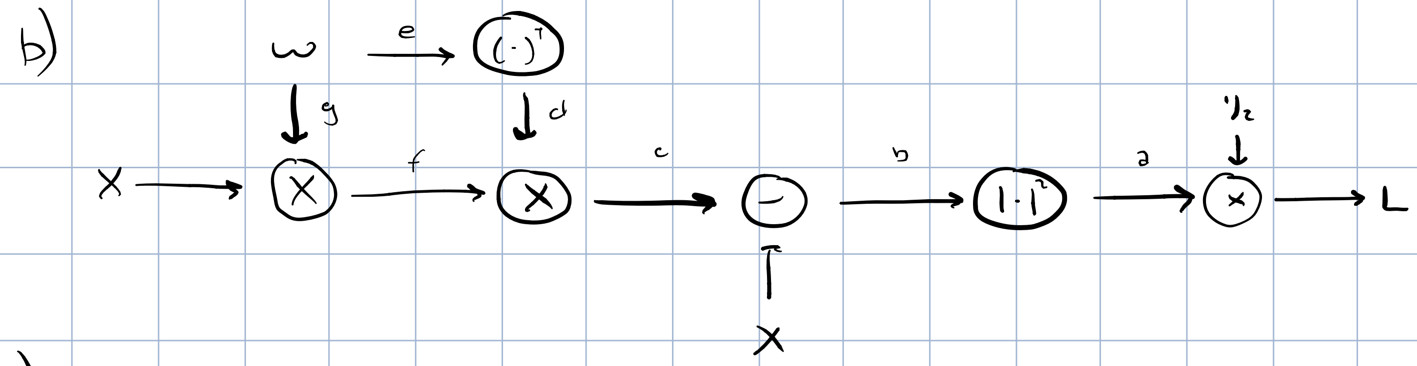
1) a) This minimization finds a W that preserves info about x because it is equivalent to PCA where $h = f(x) = W^T x$, $g(h) = Wh$. Also, we can view it in terms of SVD and the transpose - the rotations get reversed and only the scaling is left. Also $W^T W = I$ if W is orthogonal b/c W^T would be equal to W^{-1}.

b)



c) Take the derivatives with respect to both pathes and add them together. This is due to the total derivatives rule.

d) $L(a) = \frac{1}{2}a \quad = \frac{1}{2}|W^T W x - x|^2$

$a(b) = |b|^2$

$b(c) = c - x$

$c(d, f) = d \cdot f$

$d(e) = e^T$

$f(g) = g \cdot x$

$$\frac{\partial L}{\partial w_1} = \frac{\partial d}{\partial e}\ \frac{\partial c}{\partial d}\ \frac{\partial b}{\partial c}\ \frac{\partial a}{\partial b}\ \frac{\partial L}{\partial a}$$

$$\left(\left(+1 \quad 2(W^T W x - x)\ \tfrac{1}{2}\right)(Wx)^T\right)^T$$

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

$$= (Wx)(W^TWx - x)^T$$

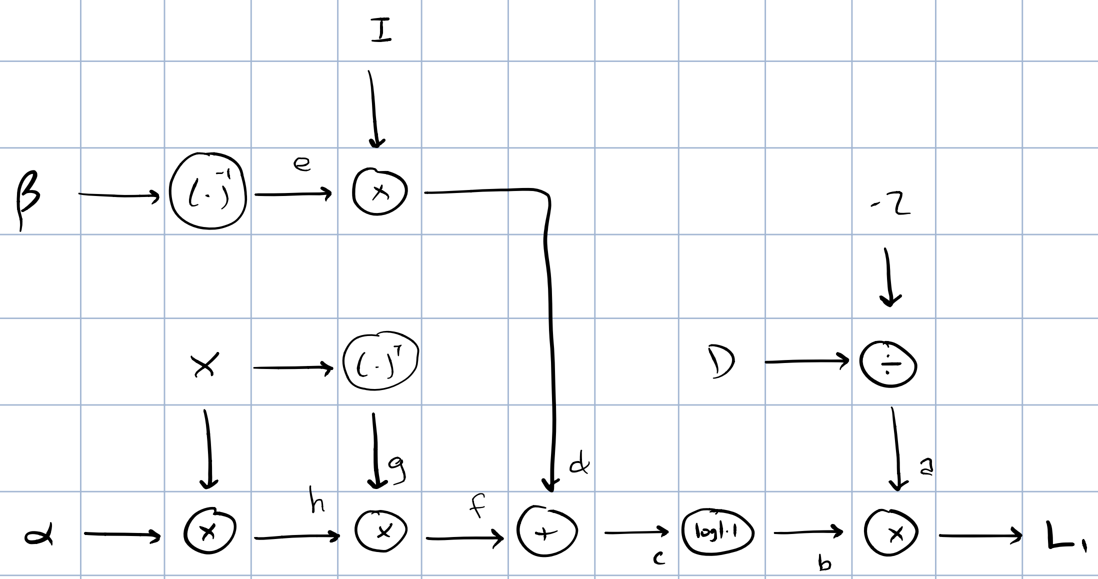
$$\frac{\partial L}{\partial w_2} = \frac{\partial f}{\partial g}\,\frac{\partial c}{\partial f}\,\frac{\partial b}{\partial c}\,\frac{\partial a}{\partial b}\,\frac{\partial L}{\partial a}$$

$$= W\,(-1)\,2(W^TWx - x)\tfrac{1}{2}\,X^T$$

$$= W(W^TWx - x)X^T$$

$$\nabla_W L = \nabla_{W_1}L + \nabla_{W_2}L = (Wx)(W^TWx - x)^T + W(W^TWx - x)X^T$$

2) a) $L_1 = -\frac{D}{2}\log|\alpha XX^T + \beta^{-1}I|$



b) $L_1 = ba$ $\qquad \frac{\partial L_1}{\partial b} = a = -\frac{D}{2}$

$b = \log|c|$ $\qquad \frac{\partial b}{\partial c} = K^{-T}$

$c = f + d$ $\qquad \frac{\partial c}{\partial f} = 1$

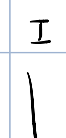
$f = h\cdot g$ $\qquad \frac{\partial f}{\partial g} = h = \alpha X$ $\qquad \frac{\partial f}{\partial h} = g^T = (X^T)^T$

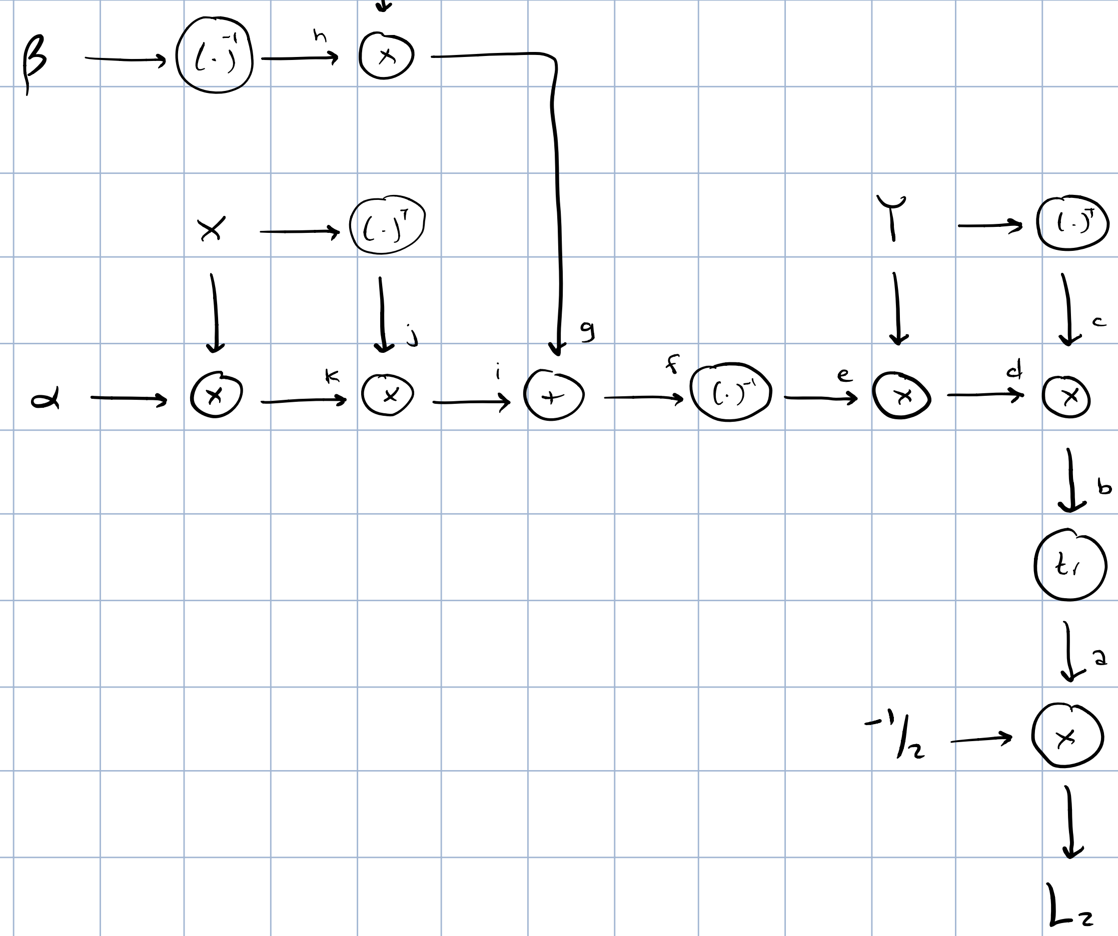
$h = \alpha X$ $\qquad \frac{\partial g}{\partial X} = I$ $\qquad \frac{\partial h}{\partial X} = \alpha$

$g = X^T$

$$\frac{\partial L_1}{\partial X} = -\frac{D}{2}K^{-T}(\alpha X + \alpha X) = -DK^{-T}\alpha X$$

c) $L_2 = -\frac{1}{2}\operatorname{tr}(\alpha XX^T + \beta^{-1}I)^{-1}YY^T$





d) $\frac{\partial L_2}{\partial X} = \frac{\partial K}{\partial X} \frac{\partial L_2}{\partial K}$ $\qquad$ $\frac{\partial L_2}{\partial K} = -K^{-T} \frac{\partial L_2}{\partial K^{-1}} K^{-T}$

$$\frac{\partial L_2}{\partial K^{-1}} = \frac{\partial}{\partial K^{-1}} \left(-1/2 \, tr\left(KYY^T\right)\right)$$

$$= -1/2 \, YY^T$$

$$\frac{\partial K}{\partial X} = \left(\alpha X + \alpha X\right)$$ $\longleftarrow$ from above

$$\frac{\partial L_2}{\partial X} = (2\alpha X)\, K^{-T} \left(1/2 \, YY^T\right) K^{-T}$$

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

e) $\frac{\partial L}{\partial X} = \frac{\partial L_1}{\partial X} + \frac{\partial L_2}{\partial X}$

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