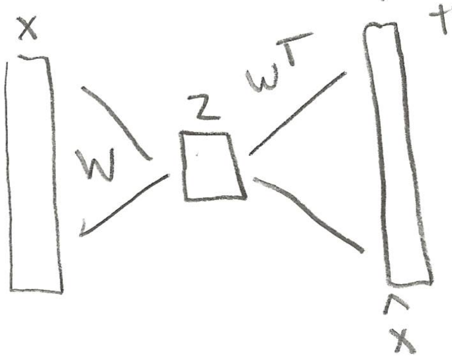
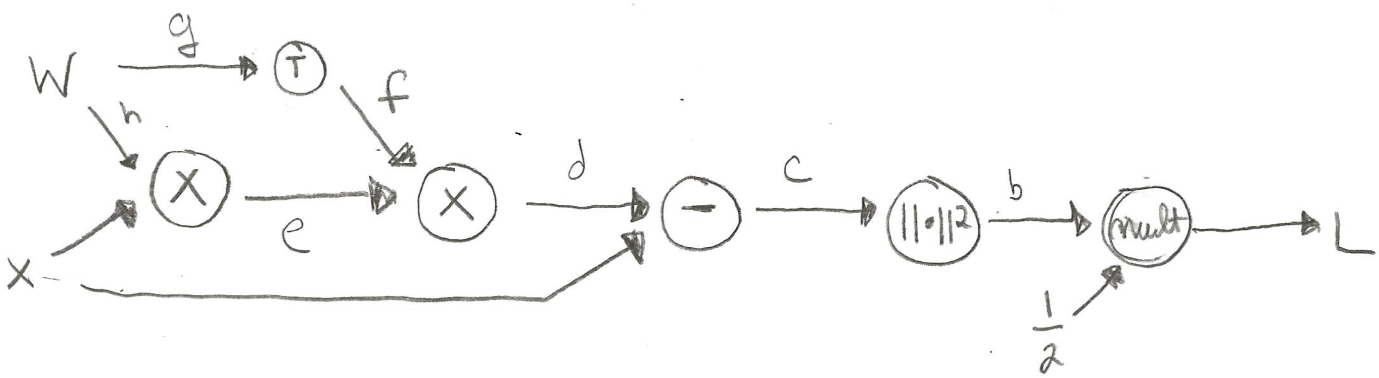
1.

a. The matrix multiplication of Wx projects x into a latent dimension z, and then the matrix multiplication of W^Tz tries to reconstruct the original projected x from z. Thus the l_2 norm penalizes entries of the reconstructed x, $\hat{x}$, that does not match x. To reconstruct well the projected x (z) must also be optimized to highly represent x.



$\hat{x} - x$ should be close, and if this does not hol[d] we penalize! $\hat{x} = W^Tz$.

b.



c. To account for these two paths to W when calculating $\nabla_W C$, you add the gradients that occur from the two paths. I.e. The law of total derivatives

$$\boxed{\begin{aligned}\nabla_W C &= \nabla_g L + \nabla_h L \\ &= \frac{dL}{dW} + \frac{dW^T}{dW}\frac{dL}{dW^T}\end{aligned}}$$

d.

$$L = \frac{1}{2} b \qquad \frac{dL}{db} = \frac{1}{2}$$

Scalar $\quad b = \|c\|^2 = c^T c$

$n\times 1 \quad c = d - x$

$n\times 1 \quad d = fe$

$m\times 1 \quad e = hx$

$n\times m \quad f = g^T$

$m\times n \quad h = W$

$m\times n \quad g = W$

$$\underset{n\times 1}{\frac{dL}{dc}} = \underset{n\times 1}{\frac{db}{dc}} \cdot \underset{1\times 1}{\frac{dL}{db}} = 2c \cdot \frac{1}{2} = c$$

$$\underset{n\times 1}{\frac{dL}{dd}} = \underset{n\times n}{\frac{dc}{dd}} \cdot \underset{n\times 1}{\frac{dL}{dc}} = I \cdot c = c$$

$$\underset{n\times m}{\frac{dL}{df}} = \underset{n\times m\times n}{\frac{dd}{df}} \cdot \underset{n\times 1}{\frac{dL}{dd}} = \underset{n\times 1}{\frac{dL}{dd}} \underset{1\times m}{e^T}$$ (trick from formal notes)

$$\frac{d d_1}{df} = \begin{bmatrix} -e^T- \\ 0 \\ | \end{bmatrix} \quad \frac{d d_2}{df} = \begin{bmatrix} 0 \\ -e^T- \\ 0 \\ | \end{bmatrix} \quad \frac{d d_n}{df} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ -e^T- \end{bmatrix}$$

$$\underset{m\times 1}{\frac{dL}{de}} = \underset{m\times n}{\frac{dd}{de}} \cdot \underset{n\times 1}{\frac{dL}{dd}} = \underset{m\times n}{f^T} \underset{n\times 1}{c}$$

$$\underset{m\times n}{\frac{dL}{dg}} = \underset{(m\times n)\times(n\times m)}{\frac{df}{dg}} \cdot \underset{(n\times m)\times 1}{\frac{dL}{df}} \quad \frac{df_{1,1}}{dg} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \ddots & & \\ 0 & & & \\ 0 & & & \end{bmatrix} \quad \frac{df_{1,2}}{dg} = \begin{bmatrix} 0 & - & - \\ 1 & \ddots & \\ 0 & & \\ | & & \end{bmatrix} \quad \frac{df_{2,1}}{dg} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & \ddots & & \\ | & & & \end{bmatrix}$$

$$= \underset{m\times n}{\left(\frac{dL}{df}\right)^T} \qquad \frac{df_{n,m}}{dg} = \begin{bmatrix} 0 & 0 & - \\ 0 & \ddots & \\ | & & 1 \end{bmatrix} \quad \frac{dL}{dg} = \left(\frac{dL}{df}\right)_{1,1} \begin{bmatrix} 1 & 0 & - \\ 0 & \ddots & \\ | & & \end{bmatrix} + \dots + \left(\frac{dL}{df}\right)_{n,m} \begin{bmatrix} 0 & - \\ | & \ddots \\ & & 1 \end{bmatrix}$$

$$\sum_{i=1}\sum_{j=1} \left(\frac{dL}{df}\right)_{i,j} \left(\frac{df_{i,j}}{dg}\right)$$

$$\frac{dL}{dh} = \underset{m\times n\times m}{\frac{de}{dh}} \cdot \underset{m\times 1}{\frac{dL}{de}} = \underset{m\times 1}{\frac{dL}{de}} \underset{1\times n}{x^T}$$ (same trick from formal notes)

$$\frac{\partial L}{\partial W} = \frac{\partial L}{\partial g} + \frac{\partial L}{\partial h}$$

$$\frac{\partial L}{\partial g} = \left(\frac{\partial L}{\partial f}\right)^T = \left(\frac{\partial L}{\partial d} e^T\right)^T = \left(c e^T\right)^T = \left((d-x)e^T\right)^T = \left((fe-x)e^T\right)^T = \left((W^T e - x)e^T\right)^T$$

$$= \left((W^T W x - x)(Wx)^T\right)^T = \underset{m\times n \;\; n\times 1}{Wx}\,(\overbrace{\underset{n\times 1}{W^T W x} - \underset{n\times 1}{x}}^{1\times n})^T$$

$$(m\times 1)\cdot(1\times n) = m\times n$$

$$\frac{\partial L}{\partial h} = \frac{\partial L}{\partial c} x^T = (f^T c)x^T = (f^T(d-x))x^T = (f^T(W^T e - x))x^T = (W(W^T W x - x))x^T$$

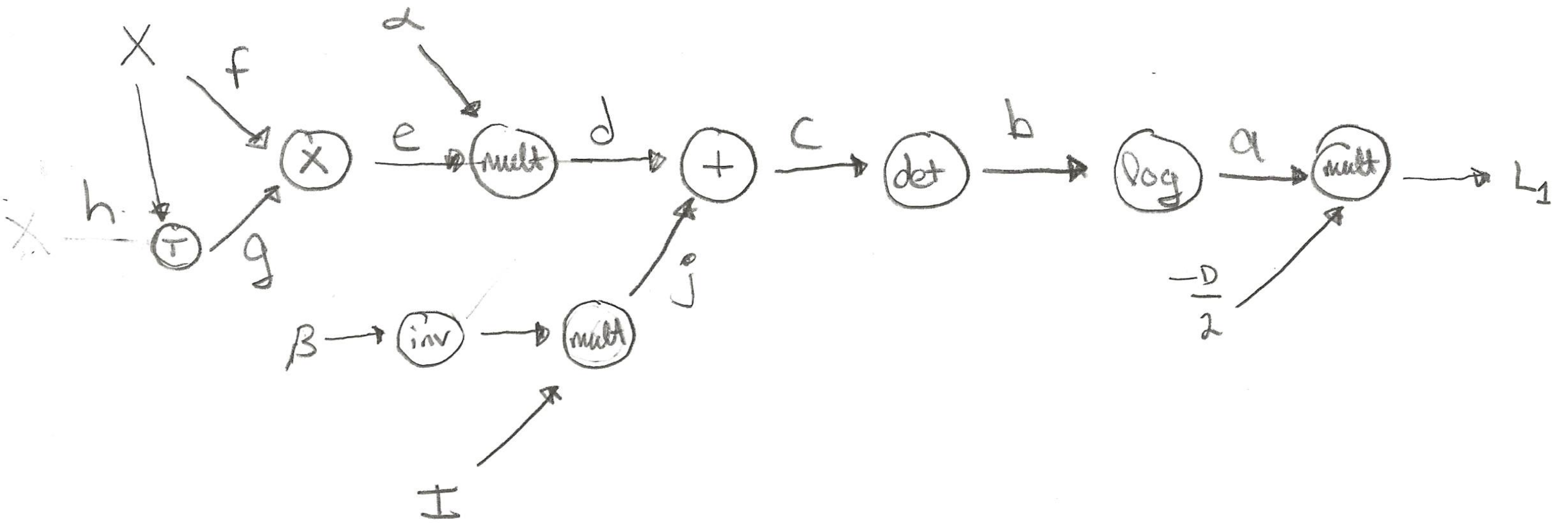
$$= (\underset{m\times n}{W}\underbrace{(W^T W x - x)}_{1\times 1})\underset{1\times n}{x^T}$$

$$= m\times n$$

$$\boxed{\frac{\partial L}{\partial W} = \frac{\partial L}{\partial h} + \frac{\partial L}{\partial g} = Wx(W^T W x - x)^T + (W(W^T W x - x))x^T}$$

2.

a.



(b)

scalar

$L_1 = \frac{-D}{2} \cdot a$

scalar $\frac{dL}{da} = \frac{-D}{2}$

scalar 1x1 $a = \log b$

scalar $\frac{dL}{db} = \frac{da}{db} \cdot \frac{dL}{da} = \frac{1}{b} \cdot \left(-\frac{D}{2}\right) = \frac{-D}{2b}$

scalar 1x1 $b = \det c$

2D matrix $\underset{m\times m}{\frac{dL}{dc}} = \underset{m\times m}{\frac{db}{dc}} \cdot \underset{1\times 1}{\frac{dL}{db}} = \underbrace{\det(c)(c^{-1})^T}_{\text{matrix (from cookbook)}} \cdot \underbrace{\frac{dL}{db}}_{\text{scalar}}$

matrix mxm $C = j + d$

matrix mxm $d = \alpha e$

2D matrix $\underset{m\times m}{\frac{dL}{dd}} = \underbrace{\frac{dc}{dd}}_{\substack{\text{4D matrix}\\ m\times m\times(m\times m)}} \cdot \underbrace{\frac{dL}{dc}}_{\substack{\text{2D matrix}\\ (m\times m)\times 1}} = \left(\frac{dL}{dc}\right)$ (showed derivations in last problem)

matrix mxm $e = f g$

matrix nxm $g = h^T$

2D matrix mxm $\frac{dL}{de} = \underbrace{\frac{dd}{de}}_{\substack{\text{4D matrix}\\ m\times m\times m\times m}} \cdot \underbrace{\frac{dL}{dd}}_{\substack{\text{2D matrix}\\ (m\times m)\times 1}} = \alpha\left(\frac{dL}{dd}\right)$ (showed derivations in last problem)

matrix mxn $h = X$

$f = X$

$X \in m\times n$

$I \in m\times m$

2D matrix $\underset{n\times m}{\frac{dL}{dg}} = \underbrace{\frac{de}{dg}}_{\substack{\text{4D matrix}\\ n\times m\times m\times m}} \cdot \underbrace{\frac{dL}{de}}_{\substack{\text{2D matrix}\\ m\times m}} = \underset{m\times m}{f^T} \underset{m\times m}{\frac{dL}{de}}$

$$\underset{m\times n}{\frac{\partial L}{\partial h}} = \underset{m\times n\times n\times m}{\frac{\partial g}{\partial h}} \circ \underset{n\times m}{\frac{\partial L}{\partial g}} = \left(\frac{\partial L}{\partial g}\right)^T \rightarrow \text{trick from before}$$

$$\underset{m\times n}{\frac{\partial L}{\partial f}} = \underset{m\times n\times m\times m}{\frac{\partial e}{\partial f}} \cdot \underset{m\times m}{\frac{\partial L}{\partial e}} = \underset{m\times m}{\frac{\partial L}{\partial e}} \cdot \underset{m\times n}{g^T}$$

$$\frac{\partial L_1}{\partial X} = \frac{\partial L}{\partial f} + \frac{\partial L}{h}$$

$$\frac{\partial L}{\partial f} = \frac{\partial L}{\partial e} \cdot g^T = \alpha \frac{\partial L}{\partial d} \cdot g^T = \alpha \frac{\partial L}{\partial c} g^T = \alpha \det(c)(c^{-1})^T \cdot \frac{\partial L}{\partial b} \cdot g^T$$

$$= \alpha \det(c)(c^{-1})^T \cdot \left(\frac{-D}{2b}\right) \cdot g^T$$

$$= \alpha \det(j+d)((j+d)^{-1})^T \cdot \frac{-D}{2\det c} \cdot X$$

$$= \left(\alpha \det\left(\frac{1}{\beta}I + \alpha XX^T\right)\underbrace{\left(\left(\frac{1}{\beta}I + \alpha XX^T\right)^{-1}\right)^T}_{m\times m} \cdot \frac{-D}{2\det\left(\frac{1}{\beta}I + \alpha XX^T\right)}\right) \cdot \underset{m\times n}{X}$$

$$\frac{\partial L}{\partial h} = \left(\frac{\partial L}{\partial g}\right)^T = \frac{\partial L}{\partial e}^T f = \left(\frac{-D\alpha}{2\det\left(\frac{1}{\beta}I + \alpha XX^T\right)} \det\left(\frac{1}{\beta}I + \alpha XX^T\right)\left(\left(\frac{1}{\beta}I + \alpha XX^T\right)^{-1}\right)^T\right)^T X$$

$$K = \alpha X X^T + \beta^{-1} I$$

$$\frac{dL_1}{dX} = \frac{dL_1}{dh} + \frac{dL_1}{df}$$

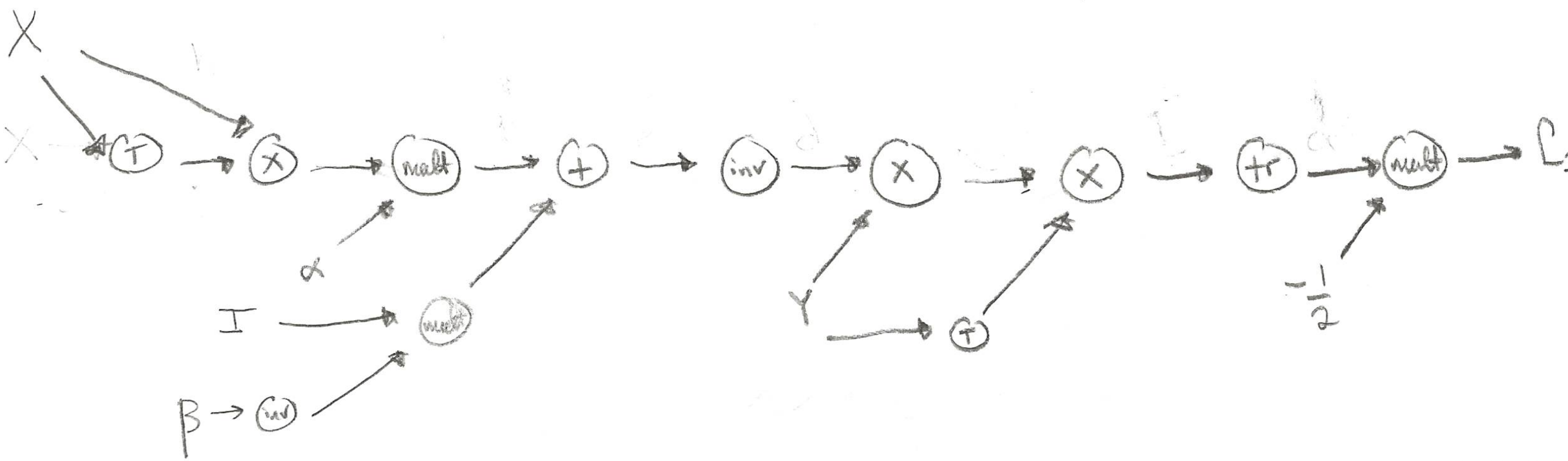
$$= \alpha \det(K)(K^{-1})^T \left(\frac{-D}{2\det(K)}\right) \cdot X + \left(\frac{-D\alpha}{2\det(K)} \det(K)\left((K^{-1})^T\right)^T\right) X$$

$$\frac{-D\alpha}{2} (K^{-1})^T X + \frac{-D\alpha}{2} (K^{-1}) X$$

$$\boxed{\frac{dL_1}{dX} = \frac{-D\alpha}{2}\left[(K^{-1})^T + (K^{-1})\right] X = -D\alpha K^{-1} X}$$

$K^T = K$

(c)



d. $\frac{\partial L}{\partial K} = -K^{-T} \frac{\partial L}{\partial K^{-1}} K^{-T}$ $\qquad \frac{\partial \operatorname{Tr}(XA)}{\partial X} = X^T$ $\quad \frac{\partial \operatorname{Tr}(K^{-1} YY^T m)}{\partial K^{-1}} = YY^T$

$$\frac{\partial L}{\partial X} = \underbrace{\frac{\partial K}{\partial X}}_{m\times n \times m \times m} \underbrace{\frac{\partial L}{\partial K}}_{m \times m} = \boxed{\alpha K^{-T} YY^T K^{-T} X = \frac{\partial L_2}{\partial X}.}$$

$$\frac{\partial K}{\partial X}\frac{\partial L}{\partial K} = \sum_{ij} \left(\frac{\partial K_{ij}}{\partial X}\right)\left(\frac{\partial L}{\partial K}\right)_{ij}$$

if $\frac{\partial L}{\partial K}$ is symmetric... train of thought

$$\underset{X}{\begin{bmatrix} x_1 & x_2 \\ x_3 & x_4 \end{bmatrix}} \underset{X^T}{\begin{bmatrix} x_1 & x_3 \\ x_2 & x_4 \end{bmatrix}} = \begin{bmatrix} x_1^2 + x_2^2 & x_1x_3 + x_2x_4 \\ x_1x_3 + x_2x_4 & x_3^2 + x_4^2 \end{bmatrix} \qquad \begin{bmatrix} a & b \\ b & d \end{bmatrix}$$

$$\begin{bmatrix} 2x_1 & 2x_2 \\ 0 & 0 \end{bmatrix} a + \begin{bmatrix} x_3 & x_4 \\ x_1 & x_2 \end{bmatrix} b + \begin{bmatrix} x_3 & x_4 \\ x_1 & x_2 \end{bmatrix} b + \begin{bmatrix} 0 & 0 \\ 2x_3 & 2x_4 \end{bmatrix} d = \begin{bmatrix} 2ax_1 + 2x_3b & 2ax_2 + 2bx_4 \\ 2x_3d + 2x_1b & 2x_4d + 2x_2b \end{bmatrix} = 2\begin{bmatrix} a & b \\ b & d \end{bmatrix}\begin{bmatrix} x_1 & x_2 \\ x_3 & x_4 \end{bmatrix}$$

one 2x2 slice is $\frac{\partial K_{ij}}{\partial X}$

e.

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

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

$$\frac{\partial L}{\partial X} = \frac{-D\alpha}{2}\left[(K^{-1})^T + (K^{-1})\right]X + \alpha K^{-T}YY^TK^{-T}X$$

$$= \alpha\left[\frac{-D}{2}\left[(K^{-T}) + K^{-1}\right] + K^{-T}YY^TK^{-T}\right]X$$

$$= \alpha\left[DK^{-1} + K^{-1}YY^TK^{-1}\right]X$$