#10q

(1) $$Df(x_1,x_2) = \begin{bmatrix} 2x_1x_2 & x_1^2 \\ x_2^3 & 3x_1x_2^2 \end{bmatrix}$$

$$Dg(y_1,y_2) = \begin{bmatrix} \cos(y_1y_2)\cdot y_2 & \cos(y_1y_2)\cdot y_1 \\ 1 & 1 \end{bmatrix}$$

$h = g \circ f$ so $D h(\vec{x}) = Dg(f(\vec{x}))\cdot Df(\vec{x})$

$$= \begin{bmatrix} \cos(x_1^3x_2^4)\,x_1x_2^3 & \cos(x_1^3x_2^4)\,x_1^2x_2 \\ 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} 2x_1x_2 & x_1^2 \\ x_2^3 & 3x_1x_2^2 \end{bmatrix} =$$

$$\begin{bmatrix} 3x_1^2x_2^4\cos(x_1^3x_2^4) & 4x_1^3x_2^3\cos(x_1^3x_2^4) \\ 2x_1x_2+x_2^3 & x_1^2+3x_1x_2^2 \end{bmatrix}$$

#10b

$$\begin{bmatrix} \cos(x_1^3 x_2^4)\, 2 x_1^2 x_2^4 + \cos(x_1^3 x_2^4)\, x_2^4 x_1^2 \quad 3x_1^2 x_2^4 \\ \cos(x_1^3 x_2^4)\, x_1^3 x_2^3 + \cos(x_1^3 x_2^4)\, 3 x_1^3 x_2^3 \quad 4 x_1^3 x_2^3 \\ \text{OK} \\ 2x_1 x_2 + x_2^3 \qquad x_1^2 + 3 x_1 x_2^2 \end{bmatrix}$$

(2) (a) $F(x, \eta) = \gamma_1 \sigma(\omega_{11} x_1 + \omega_{21} x_2 + b_1)$
$+ \gamma_2 \sigma(\omega_{12} x_1 + \omega_{22} x_2 + b_2)$

$$\nabla_\eta F = \left[\frac{\partial F}{\partial \omega_{11}}, \frac{\partial F}{\partial \omega_{21}}, \frac{\partial F}{\partial \omega_{12}}, \frac{\partial F}{\partial \omega_{22}}, \frac{\partial F}{\partial b_1}, \frac{\partial F}{\partial b_2}, \frac{\partial F}{\partial \gamma_1}, \frac{\partial F}{\partial \gamma_2}\right]$$

Letting $z_1 = \omega_{11} x_1 + \omega_{21} x_2 + b_1$
$z_2 = \omega_{12} x_1 + \omega_{22} x_2 + b_2$

$$\frac{\partial F}{\partial \omega_{11}} = \gamma_1 \sigma'(z_1) \cdot x_1, \quad \frac{\partial F}{\partial \omega_{21}} = \gamma_1 \sigma'(z_1)\, x_2$$

$$\frac{\partial F}{\partial \omega_{12}} = \gamma_2 \sigma'(z_2)\, x_1, \quad \frac{\partial F}{\partial \omega_{22}} = \gamma_2 \sigma'(z_2)\, x_2$$

$$\frac{\partial F}{\partial b_1} = \gamma_1 \sigma'(z_1), \quad \frac{\partial F}{\partial b_2} = \gamma_2 \sigma'(z_2)$$

$$\frac{\partial F}{\partial \gamma_1} = \sigma(z_1), \quad \frac{\partial F}{\partial \gamma_2} = \sigma(z_2)$$

3 (a) $\Phi(x) = \|Ax-b\|_2^2 = (Ax-b)^T(Ax-b)$

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

$- x^T A^T b + b^T b$

Now notice $b^T A x = (x^T A^T b)^T$ and they are both numbers, so $b^T A x = x^T A^T b$ so

$\Phi(x) = x^T A^T A x - 2b^T A x + b^T b$

as we computed in a previous homework

HW 4 #3

$\nabla \Phi(x) = 2A^T A x - 2b^T A$

and $H\Phi(x) = 2A^T A$

(b) $A = U \Sigma V^T$, $\Sigma = \text{diag}(\sigma_1, \cdot, \sigma_n)$

with $\sigma_n > 0$ since A is full rank.

so $A^T A = V \Sigma^T U^T U \Sigma V^T$

$= V \Sigma^2 V^{-1}$

#10d

so the eigenvalues of A^TA are $\sigma_1^2, \cdot, \sigma_n^2$

all >0, so A^TA is pos def

(c) Critical points are when

$$0 = \nabla\Phi(x) = 2A^TAx - 2b^TA$$

or when $A^TA\,x = b^TA$, But we know

A^TA is invertible, so the unique soln and thus

the unique crit pt is $x_0 = (A^TA)^{-1}b^TA$.

Now at that point, and every other point,

$$H(\Phi) = 2A^TA \quad \text{which is pos def.}$$

So by the 2nd deriv test, x_0 is a local

minimum. There are no other critical points

and thus no other loc. min, so x_0 is the

global min.

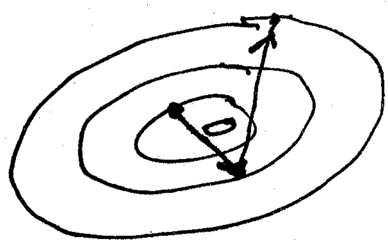
(3g) Up a a point, longer jumps get you

to the minimum with less steps.

However, if the magnitude of $h\nabla\Phi(x_0)$

#10e

is too large, you jump accross the critical point to a place where $h\nabla\Phi(x_{k+1})$ is even larger



and then you zig zag off to infinity.

So the ~~~~ choice of the value of h is crucial in using gradient descent and as mentioned in lecture there are many sophisticated tools for choosing and altering h during runtime.