Question 5:-

Given the cumulative density fnc (cdf):-

$$g(s) = \frac{1}{1+e^{-\alpha s}}$$

$$g'(s) = \frac{\partial g}{\partial s} = \frac{\alpha e^{-\alpha s}}{(1+e^{-\alpha s})^2} = \alpha g(1-g)$$

$$\Rightarrow g' = \alpha g(1-g) = \text{p.d.f (probability density fnc.)}$$

For our ICA problem, the log likelihood fnc. was,
m = no. of sources, n = no. of mic. recordings, t = no. of samples

$$\log(\mathcal{L}) = \sum_{i=1}^{m} \sum_{j=1}^{t} \log\left(g'\left(\sum_{k=1}^{n} W_{ik} X_{kj}\right)\right) + t\log(|W|) \longrightarrow (1)$$

W: (m×n) unmixing matrix. X: (n×t) mixture

Now,

$$\frac{\partial g'(s)}{\partial \alpha} = g(1-g) + \alpha \frac{\partial g}{\partial \alpha}(1-g) + \alpha g \times -\frac{\partial g}{\partial \alpha}$$

$$= g(1-g) + \alpha\, g(1-g)(1-g)s - \alpha g\, g(1-g)s$$

$$\frac{\partial g'(s)}{\partial \alpha} = g(1-g) + \alpha s g(1-g)(1-2g)$$

$$\therefore \quad \frac{\partial g'}{\partial \alpha} = \frac{1}{\alpha} g' + g' \times s(1-2g) \Rightarrow \frac{\left(\frac{\partial g'}{\partial \alpha}\right)}{g'} = \frac{1}{\alpha} + (1-2g)s \longrightarrow ②$$

$\frac{\partial}{\partial \alpha}$ ① gives

$$\frac{1}{\mathcal{L}} \frac{\partial \mathcal{L}}{\partial \alpha} = \sum_{i=1}^{m} \sum_{j=1}^{t} \left(\frac{\frac{\partial g'\left(\sum_{k=1}^{n} W_{ik} X_{kj}\right)}{\partial \alpha}}{g'\left(\sum_{k=1}^{n} W_{ik} X_{kj}\right)} \right) + 0$$

from ②,

$$\frac{1}{\mathcal{L}} \frac{\partial \mathcal{L}}{\partial \alpha} = \sum_{i=1}^{m} \sum_{j=1}^{t} \left(\frac{1}{\alpha} + \left(1-2g\left(\sum_{k=1}^{n} W_{ik} X_{kj}\right)\right) \sum_{k=1}^{n} (W_{ik} X_{kj}) \right)$$

$$= \frac{mt}{\alpha} + \sum_{i=1}^{m} \sum_{j=1}^{t} \left[\left(1-2g\left((WX)_{ij}\right)\right) (WX)_{ij} \right]$$

$$= \frac{mt}{\alpha} + \sum_{i=1}^{m} \underbrace{\sum_{j=1}^{t} \left(1-2g(WX)\right)_{ij} \left((WX)^T\right)_{ji}}$$

$$= \frac{mt}{\alpha} + \sum_{i=1}^{m} \left((1-2g(WX)) (WX)^T \right)_{ii}$$

$$\frac{1}{\mathcal{L}} \frac{\partial \mathcal{L}}{\partial \alpha} = \frac{mt}{\alpha} + tr\left((1-2g(WX)) (WX)^T \right) \longrightarrow ③$$

tr() → trace.

Clearly $g(WX)$ contains an 'α' term, thus equating

$\frac{\partial \mathcal{L}}{\partial \alpha} = 0$ may not give a good expression to find 'α'

So we gradient descent algorithm to find the optimal

'α' value.

$$\frac{\partial \mathcal{L}}{\partial \alpha} = \mathcal{L} \times \left(\frac{mt}{\alpha} + tr\left((1-2g(WX))(WX)^T \right) \right), \left\{ \mathcal{L} = \prod_{i=1}^{m} \prod_{j=1}^{t} g'((WX)_{ij}) |W| \right.$$

$$\Rightarrow \boxed{\alpha_{k+1} = \alpha_k + \eta \cdot \mathcal{L} \times \left(\frac{mt}{\alpha_k} + tr\left((1-2g(WX))(WX)^T \right) \right)}, \mathcal{L} \text{ is know}$$

If α is a vector, i.e. each audio signal had it's own α value.

$$\alpha = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_v \\ \vdots \\ \alpha_n \end{bmatrix}, \text{ i.e., } g(\alpha_{ij} (WX)_{ij}) = \frac{1}{1 + e^{-\alpha_i (WX)_{ij}}}$$

From previous eqs., ②, we can show that

$$\frac{\partial g'(\alpha_v, (WX)_{vj})}{\partial \alpha_v} = \left(\frac{1}{\alpha_v} + (1-2g(\alpha_v, (WX)_{vj}))(WX)_{vj} \right) g'(\alpha_v, (WX)_{vj})$$

$$\therefore \frac{\left(\frac{\partial g'(\alpha_v,(WX)_{vj})}{\partial \alpha}\right)}{g(\alpha_v,(WX)_{vj})} = \frac{1}{\alpha_v} + \left(1-2g(\alpha_v,(WX)_{vj})\right)(WX)_{vj}$$

$$\log(\mathcal{L}) = \sum_{i=1}^{m}\sum_{j=1}^{t} \log(g'(\alpha_i,(WX)_{ij})) + t\log|W| \longrightarrow 4$$

differentiating wrt. each component, α_v,

$$\frac{1}{\mathcal{L}}\frac{\partial \mathcal{L}}{\partial \alpha_v} = \sum_{j=1}^{t} \frac{\left(\frac{\partial g'(\alpha_v,(WX)_{vj})}{\partial \alpha_v}\right)}{g'(\alpha_v,(WX)_{vj})} + 0$$

$$\therefore \frac{1}{\mathcal{L}}\frac{\partial \mathcal{L}}{\partial \alpha_v} = \sum_{j=1}^{t}\left(\frac{1}{\alpha_v} + (1-2g(\alpha_v,(WX)_{vj})(WX)_{vj}\right) \longrightarrow 5$$

Now define $G(s) = \frac{1}{1+e^{-s}}$, Also define Diagonal matrix D_α as

$$D_\alpha = \begin{bmatrix} \alpha_1 & 0 & 0 & \cdots & \\ 0 & \alpha_2 & 0 & \cdots & \\ 0 & 0 & \alpha_3 & & \\ \vdots & \vdots & & \ddots & \\ & & & & \alpha_m \end{bmatrix}_{m\times m}$$

Therefore (5) becomes,

$$\frac{1}{\mathcal{L}}\frac{\partial \mathcal{L}}{\partial \alpha_v} = \frac{t}{\alpha_v} + \sum_{j=1}^{t}\left(1-2G\left((D_\alpha WX)_{vj}\right)\right)\left((WX)^T\right)_{jv}$$

$$= \frac{t}{\alpha_v} + \sum_{j=1}^{t}\left(1-2G(D_\alpha WX)\right)_{vj}\left((WX)^T\right)_{jv}$$

$$\frac{1}{\mathcal{L}}\frac{\partial \mathcal{L}}{\partial \alpha_v} = \frac{t}{\alpha_v} + \left(\left(1-2G(D_\alpha WX)\right)(WX)^T\right)_{vv}$$

$$\Rightarrow \quad \frac{\partial \mathcal{L}}{\partial \alpha_v} = \mathcal{L}\,\underbrace{\left[tD_\alpha^{-1} + \left(1-2G(D_\alpha WX)\right)(WX)^T\right]_{vv}}_{H}$$

Thus we can compute the diagonal elements of matrix H and use it for a gradient descent on the α vector