Note, to eq. (4.20)

$$\nabla_M E(M) = \begin{bmatrix} \frac{\partial E}{\partial M_{11}} & \cdots & \frac{\partial E}{\partial M_{1n}} \\ \vdots & & \\ \frac{\partial E}{\partial M_{m1}} & \cdots & \frac{\partial E}{\partial M_{mn}} \end{bmatrix}$$

$$M_{ij} = \left[\sum_{k=1}^{p} W(k)\right]_{ij} = \left[\sum_{k=1}^{p} y \cdot x^T\right]_{ij}$$

$$E = \frac{1}{2}\| y - \hat{y} \|_2^2 = \frac{1}{2}\| y_{m\times 1} - M_{m\times n} \cdot x_{n\times 1} \|_2^2$$

$$= \frac{1}{2}\left[\left(y_1 - \sum_{p=1}^{n} M_{1p} \cdot x_p\right)^2 + \cdots + \left(y_m - \sum_{p=1}^{n} M_{mp} \cdot x_p\right)^2\right]$$

$$\frac{\partial E}{\partial M_{ij}} = \left(y_i - \sum_{p=1}^{n} M_{ip} x_p\right)(-x_j)$$

$$\nabla_M E(M) = \begin{bmatrix} (y_1 - \underbrace{(\sum_p M_{1p} x_p}_{\hat{y}_1})(-x_1) & \cdots & (y_1 - \sum_p M_{1p} x_p)(-x_n) \\ \vdots & & \\ (y_m - \underbrace{\sum_p M_{mp} x_p}_{\hat{y}_m})(-x_1) & \cdots & (y_m - \sum_p M_{mp} x_p)(-x_n) \end{bmatrix}$$

$$= \underbrace{\begin{bmatrix} (y_1 - \hat{y}_1) \\ (y_2 - \hat{y}_2) \\ \vdots \\ (y_m - \hat{y}_m) \end{bmatrix}}_{\underset{\sim}{e}} \left[(-x_1), \quad \cdots, (-x_n)\right] = -\underset{\sim}{e} \cdot \underset{\sim}{x}^T$$