$$Y = \begin{pmatrix} y_{11} & y_{12} & \cdots & y_{1n} \\ y_{21} & y_{22} & \cdots & \vdots \\ \vdots & & & \vdots \\ y_{m1} & y_{m2} & \cdots & y_{mn} \end{pmatrix}_{m\times n} \; ; \; X = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & \vdots \\ \vdots & & & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix}_{n\times p} \; ; \; YX = \begin{pmatrix} \sum_{i=1}^{n} y_{1i}x_{i1} & \sum_{i=1}^{n} y_{1i}x_{i2} & \cdots & \sum_{i=1}^{n} y_{1i}x_{ip} \\ \sum_{i=1}^{n} y_{2i}x_{i1} & \sum_{i=1}^{n} y_{2i}x_{i2} & \cdots & \sum_{i=1}^{n} y_{2i}x_{ip} \\ \vdots & \vdots & & \vdots \\ \sum_{i=1}^{n} y_{mi}x_{i1} & \sum_{i=1}^{n} y_{mi}x_{i2} & \cdots & \sum_{i=1}^{n} y_{mi}x_{ip} \end{pmatrix}_{m\times p} \; , \; \underset{m\times p}{C} = \begin{pmatrix} c_{11} & c_{12} & \cdots & c_{1p} \\ c_{21} & c_{22} & \cdots & c_{2p} \\ \vdots & & & \\ c_{m1} & c_{m2} & \cdots & c_{mp} \end{pmatrix}.$$

Let $\sigma = \tanh$.

$$\sigma(YX)\sigma(YX)^T = \begin{pmatrix} \sum_{j=1}^{p} \sigma^2\left(\sum_{i=1}^{n} y_{1i}x_{ij}\right) & & & \\ & \sum_{j=1}^{p} \sigma^2\left(\sum_{i=1}^{n} y_{2i}x_{ij}\right) & & \\ & & \ddots & \\ & & & \sum_{j=1}^{p} \sigma^2\left(\sum_{i=1}^{n} y_{mi}x_{ij}\right) \end{pmatrix}_{m\times m}$$

// only write diagonal entries since we will apply trace

$$\underset{m\times p}{C}\,\underset{p\times m}{\sigma(YX)^T} = \begin{pmatrix} \sum_{j=1}^{p}\left[c_{1j}\,\sigma\left(\sum_{i=1}^{n} y_{1i}x_{ij}\right)\right] & & & \\ & \sum_{j=1}^{p}\left[c_{2j}\,\sigma\left(\sum_{i=1}^{n} y_{2i}x_{ij}\right)\right] & & \\ & & \ddots & \\ & & & \sum_{j=1}^{p}\left[c_{mj}\,\sigma\left(\sum_{i=1}^{n} y_{mi}x_{ij}\right)\right] \end{pmatrix}_{m\times m}$$

$$f(X) = \frac{1}{2}\|\sigma(YX) - C\|_F^2 = \frac{1}{2}\operatorname{Tr}\left\{(\sigma(YX)-C)(\sigma(YX)-C)^T\right\} = \frac{1}{2}\operatorname{Tr}\left\{(\sigma(YX)-C)(\sigma(YX)^T - C^T)\right\}$$

$$= \frac{1}{2}\operatorname{Tr}\left\{\sigma(YX)\sigma(YX)^T - \sigma(YX)C^T - C\sigma(YX)^T + CC^T\right\}$$

$$= \frac{1}{2}\operatorname{Tr}\left\{\sigma(YX)\sigma(YX)^T - (C\sigma(YX)^T)^T - C\sigma(YX)^T + CC^T\right\} \qquad // \operatorname{Tr}(A) = \operatorname{Tr}(A^T)$$

$$= \frac{1}{2}\left[\operatorname{Tr}\left\{\sigma(YX)\sigma(YX)^T\right\} - 2\operatorname{Tr}\left\{C\sigma(YX)^T\right\} + \operatorname{Tr}\left\{CC^T\right\}\right].$$

$$= \frac{1}{2}\left[\sum_{k=1}^{m}\sum_{j=1}^{p}\sigma^2\left(\sum_{i=1}^{n} y_{ki}x_{ij}\right) - 2\sum_{k=1}^{m}\sum_{j=1}^{p} c_{kj}\,\sigma\left(\sum_{i=1}^{n} y_{ki}x_{ij}\right) + \operatorname{Tr}(CC^T)\right].$$

$$\frac{\delta f}{\delta x_{ij}} = \frac{1}{2}\left[\sum_{k=1}^{m}\left\{2\sigma\left(\sum_{i=1}^{n} y_{ki}x_{ij}\right)\left(1 - \sigma^2\left(\sum_{i=1}^{n} y_{ki}x_{ij}\right)\right) y_{ki}\right\} - 2\sum_{k=1}^{m}\left\{c_{kj}\left(1 - \sigma^2\left(\sum_{i=1}^{n} y_{ki}x_{ij}\right)\right) y_{ki}\right\}\right]$$

$$= \sum_{k=1}^{m} y_{ki}\left[1 - \sigma^2\left(\sum_{i=1}^{n} y_{ki}x_{ij}\right)\right]\left[\sigma\left(\sum_{i=1}^{n} y_{ki}x_{ij}\right) - c_{kj}\right]$$

// relabel indices to be less confusing

$$= \sum_{k=1}^{m} y_{ki}\left[1 - \sigma^2\left(\sum_{l=1}^{n} y_{kl}x_{lj}\right)\right]\left[\sigma\left(\sum_{l=1}^{n} y_{kl}x_{lj}\right) - c_{kj}\right].$$

// i,j are the fixed row, col so this is i,j^{th} entry in our gradient matrix ∇f

$n\times p$ matrix (same dimension as X)