$$\|Ax-b\|^2 = (Ax-b)\cdot(Ax-b)$$

$Ax-b \rightarrow A_{ij}x_j - b_i \rightarrow$ Einstein's notation

$$(Ax-b)\cdot(Ax-b) = (A_{ij}x_j - b_i)(A_{ij}x_j - b_i)$$

e.i $\rightarrow a\cdot b = \sum_i a_i b_i = \underset{\text{repeated} = \sum_i}{a_i b_i}$

$$\underset{\arg x}{\min} \|Ax-b\|^2 \rightarrow \nabla_x(\|Ax-b\|^2) = 0$$

$$\frac{\partial}{\partial x_k}\left[(A_{ij}x_j - b_i)(A_{ij}x_j - b_i)\right] = 0$$

$$2(A_{ij}x_j - b_i)(A_{ik}) = 0$$

$$\underbrace{2A_{ik}A_{ij}x_j - 2A_{ik}b_i}_{F_k(x)=0} = 0 \equiv \overbrace{2\sum_i\sum_j A_{ik}A_{ij}x_j - 2\sum_i A_{ik}b_i = 0}$$

$F_k(x) = 0 \rightarrow$ Eqn system in variables $x_j \rightarrow$ Newton's Method

$$J(F_k(x))\Delta x = -F_k(x)$$

$$J(F_k(x)) = \frac{\partial F_k}{\partial x_l} \rightarrow J_{kl} = \frac{\partial F_k}{\partial x_l} = 2A_{ik}A_{il}$$

Newton's $\rightarrow \frac{\partial F_k}{\partial x_l}\Delta x_l = -F_k(x)$

$$\cancel{2}\cancel{A_{ik}}A_{il}\Delta x_l = -\cancel{2}\cancel{A_{ik}}A_{ij}x_j + \cancel{2}\cancel{A_{ik}}b_i$$

$$A_{il}\Delta x_l = -A_{ij}x_j + b_i \equiv A_{ij}\Delta x_j = -A_{ij}x_j + b_i$$

$A_{ij}(x_j + \Delta x_j) = b_i \rightarrow$ original system

To minimize using Newton's method is the same as to solve the original system!!!

Therefore SDG is just capricious!!!

Now the Hessian $\rightarrow H(f) = \frac{\partial^2 f}{\partial x_m \partial x_l} = H_{ml}$

$$f = \|Ax - b\|^2 = (A_{ij}x_j - b_i)(A_{ij}x_j - b_i)$$

$$\frac{\partial f}{\partial x_m} = 2A_{im}(A_{ij}x_j - b_i) = 2A_{im}A_{ij}x_j - 2A_{im}b_i$$

$$\frac{\partial^2 f}{\partial x_m \partial x_l} = 2A_{im}A_{il} \equiv H_{ml} = 2A_{im}A_{il} \Rightarrow$$

$$H_{kl} = 2A_{ik}A_{il}$$

$$H_{kl} = J_{kl}$$

<u>conclusion</u>: The hessian of the error is the same as the Jacobian of the minimization problem which was proved is the same as the original system solved by Newton's method

$$A^T A : A^T_{ik} A_{kj} = A_{ki} A_{kj}$$

$$\mathrm{diag}(A^T A) = \sum_k A_{ki} A_{kj} = \sum_k A_{ki}^2$$

$$A_i^T A_i = \begin{bmatrix} A_{i1} \\ A_{i2} \\ \vdots \\ A_{in} \end{bmatrix} \begin{bmatrix} A_{i1} & A_{i2} & \cdots \end{bmatrix} = \begin{pmatrix} A_{i1}^2 & & & \\ & A_{i2}^2 & & \\ & & A_{i3}^2 & \\ & & & \ddots \end{pmatrix}$$

$$E = \frac{1}{2} \|Ax - b\|^2$$

$$E = \frac{1}{2} (A_{ij} x_j - b_i)(A_{ij} x_j - b_i)$$

$$\frac{\partial E}{\partial x_k} = \text{grad} E = \frac{1}{2} (A_{ij} x_j - b_i) A_{ik} + \frac{1}{2} A_{ik} (A_{ij} x_j - b_i)$$

$$= A_{ik} (A_{ij} x_j - b_i) = A_{ik} e_i$$

$$= A^T (Ax - b)$$

$$= \sum_i A_{ik} e_i = \frac{\partial E}{\partial x_k}$$

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} a_{11} b_1 + a_{12} b_2 \\ a_{21} b_1 + a_{22} b_2 \end{pmatrix} = b_1 \begin{pmatrix} a_{11} \\ a_{21} \end{pmatrix} + b_2 \begin{pmatrix} a_{12} \\ a_{22} \end{pmatrix}$$

Gradient descent $\rightarrow$

$$x^{k+1} = x^k - \eta \nabla f(x^k)$$

Newton Direction $\rightarrow$

$$x^{k+1} = x^k + \delta x^k$$

$$H(x^k)\, \delta x^k = -\nabla f(x_k)$$

$$x^{k+1} = x^k - H^{-1} \nabla f(x^k)$$

Also the local minimization model changes, thus

GD $\rightarrow$ model $\rightarrow$ $f(x) \approx f(x^k) + \nabla f \cdot (x - x^k)$

Newton model $\rightarrow$

$$f(x) \approx f(x^k) + \nabla f(x^k)(x - x^k) + (x - x^k)^T H (x - x^k)$$

$$\frac{\partial}{\partial x_k} (A_{ij} x_j - b_i)^2 = 2 (A_{ij} x_j - b_i) A_{ik}$$

$$\frac{\partial^2}{\partial x_k \partial x_l} (A_{ij} x_j - b_i)^2 = 2 A_{ik} A_{il}$$

$$f = (a_1 x_1 + a_2 x_2 - b)^2$$

$$\frac{\partial f}{\partial x_1} = 2(a_1 x_1 + a_2 x_2 - b) a_1$$

$$\frac{\partial f}{\partial x_2} = 2(a_1 x_1 + a_2 x_2 - b) a_2$$

$$H = \begin{pmatrix} 2a_1^2 & 2a_1 a_2 \\ 2a_1 a_2 & 2a_2^2 \end{pmatrix}$$

$$\det H = 4a_1^2 a_2^2 - 4a_1^2 a_2^2 = 0 ??$$