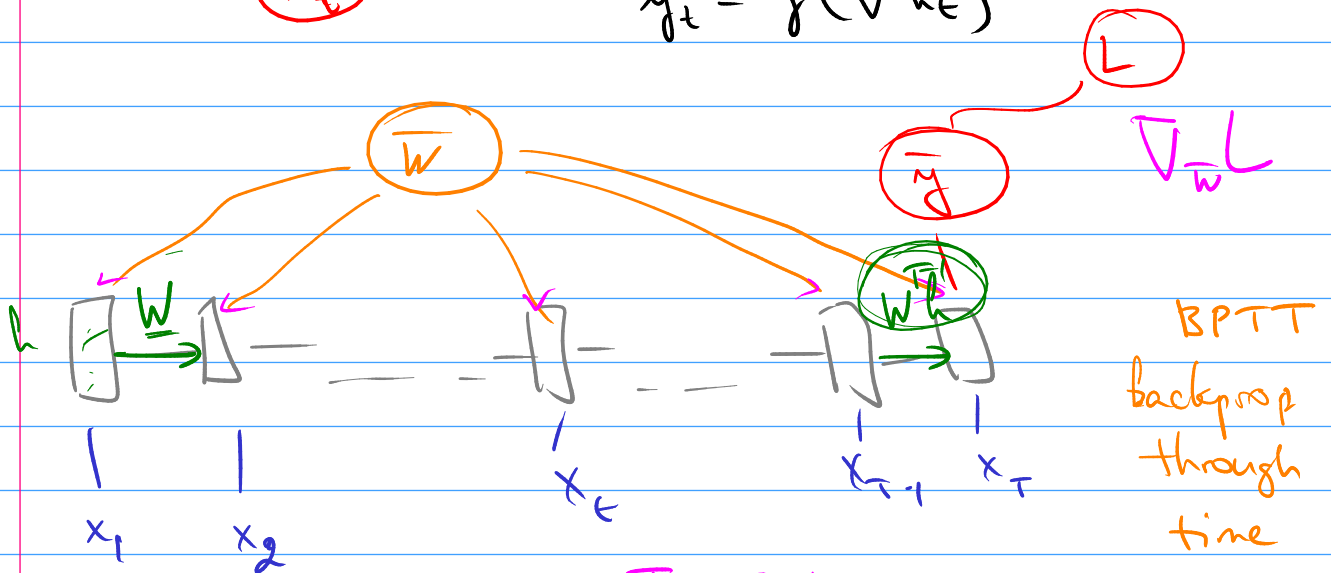


$$h_t = f(W h_{t-1} + U x_t)$$

$$y_t = g(V h_t)$$



$$\frac{\partial L}{\partial w_i} = \sum_{t=1}^T \frac{\partial L}{\partial w_i(t)}$$

$t = T - 10$ ← Truncated BPTT

Problems RNN:

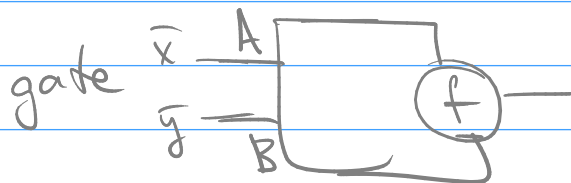
1) exploding gradients

$$\| \nabla_w L \| \leq c$$

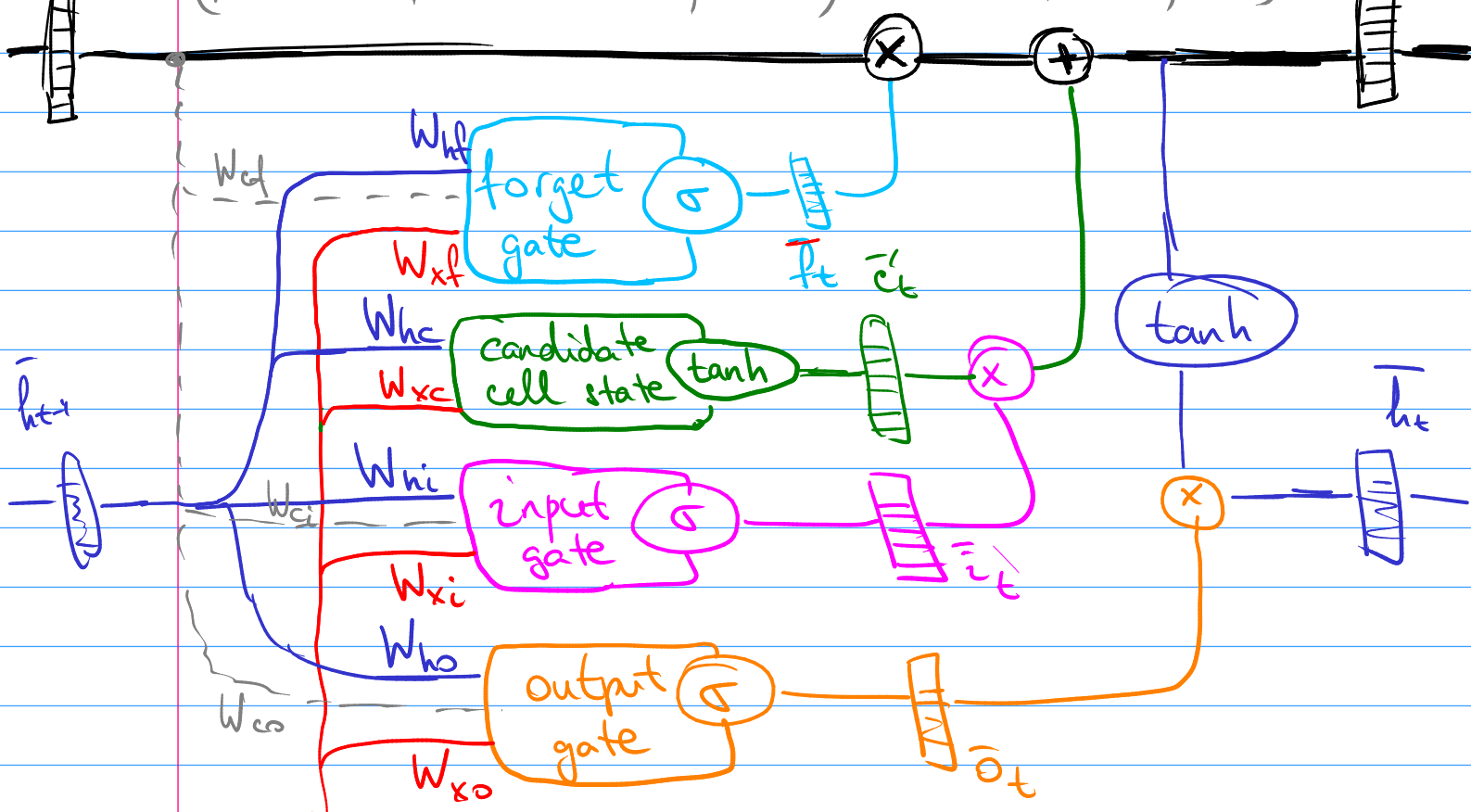
clipping

- clipnorm
- clip value

2) vanishing gradients



LSTM long short-term memory
 (Hochreiter, Schmidhuber, 1995) (Ges, Schim, 2000)



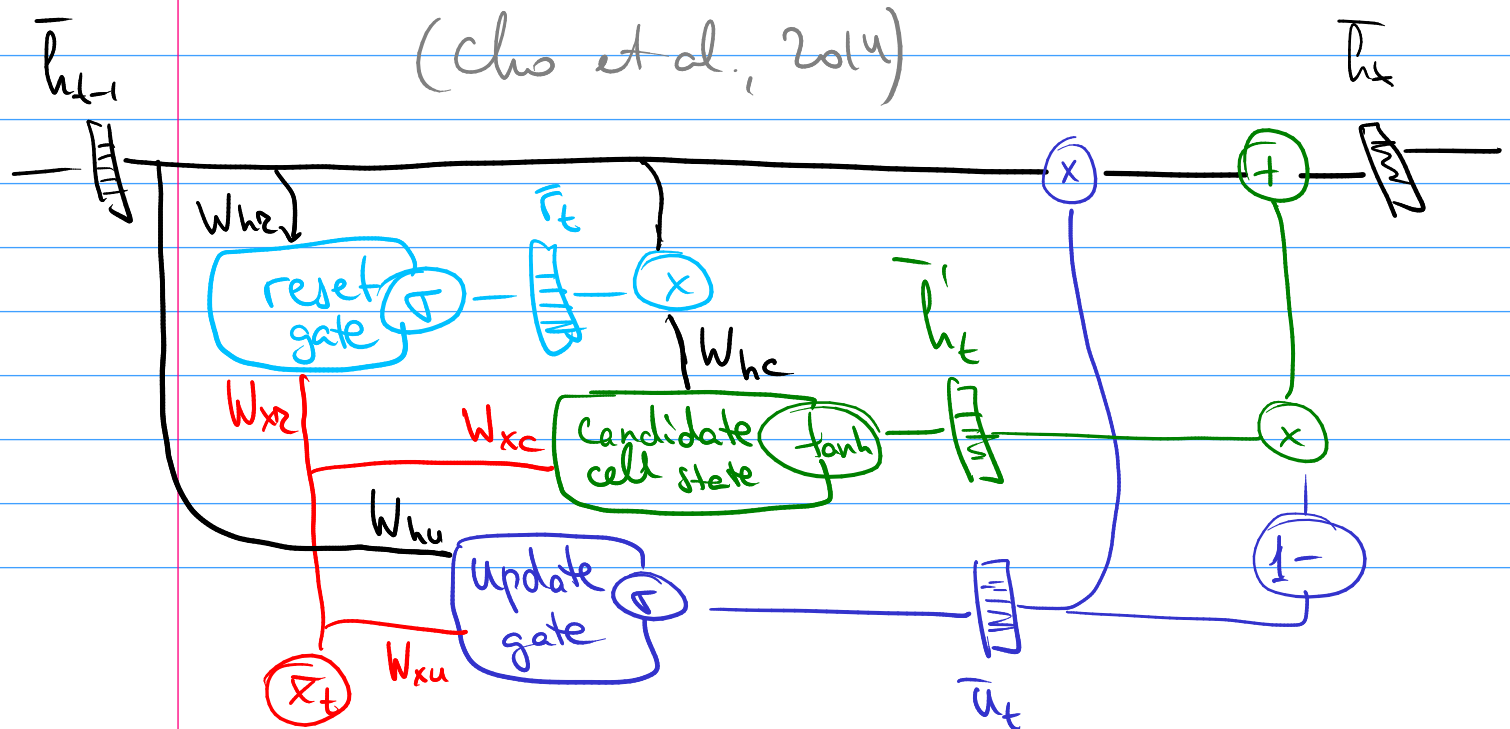
$$\bar{f}_t = \sigma(W_{hf}\bar{h}_{t-1} + W_{xf}\bar{x}_t + W_{fo})$$

$$\bar{c}_t = \tanh(\dots)$$

$$\bar{i}_t = \sigma(\dots)$$

$$\bar{o}_t = \sigma(\dots)$$

GRU (gated recurrent unit)
 (Cho et al., 2014)



Bidirectional RNN

