Problem 1

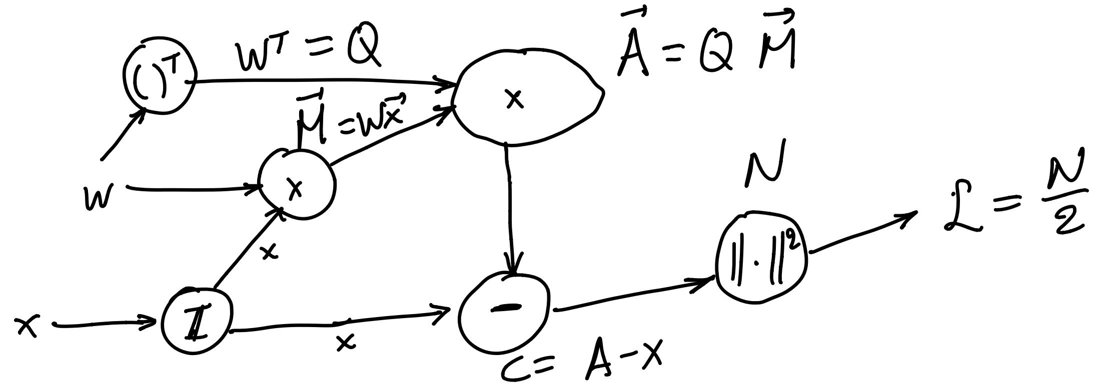
$x \in \mathbb{R}^n$

$W \in \mathbb{R}^{m \times n}$ $\quad m < n$

$$\mathcal{L} = \frac{1}{2} \| W^T W x - x \|^2$$

a) If we were able to drive $\mathcal{L} \geq 0$ all the way to zero, then $W^T W = \mathbb{I}_n$, which means that components of $W^T W x$ and x are the same. Moreover the image $y = Wx$ can be said to preserve some information of x when $W^T W = \mathbb{I}$ because Orthogonal matrices preserve some information about vectors since they are isometries and isometries preserve some structure of the vector space.

b) $\| W \vec{x} - \vec{x} \|^2$



c) At the top of the computational graph, we see the two paths to W.
These converging paths come from the fact that there is dependence on W at two places
If we call $Q = W^T$
we see that $\mathcal{L} = \mathcal{L}(Q(W), W, x)$

$$\frac{d\mathcal{L}}{dW} = \frac{\partial Q}{\partial W}\frac{\partial \mathcal{L}}{\partial Q} + \frac{\partial \mathcal{L}}{\partial W}$$

d) $$\frac{\partial \mathcal{L}}{\partial N} = \frac{1}{2}$$

$$\frac{\partial \mathcal{L}}{\partial \vec{C}} = \frac{\partial N}{\partial \vec{C}}\frac{\partial \mathcal{L}}{\partial N} = 2\vec{C}\,\frac{1}{2} = \vec{C}$$

$$\frac{\partial \mathcal{L}}{\partial \vec{A}} = \frac{\partial \vec{C}}{\partial \vec{A}}\frac{\partial \mathcal{L}}{\partial \vec{C}} = \mathbb{I}\,\vec{C}$$

$$\frac{\partial \mathcal{L}}{\partial \vec{M}} = \frac{\partial \vec{A}}{\partial \vec{M}}\frac{\partial \mathcal{L}}{\partial \vec{A}} = Q^T\vec{C}$$

$$\frac{\partial \mathcal{L}}{\partial Q} = \frac{\partial \vec{A}}{\partial Q}\frac{\partial \mathcal{L}}{\partial \vec{A}} = \frac{\partial \mathcal{L}}{\partial \vec{A}}M^T = \vec{C}M^T$$

$$\frac{dL}{dW} = \frac{\partial Q}{\partial W}\frac{\partial L}{\partial Q} + \frac{\partial \vec{M}}{\partial W}\frac{\partial L}{\partial \vec{M}}$$

$$= \mathrm{Tr}\left(\vec{C} M^T\right) + Q^T \vec{C}\,\vec{x}^T$$

$$= M\,\vec{c}^{\,T} + W\left(W^T W x - x\right)x^T$$

$$= Wx\left(W^T W x - x\right)^T + W\left(W^T W x - x\right)x^T$$

$$= -Wxx^T + Wxx^T W^T W + WW^T W x x^T - Wxx^T$$

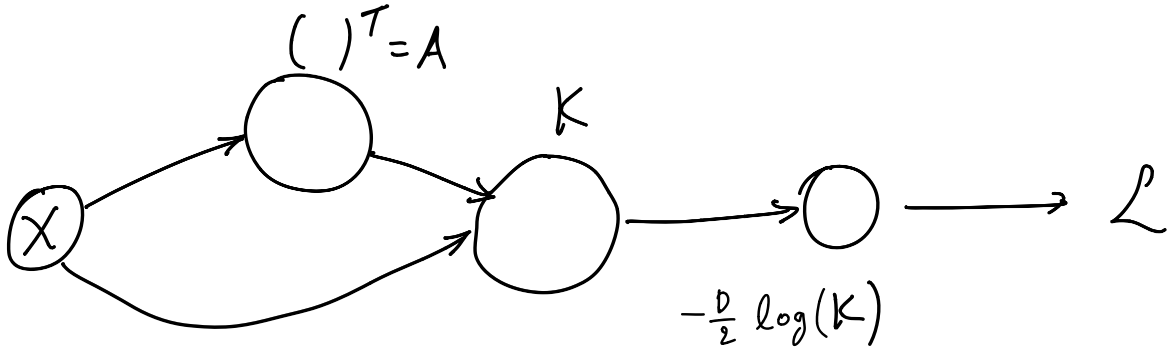
$$= \boxed{Wxx^T W^T W + WW^T W xx^T - 2Wxx^T}$$

Problem 2

$$\mathcal{L} = -c - \frac{D}{2}\log|K| - \frac{1}{2}\operatorname{tr}(K^{-1}YY^T)$$

$$\mathcal{L}_1 = -\frac{D}{2}\log|\alpha XX^T + \beta^{-1}I|$$

a)



$$\frac{\partial \mathcal{L}_1}{\partial x} = \frac{\partial K}{\partial X}\frac{\partial \mathcal{L}_1}{\partial K}$$

$$= -\frac{D}{2}\frac{dK}{dX}\frac{\partial \log|K|}{\partial K}$$

$$= -\frac{D}{2}\frac{1}{K}(\alpha 2X)$$

$$= -\frac{\alpha D}{\cancel{2}}\cancel{2}X(\alpha XX^T + \beta^{-1}I)^{-1}$$

$$= -\alpha D X(\alpha XX^T + \beta^{-1}I)^{-1}$$

here have used

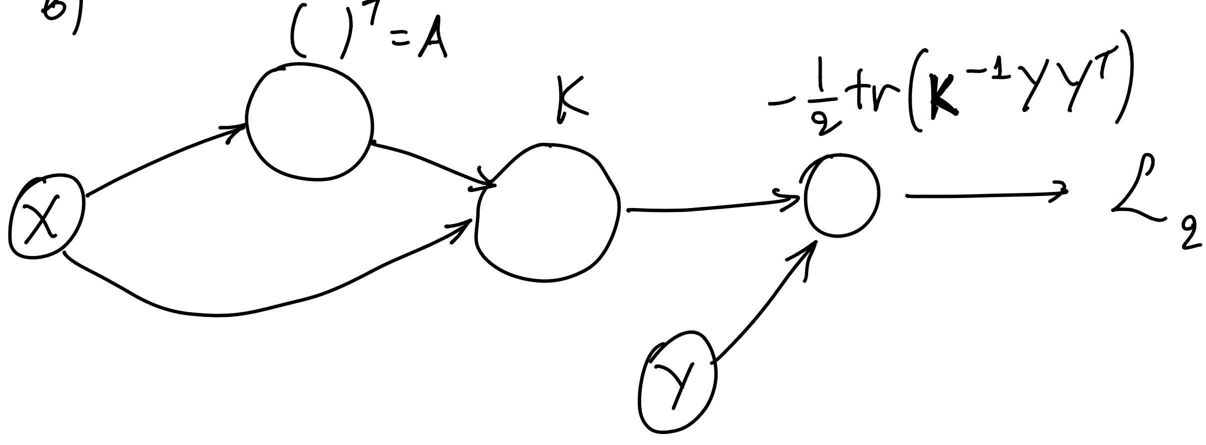
$$\frac{dK}{dX} = \frac{\partial A}{\partial X}\frac{\partial K}{\partial A} + \frac{\partial X}{\partial X}\frac{\partial K}{\partial X}$$

$$= \frac{\partial A}{\partial X}\alpha X^T + \mathbb{I}\alpha A^T$$

$$= \alpha \frac{\partial (X^T)}{\partial X} X^T + \alpha (X^T)^T$$

$$= \alpha 2 X$$

B)



$$\frac{\partial \mathcal{L}_2}{\partial X} = \frac{dK}{\partial X}\frac{\partial \mathcal{L}_2}{\partial K}$$

$$= \left(\frac{\partial A}{\partial X} \frac{\partial K}{\partial A} + \frac{\partial X}{\partial X} \frac{\partial K}{\partial X} \right) \frac{\partial \mathcal{L}_2}{\partial K}$$

$$= \left(-K^{-T} \frac{\partial \mathcal{L}_2}{\partial (K^{-1})} K^{-T} \right) \alpha\, 2X$$

$$= -\frac{1}{\cancel{2}} \alpha \cancel{2} \left(-K^{-T} \frac{\partial\, Tr(BYY^T)}{\partial B} K^{-T} \right) X$$

$$= \alpha \quad K^{-T}\, YY^T\, K^{-T} X$$

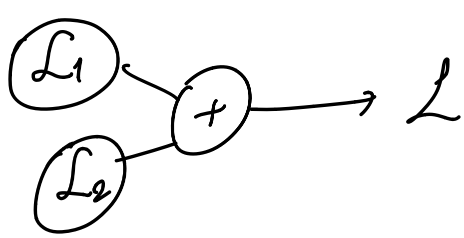
where we used $\frac{\partial}{\partial X} Tr(XA) = A^T$

However $K^{-T} = (K^T)^{-1} = K^{-1}$

hence

$$\frac{\partial \mathcal{L}_2}{\partial X} = \alpha\, K^{-1}\, YY^T K^{-1} X$$

e) Therefore from



we get

$$\frac{\partial \mathcal{L}}{\partial X} = \frac{\partial \mathcal{L}_1}{\partial X} + \frac{\partial \mathcal{L}_2}{\partial X}$$

$$= \alpha K^{-1} Y Y^T K^{-1} X$$

$$- \alpha D K^{-1} X$$