Show: 5a) 3b)

$$\cancel{A^T H A \cdot A^T = A^{-1} \cdot H}$$

Show:

$$z^{i+1} = z^i + \left(A^T H(Az^i) A\right)^{-1} \cdot A^T \ell(Az^i)$$

Show:

$$z^{i+1} = A^{-1} x^{i} + \left(A^T H(x_i) A\right)^{-1} \cdot A^T \cdot \ell(x_i)$$

Prove $\overset{?}{=} A^{-1}\left(x^i + H(x_i)^{-1} \cdot \ell(x_i)\right)$

$\Rightarrow$ Show: $\left(A^T H(x_i) A\right)^{-1} \cdot A^T = A^{-1} \cdot H^{-1}$

$$A^{-1} \cdot H(x)^{-1} \cdot \cancel{(A^T)^{-1}} \cdot \cancel{A^T} = A^{-1} \cdot H^{-1}$$

$$\Rightarrow \boxed{A^{-1} H(x)^{-1} = A^{-1} \cdot H^{-1}}$$

5a) 2p.

$$\Rightarrow \nabla_{z_i z_j} g(z) = \sum_{k=1}^{n} A_{ki} \sum_{p=1}^{n} A_{pj} \cdot \frac{d f(Az)}{x_k x_p}$$ Hesse matrix

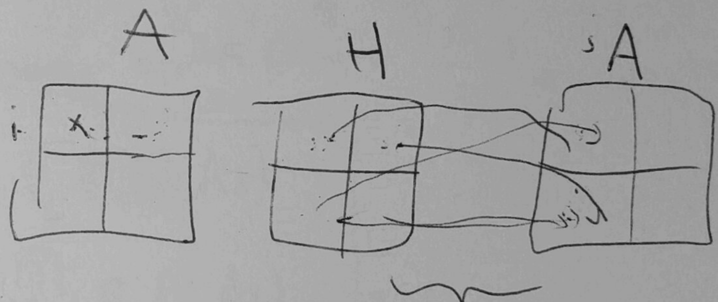
$$\cancel{\Rightarrow \nabla_{z_i z_j} g(z) = \sum_{k=1}^{n} A_{ki} \cdot A^T_{ij}\, d f(Az)}$$

$$\Rightarrow z_i z_j\, g(z) = \sum_{k=1}^{n} A_{ki} \cdot \left(d_x^2 f(Az)\right)_k \cdot A_{:,j}$$

$$= (A_{:,i})^T \Rightarrow z_i z_j\, g(z) = (A_{:,i})^T \cdot H(Az) \cdot A_{:,j}$$

$$\Rightarrow \boxed{\nabla_z^2 g(z) = A^T \; \nabla_x^2 f(Az) \cdot A}$$

$2 \times 2 \times 2 \times 2 =$



$$g_{z^2}(z)_{i,j} = \sum_{k=1}^{n} \sum_{p=1}^{n} A_{ki} \cdot A_{pj} \cdot H(Az)_{kp}$$

~~1~~ 5a) 5a) ~~1~~

$g(z) = f(Az)$

Chain rule of $\frac{dg(z)}{dz}$ where $g(z) = f(A z)$ ~~dam~~ and $\frac{df(x)}{dx}$ is known?

$\Rightarrow \nabla_z g(z) =$

$$\frac{\nabla_z g(z)}{dz_i} = \boxed{\sum_{k=1}^{n} \frac{df(Az)}{d(Az)_k} \cdot \frac{d(Az)_k}{z_i}}$$

$$= \sum_{k=1}^{n} \frac{df(Az)}{dx_k} \cdot A_{ki}$$

$$= A_{,i}^T \cdot \cancel{\frac{dl(Az)}{}} \nabla_\theta l(Az)$$

Transpose of ith column

$$\boxed{\Rightarrow \nabla_z g(z) = A^T \nabla_\theta l(\cancel{\theta}\, Az) \quad \text{or} \quad \cancel{A} \; A^T \nabla_x f(Az)}$$

Now, $\nabla_x^2 g(z)$:

$$\nabla_x^2 g(z)_{ij} = \frac{d\left(A^T \nabla_x f(Az)\right)_i}{dz_j} = \cancel{\sum_{k=1}^{n}}$$

$$= \frac{d(d_{x_i} f(Az))}{dz_j} = \sum_{k=1}^{n} \frac{d(\nabla_{x_i} f(Az))}{d(Az)_k} \cdot \frac{d(Az)_k}{dz_j}$$

5b)

~~sh~~ $g(z) = f(Az)$

$$\Rightarrow \nabla_z g(z) = \sum_{k=1}^{n} \frac{d(f(Az))}{d x_k} \cdot \frac{d(Az)}{d z_i}$$

$$= A^T \nabla_\theta l(Az)$$

$$\Rightarrow z^{i+1} = z^i + A^T \nabla_\theta l(A z_i)$$

$$x^{i+1} = x^i + \nabla_\theta l(x^i)$$

For 0 case:

$$x^1 = \nabla_\theta l(x^0)$$

$$z^1 = A^T \nabla_\theta l(0)$$

$$x^{i+1} = x^i + \nabla_\theta l(x^i)$$

$$z^{i+1} = z^i + A^T \nabla_\theta l(A z^i)$$

$A^{-1}(A^T \nabla_\theta l(0)$ is not $\nabla_\theta l(x^0)$

$\Rightarrow$ Now, with $z^i = A^{-1} x^i$:

$$A^{-1}(x^{i+1}) = A^{-1} x^i + A^{-1} \nabla_\theta l(x^i)$$

$$z^{i+1} = A^{-1} x^i + A^T \nabla_\theta l(A \cdot A^{-1} x^i)$$

but $z^{i+1} =$

$$= A^{-1} x^i + A^T \nabla_\theta l(x^i)$$

and A^{-1} not $A^T \Rightarrow$ not linear invariant!

5b)

gradient descent:

$$\theta_j := \theta_j - \sum_{i=1}^{m} \nabla_\theta \ell(x^i)$$

$$\theta_j := \theta_j + \alpha\left(y^{(i)} - h_\theta(x^{(i)})\right) \cdot x_j^{(i)}$$

Lin reg:

$$\theta_j := \theta_j + \sum_{i=1}^{m} \alpha\left(y^{(i)} - \theta^T x^i\right) \cdot x^i_j$$

* For θ_j^0: $\sum_{i=1}^{m} \alpha\left(y^{(i)} - \theta^T x^i\right) x^i_j$

One sample θ_j: $\alpha\left(y^{(i)} - 0\right) x_j$

and one dim: $\theta' = \alpha \cdot y \cdot x_j$

2 dim: θ_j

$$\theta'_1 = \alpha(y - 0)x_1 = \alpha \cdot (y) \cdot x_1$$

g(A