Problem 1.

(a) From the visual, minimal value may be at $x=1$, $y=3$ since the gradient vector is $\vec{0}$ there.

(b) $\nabla f = \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{bmatrix}$

$= \begin{bmatrix} 2x-2 \\ 2y-6 \end{bmatrix}$

(3) $\nabla f = \vec{0}$

$\begin{bmatrix} 2x-2 \\ 2y-6 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$

$\Rightarrow x=1,\ y=3$

Problem 2.

$\theta^{t+1} = \theta^t - \alpha \nabla_{\vec{\theta}} L(\vec{\theta}, X, \vec{y})$

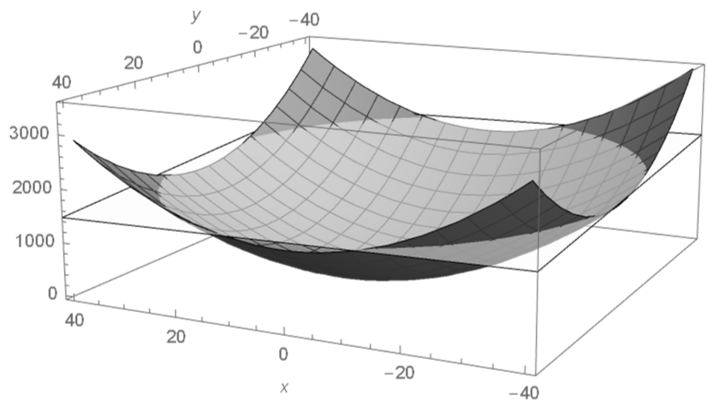
$= \theta^t - \alpha \nabla_{\vec{\theta}} \frac{1}{n} \sum (x_i^2 \theta^2 - \log y_i)$

$= \theta^t - \alpha \frac{1}{n} \sum (x_i^2 \cdot 2\theta)$

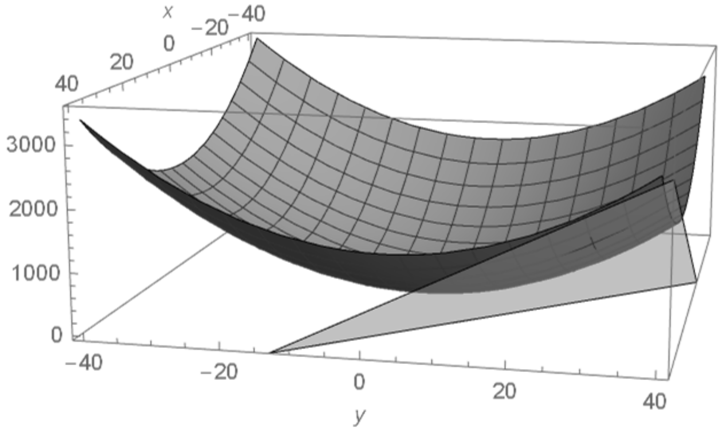
$= \theta^t - 2\alpha\theta \cdot \frac{\sum x_i^2}{n}$

Problem 3.

a. For a plane, all values either lie below or above the plane, so the function is convex.



b. For the tangential plane at point (30, 30, 1570) all values lie above the plane, so the function is convex.



c. Second derivative of f is $\nabla^2 f = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$ and is always positive wrt each variable, so the function is convex.

Problem 4.

(a). $\begin{bmatrix} A \\ B \end{bmatrix} = \nabla_{\vec{\theta}} L(\vec{\theta}, \vec{x}, \vec{y})$

$= \begin{bmatrix} \frac{\partial}{\partial \theta_0} (y_i - (0.5\theta_0 + \theta_0\theta_1 x_1 + \sin(\theta_1) \cdot x_2)) \\ \frac{\partial}{\partial \theta_1} (y_i - (0.5\theta_0 + \theta_0\theta_1 x_1 + \sin(\theta_1) x_2)) \end{bmatrix}$

$= \begin{bmatrix} -(0.5 + \theta_1 x_1) \\ -(\theta_0 x_1 + \cos(\theta_1) \cdot x_2) \end{bmatrix}$

$\therefore A = -(0.5 + \theta_1 x_1) = -0.5 - \theta_1$

$B = -(\theta_0 x_1 + \cos(\theta_1) \cdot x_2) = -\theta_0 - 2\cos\theta_1$

(b) $\vec{\theta}^{(t+1)} = \vec{\theta}^{(t)} - \alpha \nabla_{\vec{\theta}} L(\vec{\theta}, \vec{x}, \vec{y})$

$\vec{\theta}^{(1)} = \begin{bmatrix} 0.5\alpha \\ \alpha x_2 \end{bmatrix}$

$\therefore \theta_0^{(1)} = 0.5\alpha \quad \theta_1^{(1)} = \alpha x_2$

(c) $\vec{\theta}^{(t)}$ will converge close to zero when $t \to \infty$ or will fluctuate around zero if α is too large.