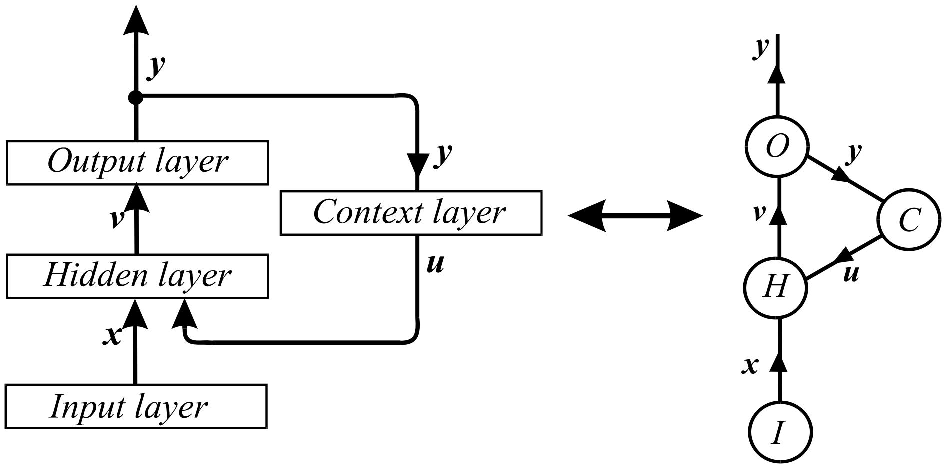
# Jordan's recurrent neural networks

*Vladimír Kvasnička*

Department of Mathematics
Slovak Technical University
812 37 Bratislava, Slovakia

March 1999

# Jordan's RNN



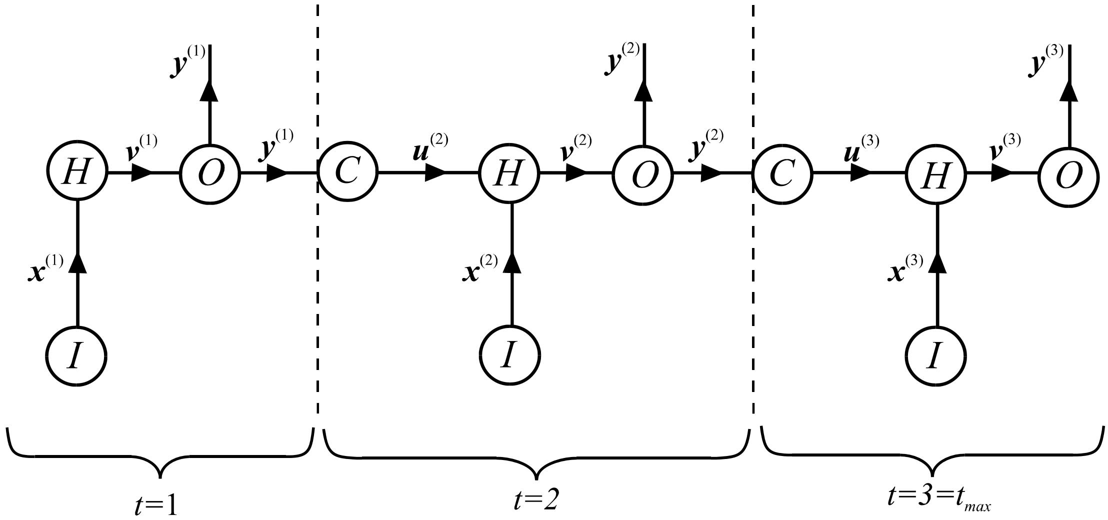

Its activities of single layers are determined as

$$\left.\begin{array}{l} \boldsymbol{x} = input \in \{a,b,c,d\}^{*} \\ \boldsymbol{v} = H(\boldsymbol{x},\boldsymbol{u}) \\ \boldsymbol{y} = O(\boldsymbol{v}) \\ \boldsymbol{u} = C(\boldsymbol{y}) \\ \boldsymbol{y} = O(\boldsymbol{v}) \in (0,1)^{n} \end{array}\right\} \Rightarrow \left\{\begin{array}{l} \boldsymbol{x} = input \\ v = H\big(\boldsymbol{x},C\big(O(v)\big)\big) \\ \boldsymbol{y} = O(\boldsymbol{v}) \in (0,1)^{n} \end{array}\right.$$

Iterative solution gives the activities specified in a recurrent form

$$\left.\begin{aligned}
\boldsymbol{x}^{(t)} &= \text{input} \\
\boldsymbol{u}^{(t)} &= \begin{cases} 0 & (t=1) \\ C\left(\boldsymbol{y}^{(t-1)}\right) & (t \geq 2) \end{cases} \\
\boldsymbol{v}^{(t)} &= H\left(\boldsymbol{x}^{(t)}, \boldsymbol{u}^{(t)}\right) \\
\boldsymbol{y}^{(t)} &= O\left(\boldsymbol{v}^{(t)}\right)
\end{aligned}\right\} \quad \text{for } t = 1,2,\ldots,t_{max}$$

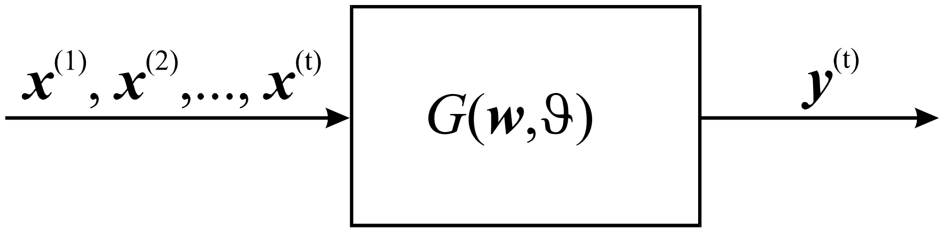
# Unfolded recurrent neural network



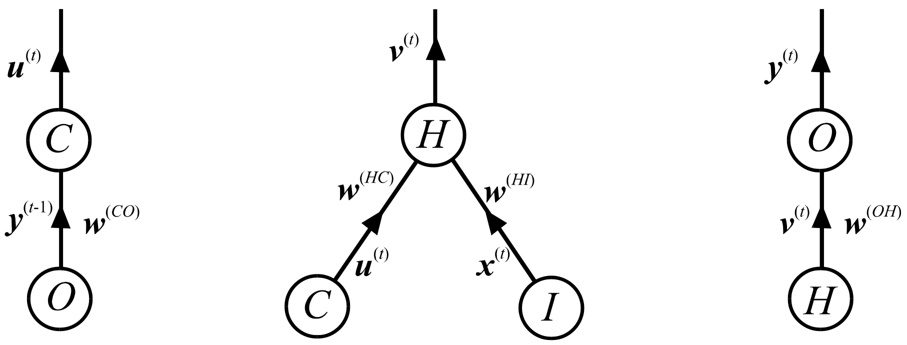

Unfolded Jordan's recurrent neural network may be considered as a parametric mapping that maps a sequence of input vectors onto an output vector

$$\boldsymbol{y}^{(t)} = G\left(\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}, \ldots, \boldsymbol{x}^{(t)}; \boldsymbol{w}, \vartheta\right)$$

for $t$=1,2,...,$t_{\max}$.



Activities are explicitly determined by




$$x_i^{(t)} = \text{input}$$

$$u_i^{(t)} = \begin{cases} 0 & (\text{for } t = 1) \\ t\left(\vartheta_i^{(C)} + \sum_{j=1}^{N_O} w_{ij}^{(CO)} y_j^{(t-1)}\right) & (\text{otherwise}) \end{cases}$$

$$v_i^{(t)} = t\left(\vartheta_i^{(H)} + \sum_{j=1}^{N_C} w_{ij}^{(HC)} u_j^{(t)} + \sum_{j=1}^{N_I} w_{ij}^{(HI)} x_j^{(t)}\right)$$

$$y_i^{(t)} = t\left(\vartheta_i^{(O)} + \sum_{j=1}^{N_H} w_{ij}^{(OH)} v_j^{(t)}\right)$$

## **Adaptation** (**learning**) of the recurrent neural network

$$A_{train} = \left\{ \left( \boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}, \ldots, \boldsymbol{x}^{(t_{max})} \right) \Big/ \left( \boldsymbol{y}_{req}^{(1)}, \boldsymbol{y}_{req}^{(2)}, \ldots, \boldsymbol{y}_{req}^{(t_{max})} \right) \right\}$$

$$E = \frac{1}{2} \sum_{t=1}^{t_{max}} \omega_t \left( \boldsymbol{y}^{(t)} - \boldsymbol{y}_{req}^{(t)} \right)^2$$

An adaptation is equivalent to a minimization of the objective function $E$ with respect to weight and threshold coefficients

$$\boxed{\left( w_{opt}, \vartheta_{opt} \right) = \arg \min_{(w,\vartheta)} E(w,\vartheta)}$$

This optimization problem is most frequently solved by the so-called gradient method of **steepest descent**

$$\boldsymbol{w}_{k+1} = \boldsymbol{w}_k - \lambda \, grad \; E(\boldsymbol{w}_k)$$

where $\lambda>0$.

In general, partial derivatives of the objective function $E$ are determined as follows

$$\left(\frac{\partial E}{\partial \vartheta_i}\right) = t'(\xi_i)\left(g_i + \sum_k \frac{\partial E}{\partial \vartheta_k} w_{ki}\right)$$

$$t'(\xi_i) = t(\xi_i)\left[1 - t(\xi_i)\right]$$

$$g_i = \begin{cases} x_i - x_{i,req} & (i \in O) \\ 0 & (i \notin O) \end{cases}$$

$$\frac{\partial E}{\partial w_{ij}} = \frac{\partial E}{\partial \vartheta_i} x_j$$

These general formulae (they form a background of the so-called **back-propagation approach**) are immediately applicable to unfolded recurrent neural networks for calculation of partial derivatives of the objective function $E$.

## (1) Initialization, $t$=$t_{\max}$

$$\left(\frac{\partial E}{\partial \vartheta_i^{(O)}}\right)^{(t_{max})} = y_i^{(t_{max})}\left(1-y_i^{(t_{max})}\right)\omega_{t_{max}}\left(y_i^{(t_{max})}-y_{i,req}^{(t_{max})}\right)$$

$$\left(\frac{\partial E}{\partial \vartheta_i^{(H)}}\right)^{(t_{max})} = v_i^{(t_{max})}\left(1-v_i^{(t_{max})}\right)\sum_{j=1}^{N_O}\left(\frac{\partial E}{\partial \vartheta_j^{(O)}}\right)^{(t_{max})} w_{ji}^{(OH)}$$

$$\left(\frac{\partial E}{\partial \vartheta_i^{(C)}}\right)^{(t_{max})} = u_i^{(t_{max})}\left(1-u_i^{(t_{max})}\right)\sum_{j=1}^{N_H}\left(\frac{\partial E}{\partial \vartheta_j^{(H)}}\right)^{(t_{max})} w_{ji}^{(HC)}$$

(2) Iteration, $1 \le t < t_{\max}$

$$\left(\frac{\partial E}{\partial \vartheta_i^{(O)}}\right)^{(t)} = y_i^{(t)}\left(1-y_i^{(t)}\right)\begin{pmatrix}\omega_t\left(y_i^{(t)}-y_{i,req}^{(t)}\right)+ \\ \sum\limits_{j=1}^{N_C}\left(\frac{\partial E}{\partial \vartheta_k^{(C)}}\right)^{(t+1)} w_{ji}^{(CO)}\end{pmatrix}$$

$$\left(\frac{\partial E}{\partial \vartheta_i^{(H)}}\right)^{(t)} = v_i^{(t)}\left(1-v_i^{(t)}\right)\sum_{j=1}^{N_O}\left(\frac{\partial E}{\partial \vartheta_j^{(O)}}\right)^{(t)} w_{ji}^{(OH)}$$

$$\left(\frac{\partial E}{\partial \vartheta_i^{(C)}}\right)^{(t)} = u_i^{(t)}\left(1-u_i^{(t)}\right)\sum_{j=1}^{N_H}\left(\frac{\partial E}{\partial \vartheta_j^{(H)}}\right)^{(t)} w_{ji}^{(HC)}$$

Partial derivatives with respect to weight coefficients are

$$\left(\frac{\partial E}{\partial w_{ij}^{(OH)}}\right)^{(t)} = \left(\frac{\partial E}{\partial \vartheta_i^{(O)}}\right)^{(t)} v_j^{(t)} \quad (1 \le t \le t_{max})$$

$$\left(\frac{\partial E}{\partial w_{ij}^{(HC)}}\right)^{(t)} = \left(\frac{\partial E}{\partial \vartheta_i^{(H)}}\right)^{(t)} u_j^{(t)} \quad (2 \le t \le t_{max})$$

$$\left(\frac{\partial E}{\partial w_{ij}^{(HI)}}\right)^{(t)} = \left(\frac{\partial E}{\partial \vartheta_i^{(H)}}\right)^{(t)} x_j^{(t)} \quad (1 \le t \le t_{max})$$

$$\left(\frac{\partial E}{\partial w_{ij}^{(CO)}}\right)^{(t)} = \left(\frac{\partial E}{\partial \vartheta_i^{(C)}}\right)^{(t)} y_j^{(t-1)} \quad (2 \le t \le t_{max})$$

Total partial derivatives are determined as follows

$$\frac{\partial E}{\partial \vartheta_i^{(O)}}=\sum_{t=1}^{t_{max}}\left(\frac{\partial E}{\partial \vartheta_i^{(O)}}\right)^{(t)}, \quad \frac{\partial E}{\partial \vartheta_i^{(H)}}=\sum_{t=1}^{t_{max}}\left(\frac{\partial E}{\partial \vartheta_i^{(H)}}\right)^{(t)}$$

$$\frac{\partial E}{\partial \vartheta_i^{(C)}}=\sum_{t=2}^{t_{max}}\left(\frac{\partial E}{\partial \vartheta_i^{(C)}}\right)^{(t)}$$

$$\frac{\partial E}{\partial w_{ij}^{(OH)}}=\sum_{t=1}^{t_{max}}\left(\frac{\partial E}{\partial w_{ij}^{(OH)}}\right)^{(t)}=\sum_{t=1}^{t_{max}}\left(\frac{\partial E}{\partial \vartheta_i^{(O)}}\right)^{(t)} v_j^{(t)}$$

$$\frac{\partial E}{\partial w_{ij}^{(HC)}}=\sum_{t=2}^{t_{max}}\left(\frac{\partial E}{\partial w_{ij}^{(HC)}}\right)^{(t)}=\sum_{t=2}^{t_{max}}\left(\frac{\partial E}{\partial \vartheta_i^{(H)}}\right)^{(t)} u_j^{(t)}$$

$$\frac{\partial E}{\partial w_{ij}^{(HI)}}=\sum_{t=1}^{t_{max}}\left(\frac{\partial E}{\partial w_{ij}^{(HI)}}\right)^{(t)}=\sum_{t=1}^{t_{max}}\left(\frac{\partial E}{\partial \vartheta_i^{(H)}}\right)^{(t)} x_j^{(t)}$$

$$\frac{\partial E}{\partial w_{ij}^{(CO)}}=\sum_{t=2}^{t_{max}}\left(\frac{\partial E}{\partial w_{ij}^{(CO)}}\right)^{(t)}=\sum_{t=2}^{t_{max}}\left(\frac{\partial E}{\partial \vartheta_i^{(C)}}\right)^{(t)} y_j^{(t-1)}$$