

Επεξεργασία Φωνής & Φυσικής Γλώσσας - ΡΟΗ Σ 7<sup>ο</sup> Εγαύνο

2<sup>η</sup> Σειρά Αναλυτικών Ασκήσεων - ΑΚ. Έτος 2020-2021

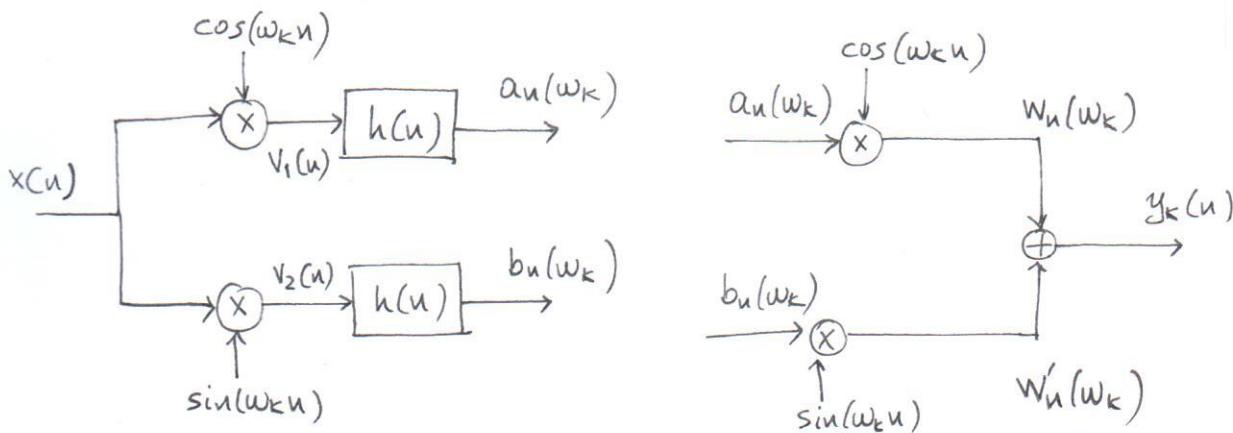
Ονοματεπώνυμο: Χρήστος Δημόπουλος

Αρ. Μητρώου: 031 17 037

E-mail: chrisdim1999@gmail.com

Άσκηση 1:

Θεωρούμε την ανάλυση και σύνθεση του σηματού  $x(n) = \cos(\omega_0 n)$ . Το δίκτυο για το k-οςτό κανάλι φαίνεται παρακάτω:



) Τα σηματα  $v_1(n)$  και  $v_2(n)$  ορίζονται ως:

$$v_1(n) = x(n) \cdot \cos(\omega_k n) = \cos(\omega_0 n) \cos(\omega_k n)$$

$$v_2(n) = x(n) \cdot \sin(\omega_k n) = \cos(\omega_0 n) \sin(\omega_k n)$$

Χρησιμοποιώντας τις τριγωνομετρικές ταυτότητες:

$$2 \cos A \cos B = \cos(A+B) + \cos(A-B)$$

$$2 \cos A \sin B = \sin(A+B) - \sin(A-B)$$

λαμβάνουμε:

$$V_1(n) = \frac{1}{2} \left\{ \cos([\omega_0 + \omega_k]n) + \cos([\omega_0 - \omega_k]n) \right\}$$

$$V_2(n) = \frac{1}{2} \left\{ \sin([\omega_0 + \omega_k]n) - \sin([\omega_0 - \omega_k]n) \right\}$$

Επειτα, δεωρίζοντας τα συστήματα κρονικής απόκρισης  $h(n)$  γραφικά & χρησικά αναλογιώτα, λαμβάνουμε:

$$a_n(\omega_k) = V_1(n) * h(n) = \sum_{m=-\infty}^{+\infty} x(m) \cdot \cos(\omega_k m) \cdot h(n-m)$$

$$b_n(\omega_k) = V_2(n) * h(n) = \sum_{m=-\infty}^{+\infty} x(m) \cdot \sin(\omega_k m) \cdot h(n-m)$$

Μεταβούντας στο πέδιο των συχνότητων μέσω M. Fourier:

$$V_1(\omega) = \frac{\pi}{2} \left\{ \delta(\omega + \omega_0 + \omega_k) + \delta(\omega - \omega_0 - \omega_k) + \delta(\omega + \omega_0 - \omega_k) + \delta(\omega - \omega_0 + \omega_k) \right\}$$

$$V_2(\omega) = \frac{\pi j}{2} \left\{ \delta(\omega + \omega_0 + \omega_k) - \delta(\omega - \omega_0 - \omega_k) + \delta(\omega + \omega_0 - \omega_k) + \delta(\omega - \omega_0 + \omega_k) \right\}$$

και:

$$\mathcal{F}\{a_n(\omega_k)\} = V_1(\omega) \cdot H(\omega) = A_{\omega_k}(\omega)$$

$$\mathcal{F}\{b_n(\omega_k)\} = V_2(\omega) \cdot H(\omega) = B_{\omega_k}(\omega)$$

όπου  $H(\omega)$  η συναρτηση μεταφοράς των συστημάτων κρονικής απόκρισης  $h(n)$ .

2) Υποδέχοντας ου το λη στην είναι narrowband lowpass φίλτρο και  
 Δεωρίνοντας ου το  $(\omega_0 - \omega_K)$  πέφτει στην μέση μάκτα διέλευσης του φίλτρου, καθώς  
 και ου  $H(e^{j\omega}) \approx 1$  για τέτοιες συχνότητες, μπορούμε να αντανακτήσουμε  
 περαιτέρω τις εκφράσεις για τα  $a_n(\omega_K), b_n(\omega_K)$ .

Ειδικότερα, το βαθυπερατό φίλτρο εξαλείφει τις υψησυχες συνιστώσες  
 που εντοπίζονται στις δέσεις  $\pm (\omega_0 + \omega_K)$  και εν τέλει λαμβάνουμε:

$$\mathcal{F}\{a_n(\omega_K)\} = \frac{\pi}{2} \left\{ \delta(\omega + \omega_0 - \omega_K) + \delta(\omega - \omega_0 + \omega_K) \right\}$$

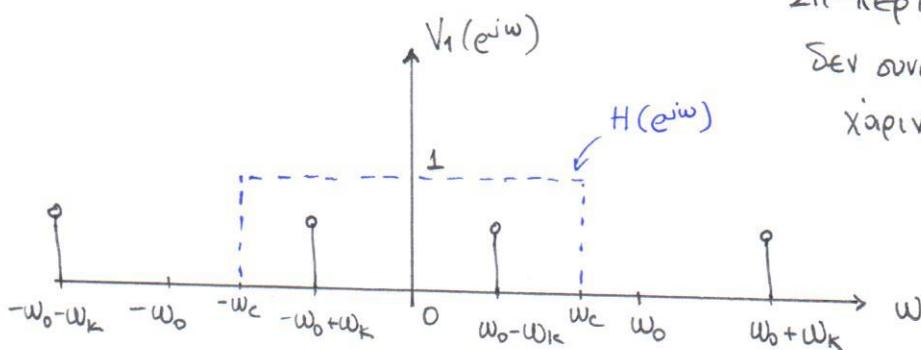
$\uparrow \downarrow \mathcal{F}^{-1}$

$$a_n(\omega_K) = \frac{1}{2} \cos[(\omega_0 - \omega_K)n]$$

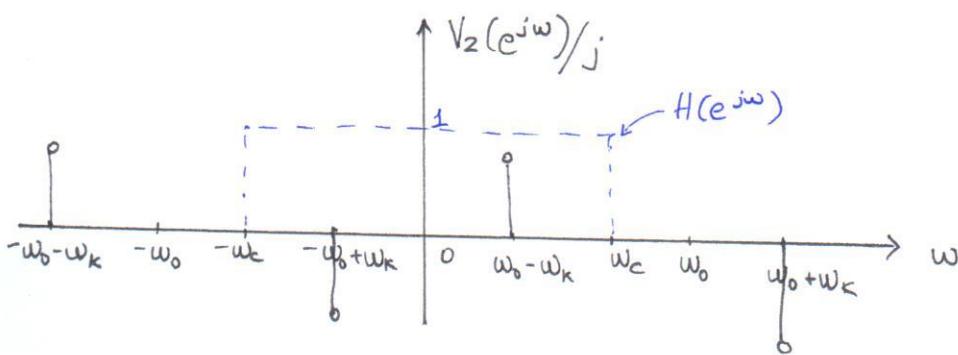
$$\mathcal{F}\{b_n(\omega_K)\} = j\frac{\pi}{2} \left\{ -\delta(\omega + \omega_0 - \omega_K) + \delta(\omega - \omega_0 + \omega_K) \right\}$$

$\uparrow \downarrow \mathcal{F}^{-1}$

$$b_n(\omega_K) = -\frac{1}{2} \sin[(\omega_0 - \omega_K)n]$$



Γνωρίζοντας ου ο DTFT είναι  
 2\pi-περιοδικός, υποδέτουμε ου  
 δεν συναρτάκει φαινόμενα aliasing  
 χάριν ευκολίας.



3) · Magnitude:

$$M_n(\omega_k) = \sqrt{a_n^2(\omega_k) + b_n^2(\omega_k)} = \sqrt{\frac{1}{4} \left[ \cos^2[(\omega_0 + \omega_k)n] + \sin^2[(\omega_0 - \omega_k)n] \right]}$$

$$\Rightarrow M_n(\omega_k) = 1/2$$

· Phase Derivative:

Av θεωρίσουμε οι εργαζόμενοι στον ουρεχή χρόνο, τότε η προώηξης της φάσης έγινε:

$$\dot{\varphi}(-\Omega_k) = \frac{d}{dt} \left[ -\arctan \left( \frac{b(t, -\Omega_k)}{a(t, -\Omega_k)} \right) \right] = \frac{d}{dt} \left[ -\arctan \left( \frac{-\sin(-\Omega_0 - \Omega_k)t}{\cos(-\Omega_0 - \Omega_k)t} \right) \right] =$$

$$= \frac{d}{dt} \left[ -\arctan \left( -\tan[(\Omega_0 - \Omega_k)t] \right) \right] \quad \text{όμως } \tan(-x) = -\tan x$$

$$= \frac{d}{dt} \left[ (\Omega_k - \Omega_0) \cdot t \right] = \Omega_0 - \Omega_k$$

Διεγματοληπτώντας τα ουρεχή σημάτα, θέτε να μεταβούμε στον διακριτό χρόνο, λαμβάνουμε:

$$\dot{\varphi}_n(\omega_k) = \omega_0 - \omega_k$$

$$4) w_n(\omega_k) = a_n(\omega_k) \cdot \cos(\omega_k n) = \frac{1}{2} \cos[(\omega_0 - \omega_k)n] \cos(\omega_k n)$$

$$w'_n(\omega_k) = b_n(\omega_k) \cdot \sin(\omega_k n) = -\frac{1}{2} \sin[(\omega_0 - \omega_k)n] \sin(\omega_k n)$$

Χρησιμοποιώντας τις τριγωνομετρικές ταυτότητες:

$$\cos A \cdot \cos B = \frac{1}{2} [\cos(A+B) + \cos(A-B)]$$

$$\sin A \cdot \sin B = \frac{1}{2} [-\cos(A+B) + \cos(A-B)]$$

λακβάρουψ.

$$w_n(\omega_k) = \frac{1}{4} \left\{ \cos(\omega_0 n) + \cos[(\omega_0 - 2\omega_k)n] \right\}$$

$$w'_n(\omega_k) = \frac{1}{4} \left\{ \cancel{\sin} \cos(\omega_0 n) - \cos[(\omega_0 - 2\omega_k)n] \right\}$$

Επομένως:

$$y_k(n) = w_n(\omega_k) + w'_n(\omega_k) = \frac{1}{4} \left\{ 2 \cos(\omega_0 n) \right\}$$

$$\Rightarrow \boxed{y_k(n) = \frac{1}{2} \cos(\omega_0 n) = \frac{1}{2} x(n)}$$

Διλαδί, το σύμα εξόδου  $y_k(n)$  είναι το σύμα εισόδου  $x(n)$  με υποδιπλασιαρχέο πήλατος.

5) Η παραγωγος φάσης  $\varphi'_n(\omega_k)$  υπολογίζεται από τον τύπο:

$$\varphi'_n(\omega_k) = \frac{b_n(\omega_k) a'_n(\omega_k) - a_n(\omega_k) b'_n(\omega_k)}{\left[a_n(\omega_k)\right]^2 + \left[b_n(\omega_k)\right]^2}$$

Δεωρινής μοναδιώς ρυθμός δεγχατοληψίας ( $T=1$ ), ώστε να μπορούμε να παραγγίσουμε στον συνεχή χρόνο, λαμβάνουμε:

$$a(t, \Omega_k) = \frac{1}{2} \cos[(\Omega_0 - \Omega_k)t] \xrightarrow{\frac{d(\cdot)}{dt}} a'(t, \Omega_k) = -\frac{1}{2} (\Omega_0 - \Omega_k) \sin[(\Omega_0 - \Omega_k)t]$$

$$b(t, \Omega_k) = -\frac{1}{2} \sin[(\Omega_0 - \Omega_k)t] \xrightarrow{\frac{d(\cdot)}{dt}} b'(t, \Omega_k) = -\frac{1}{2} (\Omega_0 - \Omega_k) \cos[(\Omega_0 - \Omega_k)t]$$

Δειγματοληπτώντας τις παραπάνω ποσότητες, τελικά παίρνουμε:

$$\varphi'_n(\omega_k) = \frac{\left(-\frac{1}{2}\right) \sin[(\omega_0 - \omega_k)t] \cdot \left(-\frac{1}{2}\right) \cdot (\omega_0 - \omega_k) \sin[(\omega_0 - \omega_k)t] - \frac{1}{2} \cos^2[(\omega_0 - \omega_k)t] \left(-\frac{1}{2}\right) (\omega_0 - \omega_k)}{\left\{\frac{1}{2} \cos[(\omega_0 - \omega_k)t]\right\}^2 + \left\{-\frac{1}{2} \sin[(\omega_0 - \omega_k)t]\right\}^2}$$

$$\Rightarrow \varphi'_n(\omega_k) = \frac{\cancel{\frac{1}{4}} (\omega_0 - \omega_k) \cdot [\sin^2(\omega_0 - \omega_k) + \cos^2(\omega_0 - \omega_k)]}{\cancel{(1/4)} \cdot [\cos^2(\omega_0 - \omega_k) + \sin^2(\omega_0 - \omega_k)]} \xrightarrow{1}$$

$$\Rightarrow \boxed{\varphi'_n(\omega_k) = \omega_0 - \omega_k}$$

Διλαδί το ίδιο ακριβώς αντιτείχει με αυτό του ερωτήματος (3).

6) Υποδέτουμε ότι οι παράγωγοι του προηγούμενου ερώτηματος υπολογίζονται ως εξής:

$$a'(w_k) \approx \frac{1}{T} (a_n(w_k) - a_{n-1}(w_k))$$

T: περίοδος δειγματοληψίας.

Επομένως:

$$\varphi'_n(w_k) = \frac{1}{T} \cdot \frac{b_n(w_k)a_n(w_k) - b_{n-1}(w_k)a_{n-1}(w_k) - a_n(w_k)b_n(w_k) + a_{n-1}(w_k)b_{n-1}(w_k)}{[a_n(w_k)]^2 + [b_n(w_k)]^2}$$

$$\Rightarrow \varphi'_n(w_k) = \frac{1}{T} \cdot \frac{a_n(w_k)b_{n-1}(w_k) - b_n(w_k)a_{n-1}(w_k)}{[a_n(w_k)]^2 + [b_n(w_k)]^2} \quad (\text{I})$$

όπου:

$$a_{n-1}(w_k) = \frac{1}{2} \cos[(\omega_0 - \omega_k)(n-1)] = \frac{1}{2} \left\{ \cos[(\omega_0 - \omega_k)_n] \cos(\omega_0 - \omega_k) + \sin[(\omega_0 - \omega_k)_n] \sin(\omega_0 - \omega_k) \right\}$$

$$b_{n-1}(w_k) = -\frac{1}{2} \sin[(\omega_0 - \omega_k)(n-1)] = -\frac{1}{2} \left\{ \sin[(\omega_0 - \omega_k)_n] \cos(\omega_0 - \omega_k) - \sin(\omega_0 - \omega_k) \cos[(\omega_0 - \omega_k)_n] \right\}$$

Αντικαθιστώντας στη σχέση (I):

$$\varphi'_n(w_k) = \frac{1}{T} \frac{(1/4)}{\cancel{(1/4)}} \left\{ -\sin[(\omega_0 - \omega_k)_n] \cos[(\omega_0 - \omega_k)_n] \cos(\omega_0 - \omega_k) + \sin(\omega_0 - \omega_k) \cdot \cos^2[(\omega_0 - \omega_k)_n] \right.$$

$$\left. + \sin[(\omega_0 - \omega_k)_n] \cos[(\omega_0 - \omega_k)_n] \cos(\omega_0 - \omega_k) + \sin^2[(\omega_0 - \omega_k)_n] \sin(\omega_0 - \omega_k) \right\}$$

$$\Rightarrow \varphi'_n(w_k) = \frac{1}{T} \cdot \sin(\omega_0 - \omega_k) \cdot \left[ \sin^2[(\omega_0 - \omega_k)_n] + \cos^2[(\omega_0 - \omega_k)_n] \right]^{\cancel{1}}$$

$$\Rightarrow \varphi_n'(\omega_k) = \frac{1}{T} \sin(\omega_0 - \omega_k)$$

Γνωρίζοντας ου  $\lim_{x \rightarrow 0} \sin x = x$ , μπορούμε να συμπεράνουμε ου η παρόχυγος φάσης είναι περίου ίδια με  $(\omega_0 - \omega_k)$ , δηλαδί το αντίτοπο του προηγουμένου ερωτήματος, εφόσον ισχύουν τα ακόλουθα:

- (i) Η περίοδος δειγματοληψίας είναι ποντιαία, δηλ.  $T=1$ .
- (ii)  $|\omega_0 - \omega_k| \rightarrow 0$ , δηλαδί οι συχνότητες  $\omega_0$  και  $\omega_k$  είναι πολύ κοντες μεταξύ τους.

## Άσκηση 2:

Δευτορίκες δύο πεπερασμένα σήματα σειρών  $x(n)$  και  $y(n)$ ,  $0 \leq n \leq N-1$  (με μιδενικές τιμές εκτός του παραδίποντος ανάλυσης). Για LPC ανάλυση με την Autocorrelation function μέθοδο χρειάζονται οι αυτοσυνχρέσεις:

$$R_x(k) = \sum_{n=0}^{N-1-k} x(n)x(n+k) \quad \text{και} \quad R_y(k) = \sum_{n=0}^{N-1-k} y(n)y(n+k)$$

Οι ονομές με τη μέθοδο Levinson-Durbin δίνουν τους ανιστοχούς βελτιστούς LPC συντετονίες:

$$\alpha_x = (\alpha_{x0}, \alpha_{x1}, \dots, \alpha_{xp}), \quad \alpha_y = (\alpha_{y0}, \alpha_{y1}, \dots, \alpha_{yp})$$

$$\mu \in \alpha_{x0} = \alpha_{y0} = -1.$$

1) Ενέργεια λιανού προβλεψης για το  $x(n)$ :

$$\begin{aligned} E_x &= \sum_{n=0}^{N-1+p} (e[n]^2) = \sum_{n=0}^{N-1+p} \left( \sum_{k=0}^p \alpha_{xk} x(n-k) \right)^2 \\ &= \sum_{n=0}^{N-1+p} \left( \sum_{k=0}^p \alpha_{xk} x(n-k) \right) \left( \sum_{k'=0}^p \alpha_{xk'} x(n-k') \right) \\ &= \sum_{k=0}^p \alpha_{xk} \sum_{k'=0}^p \alpha_{xk'} \sum_{n=0}^{N-1+p} x(n-k) x(n-k') \end{aligned}$$

Όμως:

$$\sum_{n=0}^{N-1+p} x(n-k) x(n-k') = \sum_{n=0}^{N-1+p} x(n) x(n-k'+k) = R_x(k-k')$$

Επομένως:

$$E_x = \sum_{k=0}^p a_{xk} \sum_{k'=0}^p a_{xk'} R_x(k-k') \Rightarrow E_x = a_x R a_x^\top$$

ποτου:

$$R = \begin{bmatrix} R_x(0) & R_x(1) & \dots & R_x(p) \\ R_x(1) & R_x(0) & & R_x(p-1) \\ \dots & \dots & \dots & \dots \\ R_x(p) & R_x(p-1) & \dots & R_x(0) \end{bmatrix}$$

μητρώο  $(p+1) \times (p+1)$

2) Αν κάνουμε γραμμική πρόβλεψη του σήματος  $x(n)$  με τους βέλτιστους συντελεστές του σήματος  $y(n)$ , λαμβάνουμε την ενέργεια του νέου υποδιδικού λαθούς πρόβλεψης:

$$\begin{aligned} E_{xy} &= \sum_{n=0}^{N-1+p} \left( \sum_{k=0}^p a_{yk} x(n-k) \right)^2 = \sum_{n=0}^{N-1+p} \left( \sum_{k=0}^p a_{yk} x(n-k) \right) \left( \sum_{k'=0}^p a_{yk'} x(n-k') \right) \\ &= \sum_{k=0}^p a_{yk} \sum_{k'=0}^p a_{yk'} \sum_{n=0}^{N-1+p} (x(n-k))(x(n-k')) = \sum_{k=0}^p a_{yk} \sum_{k'=0}^p a_{yk'} R_x(k-k') \end{aligned}$$

$$\Rightarrow E_{xy} = a_y R a_y^\top$$

ποτου:

$R$  το  $(p+1) \times (p+1)$  μητρώο του προηγουμένων ερωτημάτων.

3) Γνωρίζουμε ότι για το σύμβολο  $x(n)$  οι βέττιστοι  $LPC$  συντελεστές είναι ότι  $\alpha_x = (\alpha_{x0}, \alpha_{x1}, \dots, \alpha_{xp})$ . Με αυτά λόγια, η ενέργεια λίθου πρόβλεψης  $E_x$  είναι η ελάχιστη δυνατή για γραμμική πρόβλεψη με τους συντελεστές  $\alpha_x$ . Οποιοδήποτε άλλος συντελεστής γραμμικής πρόβλεψης για το σύμβολο  $x(n)$  δίνουν ενέργεια σφάλματος μεγαλύτερη - ή στην καλύτερη περίπτωση ίση - με αυτήν την δίνουν οι συντελεστές  $\alpha_x$ .

Οις εκ τούτου, αν κάνουμε γραμμική πρόβλεψη του σήματος  $x(n)$  χρησιμοποιώντας, ίσως, τους συντελεστές του σήματος  $y(n)$ , τότε η υβριδική ενέργεια σφάλματος πρόβλεψης  $E_{xy}$  θα είναι σίγουρα μεγαλύτερη ή ίση της ενέργειας  $E_x$ .

Με αυτά λόγια:

$$E_{xy} \geq E_x \xrightarrow{Ex \neq 0} \boxed{\frac{E_{xy}}{E_x} \geq 1}$$

### Άσκηση 3:

Σεωρούμε σε μια ακολουθία φωνητών τη ποικιλότηταν της εναλλαγής άφωνων ( $U = \text{unvoiced}$ ) και φωνητών ( $V = \text{voiced}$ ) μέχρι να έχει η HMM ποικιλότηταν ( $\lambda$ ) 4 καταστάσεων με τις εξής πιθανότητες:

	State 1	State 2	State 3	State 4
P(V)	0.5	0.8	0.25	0.2
P(U)	0.5	0.2	0.75	0.8

Πιθανότητες Μεταβάσεων Καταστάσεων:

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{bmatrix} = \begin{bmatrix} 0.25 & 0.2 & 0.3 & 0.25 \\ 0.2 & 0.25 & 0.3 & 0.25 \\ 0.4 & 0.2 & 0.2 & 0.2 \\ 0.25 & 0.3 & 0.2 & 0.25 \end{bmatrix}$$

Πιθανότητες αρχικής καταστάσεων:  $\pi_i = 0.25, i=1,2,3,4$

Παρατηρούμε την ακολουθία:

$$O = (UVUVVUVUVU)$$

$$S_t(i) = \max_{q_1, q_2, \dots, q_{t-1}} P[q_1, q_2, \dots, q_{t-1}, q_t=i, o_1, o_2, \dots, o_t | \lambda], \quad i=1,2,3,4, \quad t=1, \dots, 10$$

$$= \max_{q_1, \dots, q_{t-1}} \{ P(q_1) P(q_2 | q_1) \dots P(q_t=i | q_{t-1}) P(o_1 | q_1) P(o_2 | q_2) \dots P(o_t | q_t) \}$$

$$= \max_{q_1, \dots, q_{t-1}} \{ \pi_1 a_{12} a_{23} \dots a_{t-1+i} b_{q_1}(o_1) b_{q_2}(o_2) \dots b_{q_t}(o_t) \}$$

$$= b_{q_t}(o_t) \cdot \max_{q_{t-1}} \{ b_{q_{t-1}}(o_{t-1}) \cdot P(q_t=i | q_{t-1}) \cdot \max_{q_{t-2}} \{ \dots \max_{q_1} \{ b_{q_1}(o_1) a_{12} \pi_1 \} \} \}$$

Ακολουθούμε τον αλγόριθμο Viterbi :

(i) Αρχικοποίηση:  $\delta_1(i) = \pi_i \cdot b_i(O_1 = v)$

$$\delta_1(1) = 0.125, \delta_1(2) = 0.05, \delta_1(3) = 0.1875, \delta_1(4) = 0.2$$

(ii) Επαναπτυξική διαδικασία:  $\delta_{t+1}(j) = (\max_i \delta_t(i) \alpha_{ij}) \cdot b_j(O_{t+1})$

· Για  $t=2$ : ( $O_2 = v$ )

$$\delta_2(1) = \frac{0.0375}{\cancel{0.05}}$$

$$\delta_2(2) = 0.048$$

$$s_2(1) = 3$$

$$s_2(2) = 4$$

$$\delta_2(3) = 0.01$$

$$s_2(3) = 4$$

$$\delta_2(4) = 0.01$$

$$s_2(4) = 4$$

· Για  $t=3$ : ( $O_3 = v$ )

$$\delta_3(1) = 4 \cdot 8 \cdot 10^{-3}$$

$$\delta_3(2) = \frac{2,4 \cdot 10^{-3}}{\cancel{0.024}}$$

$$s_3(1) = 2$$

$$s_3(2) = 2$$

$$\delta_3(3) = 10,8 \cdot 10^{-3}$$

$$s_3(3) = 2$$

$$\delta_3(4) = 9,6 \cdot 10^{-3}$$

$$s_3(4) = 2$$

· Για  $t=4$ : ( $O_4 = v$ )

$$\delta_4(1) = 2,16 \cdot 10^{-3}$$

$$\delta_4(2) = 2,304 \cdot 10^{-3}$$

$$s_4(1) = 3$$

$$s_4(2) = 4$$

$$\delta_4(3) = 0.54 \cdot 10^{-3}$$

$$s_4(3) = 3$$

$$\delta_4(4) = 0.48 \cdot 10^{-3}$$

$$s_4(4) = 4$$

· Για  $t=5$ : ( $O_5 = v$ )

$$\delta_5(1) = 0.23 \cdot 10^{-3}$$

$$\delta_5(2) = 0.46 \cdot 10^{-3}$$

$$s_5(1) = 2$$

$$s_5(2) = 2$$

$$\delta_5(3) = 0.17 \cdot 10^{-3}$$

$$s_5(3) = 2$$

$$\delta_5(4) = 0.115 \cdot 10^{-3}$$

$$s_5(4) = 2$$

Για  $t=6$ : ( $O_6 = v$ )

$$\delta_6(1) = 0.046 \cdot 10^{-3}$$

$$\delta_6(2) = 0.092 \cdot 10^{-3}$$

$$s_6(1) = 2$$

$$s_6(2) = 2$$

$$\delta_6(3) = 0.035 \cdot 10^{-3}$$

$$s_6(3) = 2$$

$$\delta_6(4) = 0.023 \cdot 10^{-3}$$

$$s_6(4) = 2$$

Για  $t=7$  : ( $O_7 = U$ )

$$\delta_7(1) = 9,2 \cdot 10^{-6}$$

$$S_7(1) = 2$$

$$\delta_7(2) = 4,6 \cdot 10^{-6}$$

$$S_7(2) = 2$$

$$\delta_7(3) = 20,7 \cdot 10^{-6}$$

$$S_7(3) = 2$$

$$\delta_7(4) = 18,4 \cdot 10^{-6}$$

$$S_7(4) = 2$$

Για  $t=8$  : ( $O_8 = U$ )

$$\delta_8(1) = 4,14 \cdot 10^{-6}$$

$$S_8(1) = 3$$

$$\delta_8(2) = 1,104 \cdot 10^{-6}$$

$$S_8(2) = 4$$

$$\delta_8(3) = 3,1 \cdot 10^{-6}$$

$$S_8(3) = 3$$

$$\delta_8(4) = 3,68 \cdot 10^{-6}$$

$$S_8(4) = 4$$

Για  $t=9$  : ( $O_9 = V$ )

$$\delta_9(1) = 0,62 \cdot 10^{-6}$$

$$S_9(1) = 3$$

$$\delta_9(2) = 0,88 \cdot 10^{-6}$$

$$S_9(2) = 4$$

$$\delta_9(3) = 0,31 \cdot 10^{-3}$$

$$S_9(3) = 1$$

$$\delta_9(4) = 0,207 \cdot 10^{-3}$$

$$S_9(4) = 1$$

Για  $t=10$  : ( $O_{10} = U$ )

$$\delta_{10}(1) = \frac{0,088}{0,044} \cdot 10^{-6}$$

$$S_{10}(1) = 1$$

$$\delta_{10}(2) = 0,044 \cdot 10^{-6}$$

$$S_{10}(2) = 2$$

$$\delta_{10}(3) = 0,198 \cdot 10^{-6}$$

$$S_{10}(3) = 2$$

$$\delta_{10}(4) = 0,176 \cdot 10^{-6}$$

$$S_{10}(4) = 2$$

2) Επομένως, σύμφωνα με τον αλγόριθμο Viterbi, η πιο πιθανή ακολούθια καταστάσεων είναι:

$$Q^* = (4 - 2 - 4 - 2 - 2 - 2 - 4 - 4 - 2 - 3)$$

$$P^* = (O, Q^* | \lambda) = \delta_{10}(3) = 0,198 \cdot 10^{-6}$$

$$\Rightarrow P^* \approx 1,99 \cdot 10^{-7}$$

## Άσκηση 4:

Δίνεται το ακόλουθο φωνητικό λεξικό:

any	eh n iy
e.	i y
many	m eh n iy
men	m eh n
per	p er
persons	p er s uh n z
sons	s uh n z
suns	s uh n z
to	t uw
tomb	t uw m
too	t uw
two	t uw

Προκειμένου να υπολογίσουμε τη φωνολογική απόσταση μεταξύ των λέξεων του παραπάνω dictionary, χρειάζεται να υλοποιήσουμε την απόσταση Levenshtein με χρήση της βιβλιοθήκης OpenFST.

Αρχικά, δημιουργούμε το αρχείο **phones.syms** το οποίο περιλαμβάνει όλα τα φωνήματα που εμφανίζονται στο λεξικό ταξινομημένα σε αλφαριθμητική σειρά:

<eps> 0

eh 1

er 2

i y 3

m 4

n 5

p 6

s 7

t 8

uh 9

uw 10

z 11

Στη συνέχεια, δημιουργούμε τη μηχανή πεπερασμένης κατάστασης που υλοποιεί την απόσταση Levenshtein για τις λέξεις του dictionary, με χρήση του python script **make\_L.py** λαμβάνοντας υπόψη ότι η απόσταση για διαγραφή (deletion) ή πρόσθεση (insertion) φωνήματος είναι 1.3, ενώ το κόστος της αντικατάστασης (substitution) δύο τυχαίων φωνημάτων είναι 1.0. Επίσης, θεωρούμε ότι η απόσταση ανάμεσα στα φωνήματα {uh, uw} και στα φωνήματα {er, eh} είναι η μισή της απόστασης μεταξύ δύο τυχαίων φωνημάτων. Το ίδιο για τα κρουστικά σύμφωνα {r,t}, τα ένρινα {m,n} και τα συριστικά {s,z}. Το αυτόματο αποθηκεύεται στο αρχείο **L.fst**.

Τέλος, με χρήση του bash script **word\_edits.sh**, το οποίο χρησιμοποιεί και το python script **mkfstinput.py** που δημιουργεί αυτόματα που αποδέχονται φώνημα προς φώνημα λέξεις του dictionary, μπορούμε να υπολογίσουμε την ελάχιστη φωνολογική απόσταση ανάμεσα σε δύο οποιεσδήποτε λέξεις του corpus, καθώς και τα edits που απαιτούνται για να ταυτιστούν φωνολογικά μεταξύ τους. Αυτό επιτυγχάνεται με χρήση της εντολής:

```
bash word_edits.sh <word1> <word2>
```

Ενδεικτικά, παραθέτουμε τις ελάχιστες φωνολογικές αποστάσεις της λέξης 'any' με όλες τις υπόλοιπες λέξεις του corpus:

- any e

```
eh      <eps>  1.29999995  
n      <eps>  1.29999995
```

- any men

```
<eps>  m      1.29999995  
iy      <eps>  1.29999995
```

- any per

```
eh      <eps>  1.29999995  
iy      er      1  
n      p      1
```

- any persons

```
<eps>  p      1.29999995  
iy      z      1  
<eps>  uh     1.29999995  
<eps>  s      1.29999995  
eh      er     0.5
```

- any sons

<eps>	s	1.29999995
iy	z	1
eh	uh	1

- any suns

<eps>	s	1.29999995
iy	z	1
eh	uh	1

- any to

eh	<eps>	1.29999995
iy	uw	1
n	t	1

- any tomb

eh	t	1
iy	m	1
n	uw	1

- any too

eh	<eps>	1.29999995
iy	uw	1
n	t	1

- any two

eh	<eps>	1.29999995
iy	uw	1
n	t	1



*Phone Levenshtein Distance - L FST*

## Άσκηση 5:

Το λεξικό μιάς φανταστικής παιδικής γλώσσας αποτελείται από τις ακόλουθες συλλαβές: {Μπα, Ντα, Γκα, Τσα}. Ένας γλωσσολόγος συνέλεξε τα παρακάτω δεδομένα από παιδιά ηλικίας ενός έτους:

Μπα Μπα Τσα Ντα Ντα Τσα Γκα Τσα Μπα Μπα Γκα Γκα Τσα Ντα Ντα Γκα Ντα Ντα Ντα  
Μπα Μπα Τσα Γκα Τσα Ντα Ντα Μπα Μπα Τσα Μπα Μπα Ντα Ντα Τσα Τσα Μπα Μπα  
Γκα Γκα Τσα Ντα Γκα Ντα Ντα Μπα Μπα Τσα Γκα Τσα Ντα Ντα Μπα Μπα  
Τσα

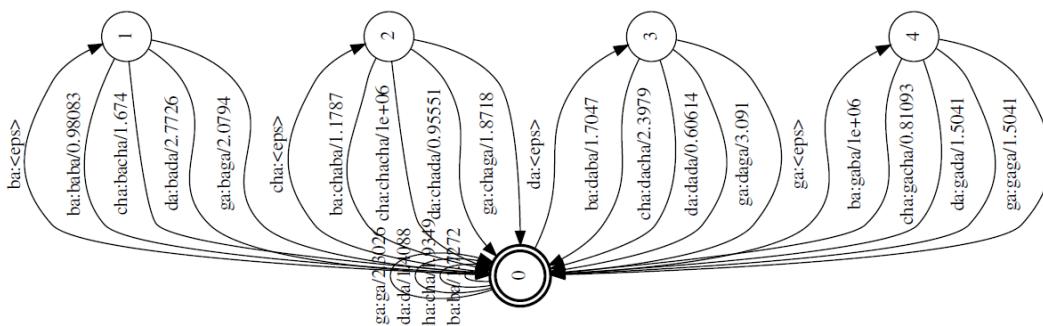
Ο γλωσσολόγος ξέρει ότι όλες οι λέξεις αυτής της παιδικής γλώσσας περιέχουν μία ή δύο συλλαβές, π.χ. Μπα, Ντα, Γκα, Τσα, ΜπαΜπα, ΜπαΝτα, ΜπαΓκα, ΜπαΤσα, ΝταΜπα, ΝταΝτα. Οι λέξεις με μία συλλαβή εμφανίζονται εξίσου συχνά με τις λέξεις με δύο συλλαβές, δηλαδή,  $P(\text{Μπα}) + P(\text{Ντα}) + P(\text{Γκα}) + P(\text{Τσα}) = 0.5$

- 1) Δεδομένου ότι οι λέξεις με μία συλλαβή εμφανίζονται εξίσου συχνά με τις λέξεις με δύο συλλαβές, μπορούμε να θεωρήσουμε ότι τα παραπάνω δεδομένα αποτελούνται εξ ολοκλήρου από δισύλλαβες λέξεις. Δηλαδή:

ΜπαΜπα ΤσαΝτα ΝταΤσα ΓκαΤσα ΜπαΜπα ΓκαΓκα ΤσαΝτα ΝταΓκα ΝταΝτα ΝταΝτα  
ΜπαΜπα ΤσαΓκα ΤσαΝτα ΝταΝτα ΜπαΜπα ΤσαΜπα ΜπαΝτα ΝταΤσα ΤσαΜπα  
ΜπαΓκα ΓκαΤσα ΝταΝτα ΓκαΝτα ΝταΝτα ΝταΜπα ΜπαΤσα ΓκαΤσα ΝταΝτα ΝταΜπα  
ΜπαΤσα

Στην προκειμένη περίπτωση, μετράμε με χρήση `python` ποιά δισύλλαβη λέξη εμφανίζεται τις περισσότερες φορές στο παραπάνω δείγμα. Βρίσκουμε ότι η **λέξη dada είναι η πιο συχνά εμφανιζόμενη δισύλλαβη**, καθώς την παρατηρούμε συνολικά 6 φορές, ενώ το κόστος εμφάνισης της εν λόγω λέξης είναι 0,606.

- 2) Αρχικά, δημιουργούμε το αρχείο **corpus.txt**, το οποίο περιλαμβάνει τα γλωσσικά δεδομένα που συνέλεξε ο γλωσσολόγος, και το αρχείο **words.syms**, το οποίο περιλαμβάνει όλες τις λέξεις του λεξιλογίου της παιδικής γλώσσας, είτε πρόκειται για μονοσύλλαβες είτε για δισύλλαβες. Στη συνέχεια, με χρήση του script "**lmodel.py**" μετράμε τη πιθανότητα εμφάνισης κάθε μονοσύλλαβης και δισύλλαβης λέξης και τη μετατρέπουμε σε κόστος με χρήση του τροπικού ημιδακτυλίου (`cost = -logP`). Με βάση τα κόστη αυτά, υλοποιούμε το FST `L.fst`:



Έπειτα, με χρήση του script “**mkfstinput.py**” δημιουργούμε έναν transducer που αποδέχεται συλλαβή προς συλλαβή το περιεχόμενο του **corpus.txt**, τον οποίο ονομάζουμε **V.fst**.

Στη συνέχεια, με την εντολή **fstcompose** παίρνουμε τη σύνθεση των δύο transducers και το αποτέλεσμα της αποθηκεύεται ως **VL.fst**. Τέλος, με χρήση της εντολής **fstshortestpath** πάνω στο **VL.fst**, βρίσκουμε το μονοπάτι ελάχιστου κόστους, δηλαδή την πιο πιθανή ακολουθία από λέξεις πάνω στα δεδομένα του γλωσσολόγου. Το αποτέλεσμα αποθηκεύεται στον transducer **path.fst**.

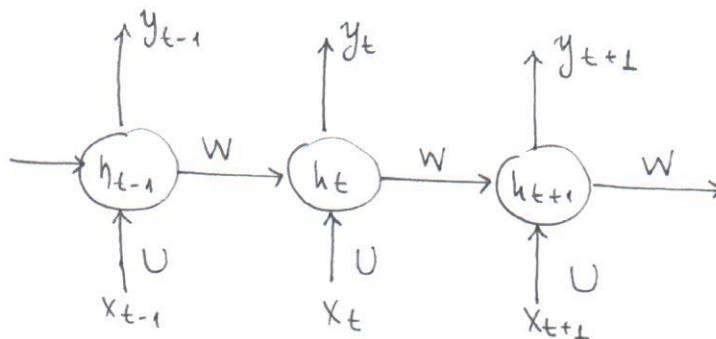
Τελικώς, η πιο πιθανή σειρά από λέξεις στην παραπάνω σειρά από συλλάβες που συνέλεξε ο γλωσσολόγος είναι:

“ΜπαΜπα ΤσαΝτα ΝταΤσα ΓκαΤσα ΜπαΜπα **Γκα** ΓκαΤσα ΝταΝτα ΓκαΝτα ΝταΝτα ΝταΜπα  
ΜπαΤσα ΓκαΤσα ΝταΝτα ΝταΜπα ΜπαΤσα ΜπαΜπα ΝταΝτα **Τσα** ΤσαΜπα ΜπαΓκα  
ΓκαΤσα ΝταΝτα ΓκαΝτα ΝταΜπα ΜπαΤσα ΓκαΤσα ΝταΝτα ΝταΜπα ΜπαΤσα”

## Άσκηση 6:

Back Propagation through time:

Δίνεται το ακόλουθο RNN



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

$$\sigma(z) = \frac{1}{1 + e^{-z}}$$

Συνάρτηση Σφράγισματος:  $L = \sum_{t=0}^T L_t$ , ονου  $L_t$  τα επικερδούσ χρονικά σφράγισα.

a) Δίνεται ότι  $y = \sigma(w \cdot x)$ , ονου  $y \in \mathbb{R}^n$ ,  $x \in \mathbb{R}^d$ ,  $w \in \mathbb{R}^{n \times d}$ .

Tότε:

$$\frac{\partial y}{\partial x} = \frac{\partial \sigma(w \cdot x)}{\partial (w \cdot x)} \circ \frac{\partial (w \cdot x)}{\partial x} \quad (\text{chain rule})$$

$$\text{όνοι: } \frac{\partial (w \cdot x)}{\partial x} = w$$

Επινήλεον,  $w \in \mathbb{R}^{n \times d}$  και  $x \in \mathbb{R}^{d \times 1}$  οπότε  $w \cdot x \in \mathbb{R}^{n \times 1}$ . Επινήλεον, η σήμανσης συνάρτησης  $\sigma(\cdot)$  εφαρμόζεται στο  $w \cdot x$  elementwise. Ως εκ τούτου, η Jacobian μήτρα δίνεται:

$$\frac{\partial \sigma(w \cdot x)}{\partial w} = \begin{bmatrix} \frac{\partial \sigma(w x_1)}{\partial w x_1} & \frac{\partial \sigma(w x_1)}{\partial w x_2} & \dots & \frac{\partial \sigma(w x_1)}{\partial w x_n} \\ \frac{\partial \sigma(w x_2)}{\partial w x_1} & \frac{\partial \sigma(w x_2)}{\partial w x_2} & \dots & \frac{\partial \sigma(w x_2)}{\partial w x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial \sigma(w x_n)}{\partial w x_1} & \frac{\partial \sigma(w x_n)}{\partial w x_2} & \dots & \frac{\partial \sigma(w x_n)}{\partial w x_n} \end{bmatrix}$$

$$\text{Προφανώς: } \frac{\partial \sigma(wx_i)}{\partial w_j} = 0, \forall i \neq j$$

οπότε:

$$\frac{\partial \sigma(wx)}{\partial (wx)} = \begin{bmatrix} \frac{\partial \sigma(wx_1)}{\partial (wx_1)} & 0 & \dots & 0 \\ 0 & \frac{\partial \sigma(wx_2)}{\partial (wx_2)} & \dots & 0 \\ \vdots & 0 & \ddots & \frac{\partial \sigma(wx_n)}{\partial (wx_n)} \end{bmatrix} = \text{diag}(\sigma')$$

Τελικώς:

$$\frac{\partial y}{\partial x} = \frac{\partial \sigma(wx)}{\partial (wx)} \cdot \frac{\partial (w \cdot x)}{\partial x} \Rightarrow \boxed{\frac{\partial y}{\partial x} = \text{diag}(\sigma') \cdot w} \in \mathbb{R}^{n \times d}$$

β) Έχουμε  $L = \sum_{t=0}^T L_t$ , οπότε  $\frac{\partial L}{\partial w} = \sum_{t=0}^T \frac{\partial L_t}{\partial w}$

Όμως:

$$\frac{\partial L_t}{\partial w} = \frac{\partial L_t}{\partial y_t} \cdot \frac{\partial y_t}{\partial h_t} \cdot \frac{\partial h_t}{\partial w} \quad (\text{όπως στην προκειμένη περίτωση } y_t = h_t)$$

$$\Rightarrow \frac{\partial L}{\partial w} = \frac{\partial L_t}{\partial h_t} \cdot \frac{\partial h_t}{\partial w}$$

Ο στόχος, ισχύει:  $h_t = \sigma(w \cdot h_{t-1} + u_t)$ , δηλαδή υπάρχει εξάρτηση ανά το  $h_{t-1}$ .

Με εφαρμογή κανονικά ανυσίδας:

$$\frac{\partial h_t}{\partial w} = \frac{\partial h_t}{\partial h_{t-1}} \cdot \frac{\partial h_{t-1}}{\partial w}$$

Όμως:

$$\frac{\partial h_{t-1}}{\partial w} = \frac{\partial h_{t-1}}{\partial h_{t-2}} \cdot \frac{\partial h_{t-2}}{\partial w} \quad \text{k.o.k.} \rightarrow \text{Προκυπτει αναδρομική εξάρτηση.}$$

Επομένως:

$$\frac{\partial L^+}{\partial w} = \frac{\partial L^+}{\partial h_t} \cdot \frac{\partial h_t}{\partial h_{t-1}} \cdot \frac{\partial h_{t-1}}{\partial w}$$

Αρα, μέχρι τη χρονική σεγκί  $t$ , μπορούμε να υπολογίσουμε το gradient και στη συέχεια με backpropagation στον χέρο από τη χρονική σεγκί  $t$  μέχρι τη χρονική σεγκί  $1$ , να υπολογίσουμε το ολικό gradient:

$$\frac{\partial L^+}{\partial w} = \sum_{k=1}^+ \frac{\partial L^+}{\partial h_t} \frac{\partial h_t}{\partial h_k} \frac{\partial h_k}{\partial w}$$

Λαμβάνοντας υπόψη τους αναδρομικούς καίρους αλυσίδας, δέχαμε ότι:

$$\frac{\partial h_t}{\partial h_k} = \frac{\partial h_t}{\partial h_{t-1}} \cdot \frac{\partial h_{t-1}}{\partial h_{t-2}} \cdots \frac{\partial h_{k+1}}{\partial h_k} = \prod_{j=k+1}^t \frac{\partial h_j}{\partial h_{j-1}}$$

Συνεπώς, το συνολικό σφάλμα, αν παραχωρίσουμε ως προς  $w$ , λαμβάνουμε:

$$\frac{\partial L}{\partial w} = \sum_{t=0}^T \sum_{k=1}^+ \frac{\partial L^+}{\partial h_t} \cdot \frac{\partial h_t}{\partial h_k} \cdot \frac{\partial h_k}{\partial w}$$

## Vanishing / Exploding Gradients:

γ) Για τιμή χρονικού ορίστρα  $T=3$ , λαμβάνουμε:

$$\begin{aligned}
 \frac{\partial L}{\partial w} &= \frac{\partial L_1}{\partial h_1} \cdot \frac{\partial h_1}{\partial w} + \frac{\partial L_2}{\partial h_2} \cdot \left[ \frac{\partial h_2}{\partial h_1} \cdot \frac{\partial h_1}{\partial w} + \frac{\partial h_2}{\partial w} \right] + \\
 &+ \frac{\partial L_3}{\partial h_3} \cdot \left[ \frac{\partial h_3}{\partial h_1} \cdot \frac{\partial h_1}{\partial w} + \frac{\partial h_3}{\partial h_2} \cdot \frac{\partial h_2}{\partial w} + \frac{\partial h_3}{\partial w} \right] \\
 &= \frac{\partial L_1}{\partial h_1} \cdot \frac{\partial h_1}{\partial w} + \frac{\partial L_2}{\partial h_2} \cdot \left[ \frac{\partial h_2}{\partial h_1} \cdot \frac{\partial h_1}{\partial w} + \frac{\partial h_2}{\partial h_1} \cdot \frac{\partial h_1}{\partial w} \right] + \\
 &+ \frac{\partial L_3}{\partial h_3} \cdot \left[ \frac{\partial h_3}{\partial h_2} \cdot \frac{\partial h_2}{\partial h_1} \cdot \frac{\partial h_1}{\partial w} + \frac{\partial h_3}{\partial h_2} \cdot \frac{\partial h_2}{\partial h_1} \cdot \frac{\partial h_1}{\partial w} + \frac{\partial h_3}{\partial h_2} \cdot \frac{\partial h_2}{\partial h_1} \cdot \frac{\partial h_1}{\partial w} \right] \\
 \Rightarrow \frac{\partial L}{\partial w} &= \frac{\partial L_1}{\partial h_1} \cdot \frac{\partial h_1}{\partial w} + \frac{\partial L_2}{\partial h_2} \cdot \left[ 2 \cdot \frac{\partial h_2}{\partial h_1} \cdot \frac{\partial h_1}{\partial w} \right] + \frac{\partial L_3}{\partial h_3} \cdot \left[ 3 \cdot \frac{\partial h_3}{\partial h_2} \cdot \frac{\partial h_2}{\partial h_1} \cdot \frac{\partial h_1}{\partial w} \right]
 \end{aligned}$$

Οπως δείχνεται στο Ερώτημα (a) για  $y = \sigma(w \cdot x)$  ισχυει  $\frac{\partial y}{\partial w} = \text{diag}(\sigma'(w)) \cdot w$ .

Στην προκειμένη περιπτώση, ονού  $h_t = \sigma(w \cdot h_{t-1} + U \cdot x_t)$  έχουμε:

$$\frac{\partial h_t}{\partial h_{t-1}} = \text{diag}(\sigma'(w^T h_{t-1} + U^T x_t)) \cdot w$$

Αν, λοιπόν, κάνουμε backpropagation για η χρονικές συρρίξεις, τότε ανδ των αναδρομικών κάνοντα αλυσίδας που δείχνεται, θα λαμβάνουμε:

$$\prod_{j=k+1}^{k+n} \frac{\partial h_j}{\partial h_{j-1}} = \prod_{j=k+1}^{k+n} \left\{ \text{diag}(\sigma'(w^T h_{t-1} + U^T x_t)) \right\} w$$

Ως αντέλεσμα, συνιστάται να είναι vanishing/exploding gradients.

δ) Κάθε διαγωνοτήσιμος πίνακας  $M$  μπορεί να αναπαραγγεί μέσω των  
ιδιοτήτων και των ιδιοδιανυσμάτων του στη μορφή:

$$M = Q \Lambda Q^{-1}$$

όπου ο πίνακας  $Q$  έχει ως i-οστή στήλην το i-οστό ιδιοδιανυσματού του  $M$   
και  $\Lambda$  διαγώνιος πίνακας με στοιχεία τους οι ιδιοτήτες του  $M$ .

Αν  $\lambda$  ιδιοτήτη του πίνακα  $M$ , τότε ισχύει:

$$Mx = \lambda \cdot x \xrightarrow{\cdot A} M^2 x = \lambda Mx = \lambda^2 x$$

Διλαδί, οι ιδιοτήτες του πίνακα  $M^2$  είναι τα τετράγωνα των ιδιοτήτων του  
πίνακα  $M$ , ενώ οι πίνακες  $M$  και  $M^2$  έχουν τα ίδια ιδιοδιανυσματα.

Άλλως:

$$M^2 = (Q \Lambda Q^{-1})^2 = Q \Lambda \cancel{Q^{-1}} \cancel{Q} \Lambda Q^{-1} = Q \Lambda^2 Q^{-1}$$

Εντελώς εναργήκα, αν υψώσουμε εις την  $k$ :

$$M^k = Q \Lambda \cancel{Q^{-1}} \cancel{Q} \Lambda Q^{-1} \cdot \cancel{Q} \cancel{Q} \Lambda Q^{-1} = Q \Lambda^k Q^{-1}$$

Διλαδί:

$$\boxed{\prod_{l=1}^n M = M^n = Q \Lambda^n Q^{-1}}$$

ε) Θεωρούμε τον πίνακα Βαρών  $W = \begin{bmatrix} 0.58 & 0.24 \\ 0.24 & 0.72 \end{bmatrix}$  με ιδιοδιαστάση:

$$W = Q \Lambda Q^{-1} = \begin{pmatrix} 0.8 & -0.6 \\ 0.6 & 0.8 \end{pmatrix} \begin{pmatrix} 0.4 & 0 \\ 0 & 0.9 \end{pmatrix} \begin{pmatrix} 0.8 & 0.6 \\ -0.6 & 0.8 \end{pmatrix}$$

Tóte :

$$\begin{aligned} W^{30} &= Q \Lambda^{30} Q^{-1} = \begin{pmatrix} 0.8 & -0.6 \\ 0.6 & 0.8 \end{pmatrix} \cdot \begin{pmatrix} 0.4^{30} & 0 \\ 0 & 0.9^{30} \end{pmatrix} \begin{pmatrix} 0.8 & 0.6 \\ -0.6 & 0.8 \end{pmatrix} = \\ &= \begin{pmatrix} 0.8 & -0.6 \\ 0.6 & 0.8 \end{pmatrix} \cdot \begin{pmatrix} 1.15 \cdot 10^{-12} & 0 \\ 0 & 0.042 \end{pmatrix} \begin{pmatrix} 0.8 & 0.6 \\ -0.6 & 0.8 \end{pmatrix} = \begin{pmatrix} 0.92 \cdot 10^{-12} & -0.0252 \\ 0.69 \cdot 10^{-12} & 0.0336 \end{pmatrix} \begin{pmatrix} 0.8 & 0.6 \\ -0.6 & 0.8 \end{pmatrix} \\ \Rightarrow W &= \begin{bmatrix} 0.01526 & -0.02034 \\ -0.02034 & 0.02713 \end{bmatrix} \end{aligned}$$

Παρατηρούμε ότι τα στοιχεία του  $W^{30}$  είναι αρκετά μικρότερα από αυτά του πίνακα  $W$ . Αυτό συμβαίνει διότι οι βιδιούμες του πίνακα  $W$  είναι ίδιες μικρότερες (κατά μέτρο) της μονάδας. Γενικότερα :

►  $|\lambda| < 1$  :

Iσχυει  $\lim_{n \rightarrow +\infty} |\lambda|^n = 0$ , αν  $|\lambda| < 1$ . Ονότε, αν για κάθε βιδιούμα στον μητρώο  $W$ , έστω  $\lambda$ , ισχυει  $|\lambda| < 1$ , τότε :

$$\lim_{n \rightarrow +\infty} W^n = \lim_{n \rightarrow +\infty} Q \begin{bmatrix} \lambda_1^n & 0 & 0 & \dots & 0 \\ 0 & \lambda_2^n & 0 & \dots & 0 \\ 0 & 0 & \ddots & \dots & 0 \end{bmatrix} Q^{-1} = 0 \rightsquigarrow \text{Ασυρτωτική Ευστάθεια}$$

►  $|\lambda| > 1$  : Iσχυει  $\lim_{n \rightarrow +\infty} |\lambda|^n = +\infty$ , αν  $|\lambda| > 1$ . Ονότε, ενεργειώς αναλόγα

Ωστισχυει  $\lim_{n \rightarrow +\infty} W^n = +\infty \rightsquigarrow \text{Αστάθεια}$

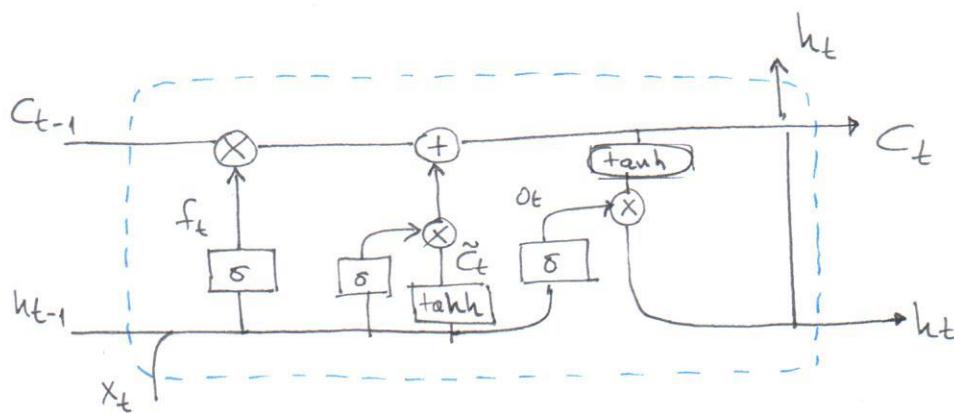
►  $|\lambda| = 1$  : Tóte :  $W = Q \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} Q^{-1} = Q \bar{Q}^{-1} = I$

Δηλαδή οριακή ευστάθεια.

στ) Όπως δείχνει, λοιπόν, για τη πολλαπλασιάσμα του μπτρώων  $\text{diag}(\sigma' w)$  αν οι ιδιοτήτες του είναι αρνητικές κατά μέτρο, τότε για  $n \rightarrow +\infty$  θα έχουμε  $\text{diag}(\sigma' w) \rightarrow 0$  (vanishing gradient), ενώ αν οι ιδιοτήτες του είναι θετικές, τότε για  $n \rightarrow +\infty$  θα έχουμε  $\text{diag}(\sigma' w) \rightarrow +\infty$  (exploding gradient).

Το πρόβλημα αυτό έρχεται να αναμετωπίσει η αρχιτεκτονική LSTM RNNs, με τιν οποια αποφασίζουμε ποια πρόσερη η πληροφορία - γύρων θέτουμε να κρατήσουμε και ποια να πετάξουμε. Η πληροφορία διακλίνεται στην χρονική ακολούθια με χρόνια των λεχόμενων κοιτασμάτων κελιών (cell states), η οποία ρέει μέσω αυτών με ελάχιστα γραμμικά interactions.

Η LSTM αρχιτεκτονική μπορεί να ελέγχει ποια πληροφορία θα προβλέψει ή να αφαιρεί από το cell state με χρόνια των gates. Ειδικότερα:



i) Forget Gate Layer:  $f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$

Μέσω αυτού το LSTM αποφασίζει ποια πληροφορία θα πετάξει από το cell state. Η απόφαση αυτή παρίσταται μέσω ενός sigmoid στρώματος, το οποίο με εισόδο τα  $x_t$  και  $h_{t-1}$ , βραγέ ως έργο έναν αριθμό ανά 0 έως 1 για κάθε αριθμό του cell state  $C_{t-1}$ . Το 1 σημαίνει "κράτα τα όλα" και το "0" "πέτα τα όλα".

## (2) Input Gate Layer:

Προκειμένου να αποφασίσουμε την καυτήρια πληροφορία που θα εισάγουμε στο cell state ακολουθούμε δύο βήματα: πρώτον ένα sigmoid layer αποφασίζει ποιές τιμές να ενημερωθούν (input gate layer) και δεύτερον ένα tanh layer δημιουργεί τις πιθανές τιμές,  $\tilde{C}_t$ , που θα μπορούσαν να προστεθούν στο cell state.

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$$

$$\tilde{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c)$$

Η ενημέρωση του cell state  $C_{t-1}$  σε  $C_t$ , δοιον, ακολουθεί τα εξής βήματα:

- Πολλαπλασιάζουμε την παλιά cell state  $C_{t-1}$  με τον όρο  $f_t$ , προκειμένου να συχνάσουμε την πληροφορία που δεν χρειαζόμαστε.
- Προσθέτουμε τον όρο  $i_t * \tilde{C}_t$  για να ενημερώσουμε με νέα πληροφορία, στον βαθμό που θέλουμε.

Δηλαδή:  $C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$

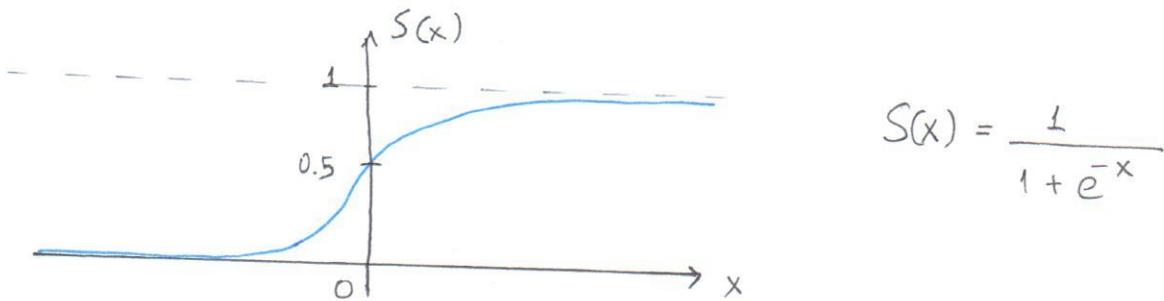
## (3) Output Gate Layer:

Χρησιμοποιείται για τον καθορισμό της πληροφορίας στην έξοδο, η οποία δικτύωσε φιλτραρισμένη εκδοχή του cell state. Πρώτον ένα sigmoid layer αποφασίζει ποια μέρη του cell state να συντηρήσουν στην έξοδο, και ένατα ένα tanh layer (με εύρους τιμών -1 έως 1) πάνω στο cell state πολλαπλασιάζει το sigmoid layer, ώστε να εξάγουμε μόνο τα κομμάτια πληροφορίας που επιθυμούμε.

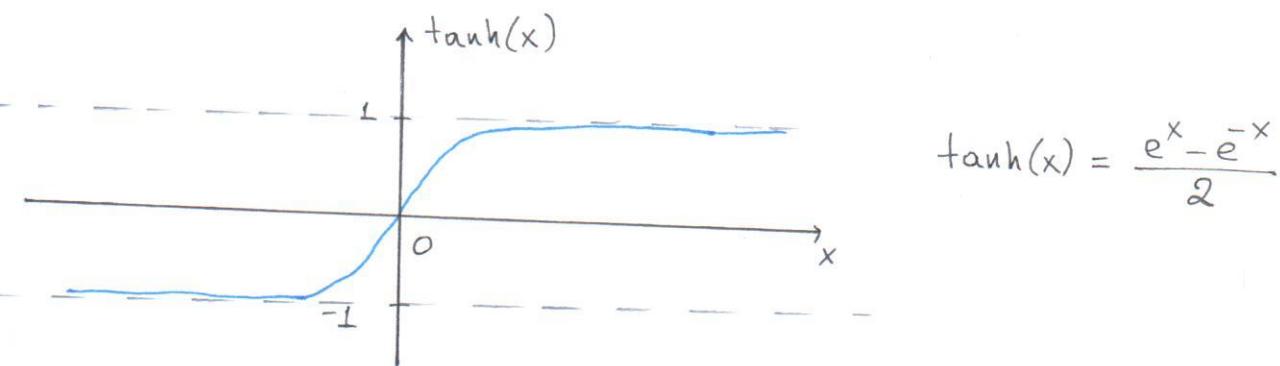
$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o)$$

$$h_t = o_t * \tanh(C_t)$$

3) Οι ποδότητες  $f_t$ ,  $i_t$ ,  $o_t$  ως sigmoid layers είναι πάντα δεκτές ή μηδέν. Παρακάτω σφίγξεται προσεγγισματικά η γραφική παράσταση της sigmoid function:



Αντίθετως, οι ποδότητες  $\tilde{C}_t$  και  $h_t = o_t \cdot \text{tanh}(\tilde{C}_t)$  ανίκουν στο διάστημα  $(-1, 1)$  λόγω του tanh layer. Παρακάτω σφίγξεται προσεγγισματικά η γραφική παράσταση της συριγμένης tanh:



) Απλονοικήσιμη Εγίσωση:

$$\frac{\partial L}{\partial W} = \sum_{t=0}^T \sum_{k=1}^t \frac{\partial L}{\partial C_t} \cdot \frac{\partial C_t}{\partial C_k} \frac{\partial C_k}{\partial W}$$

Η παραπάνω εγίσωση είναι απλονοικήσιμη, καθώς αγνούνται τις εξαρτήσεις του  $C_t$  από τους όπους  $f_t$ ,  $i_t$ ,  $\tilde{C}_t$ . Τροκενμένον να μελετήσουμε το σφανδάλιο των vanishing gradients για τα LSTMs, χρειάζεται να διάβουμε τις εξαρτήσεις αυτές υπόψη, καθώς τα  $f_t$ ,  $i_t$  και  $\tilde{C}_t$  είναι οδη συριγμέσεις του  $C_t$  και γι' αυτό πρέπει να συμπληρώνουν στον υπολογισμό της παραγώγου  $\frac{\partial C_t}{\partial C_{t-1}}$ .

δ) Δεδομένου ότι:

$$\frac{\partial C_t}{\partial C_k} = \prod_{i=k+1}^t \frac{\partial C_i}{\partial C_{i-1}}$$

Δειχνύεται ότι  $f_t = 1$  και  $i_t = 0$ . Τότε:

$$C_t = \cancel{f_t} \circ \overset{1}{C}_{t-1} + \cancel{i_t} \circ \overset{0}{\tilde{C}_t} \Rightarrow C_t = C_{t-1} \quad (\text{I})$$

Οπότε:

$$\frac{\partial C_t}{\partial C_k} = \prod_{i=k+1}^t \frac{\partial C_i}{\partial C_{i-1}} = \frac{\partial C_{k+1}}{\partial C_k} \cdot \frac{\partial C_{k+2}}{\partial C_{k+1}} \cdots \frac{\partial C_t}{\partial C_{t-1}} \stackrel{(\text{I})}{=} \frac{\partial (C_k)}{\partial C_k} \cdot \frac{\partial (C_{k+1})}{\partial C_{k+1}} \cdots \frac{\partial (C_{t-1})}{\partial C_{t-1}}$$

$$\Rightarrow \frac{\partial C_t}{\partial C_k} = 1 \cdot 1 \cdots 1 \Rightarrow \boxed{\frac{\partial C_t}{\partial C_k} = 1}$$

i) Από τα καίρια αλγόριθμα συμπληρώνων πολλών μεταβλητών έχουμε:

$$\begin{aligned} \frac{\partial C_t}{\partial C_{t-1}} &= \frac{\partial C_t}{\partial f_t} \frac{\partial f_t}{\partial h_{t-1}} \frac{\partial h_{t-1}}{\partial C_{t-1}} + \frac{\partial C_t}{\partial i_t} \frac{\partial i_t}{\partial h_{t-1}} \frac{\partial h_{t-1}}{\partial C_{t-1}} + \frac{\partial C_t}{\partial \tilde{C}_t} \frac{\partial \tilde{C}_t}{\partial h_{t-1}} \frac{\partial h_{t-1}}{\partial C_{t-1}} \\ &\quad + \frac{\partial C_t}{\partial C_{t-1}} \end{aligned}$$

Όπως:

$$\frac{\partial C_t}{\partial f_t} = \frac{\partial [f_t \circ C_{t-1} + i_t \circ \tilde{C}_t]}{\partial f_t} = C_{t-1}$$

$$\frac{\partial f_t}{\partial h_{t-1}} = \frac{\partial [\sigma(W_f h_{t-1} + U_f x_t)]}{\partial h_{t-1}} = \sigma'(\cdot) \cdot W_f \quad (\text{βλ. ερώτημα a'})$$

$$\begin{aligned} \frac{\partial h_{t-1}}{\partial C_{t-1}} &= \frac{\partial [o_{t-1} \odot \tanh(C_{t-1})]}{\partial C_{t-1}} = \frac{\partial o_{t-1}}{\partial C_{t-1}} \cdot \tanh(C_{t-1}) + o_{t-1} \odot \frac{\partial \tanh(C_{t-1})}{\partial C_{t-1}} \\ &= \frac{\partial [\sigma(W_o \cdot h_{t-2} + U_o x_{t-1})]}{\partial C_{t-1}} \cdot \tanh(C_{t-1}) + \delta = \delta \end{aligned}$$

$$\frac{\partial C_t}{\partial i_t} = \frac{\partial [f_t \circ C_{t-1} + i_t \circ \tilde{C}_t]}{\partial i_t} = \tilde{C}_t$$

$$\cdot \frac{\partial i_t}{\partial h_{t-1}} = \frac{\partial [\sigma(w_i h_{t-1} + v_i x_t)]}{\partial h_{t-1}} = \sigma'(\cdot) \cdot w_i$$

$$\cdot \frac{\partial C_t}{\partial \tilde{C}_t} = \frac{\partial [f_t \circ C_{t-1} + i_t \circ \tilde{C}_t]}{\partial \tilde{C}_t} = i_t$$

$$\cdot \frac{\partial \tilde{C}_t}{\partial h_{t-1}} = \frac{\partial [\tanh(w_c h_{t-1} + v_c x_t)]}{\partial h_{t-1}} = \tanh'(\cdot) \cdot w_c$$

και τέλος:

$$\cdot \frac{\partial C_t}{\partial C_{t-1}} = \frac{\partial [f_t \circ C_{t-1} + i_t \circ \tilde{C}_t]}{\partial C_{t-1}} = f_t$$

Συνδυάγοντας όλα τα παραπάνω, σε τελειούχη μορφή:

$$\frac{\partial C_t}{\partial C_{t-1}} = \sigma'(\cdot) w_f \delta \circ C_{t-1} + \sigma'(\cdot) w_i \delta \cdot \tilde{C}_t + i_t \circ \tanh'(\cdot) \delta + f_t$$

Συνεπώς, στα RNNs με hidden state οι παράγωγοι  $\frac{\partial h_{t+1}}{\partial h_t}$  μετά από κάποιο χρονικό διάστημα είναι θα λαμβάνουν τιμές αποκλειστικά μεγαλύτερες του 1 (exploding gradient), είτε αποκλειστικά έντος του διαστήματος  $[0, 1]$  (vanishing gradient). Ανυπέρβλητως, στα RNNs με cell states οι παράγωγοι  $\frac{\partial C_t}{\partial C_{t-1}}$ , μπορούν ανά πασα σειρικά να άλεσουν τις δεξιές δε οποιοδήποτε από τα δύο διαστήματα. Επομένως, αν βλέπουμε ότι προκειται να συγκρίνουμε στο μήδεν, μπορούμε κάθιστα να ευθύγραψουμε την αριθμό του  $f_t$  (forget gate) ώστε να λάβει μεγαλύτερη τιμή, αποφέροντας έτσι φανταχειρανούμενο vanishing gradient.