ECE 403
Assignment #2

Ashlynn Steeves
V00850631

All MATLAB code used to generate answers for this assignment can be found at

github.com/ashlynns/ECE403/tree/master/A2

1.4

a) $f(x_1,x_2) = x_1^2 - 2x_2\sin x_1 + 100$

$$\nabla f(x_1,x_2) = \begin{bmatrix} \partial f/\partial x_1 \\ \partial f/\partial x_2 \end{bmatrix} = \begin{bmatrix} 2x_1 - 2x_2\cos x_1 \\ -2\sin x_1 \end{bmatrix}$$

$$\nabla^2 f(x_1,x_2) = \begin{bmatrix} \partial^2 f/\partial x_1^2 & \partial^2 f/\partial x_1\partial x_2 \\ \partial^2 f/\partial x_1\partial x_2 & \partial^2 f/\partial x_2^2 \end{bmatrix} = \begin{bmatrix} 2+2x_2\sin x_1 & -2\cos x_1 \\ -2\cos x_1 & 0 \end{bmatrix}$$

b) $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T\mathbf{H}\mathbf{x} + \mathbf{x}^T\mathbf{b} + k$, $\mathbf{H} \in R^{N\times N}$, $\mathbf{b} \in R^{N\times 1}$, $k \in R^{1\times 1}$

Let $\mathbf{x} = [x_1, x_2, \dots x_n]^T$, $\mathbf{H} = \begin{bmatrix} \mathbf{h}_1 \\ \mathbf{h}_2 \\ \vdots \\ \mathbf{h}_n \end{bmatrix} = \begin{bmatrix} h_{1,1} & h_{1,2} & \cdots & h_{1,n} \\ h_{2,1} & h_{2,2} & \cdots & h_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ h_{n,1} & h_{n,2} & \cdots & h_{n,n} \end{bmatrix}$, $\mathbf{b} = [b_1, b_2, \dots, b_n]^T$

$$\nabla f(\mathbf{x}) = \nabla \tfrac{1}{2}\mathbf{x}^T\mathbf{H}\mathbf{x} + \nabla \mathbf{x}^T\mathbf{b} + \nabla k$$

$$= \tfrac{1}{2}\cdot \nabla \mathbf{x}^T[\mathbf{h}_1\mathbf{x}, \mathbf{h}_2\mathbf{x}, \dots, \mathbf{h}_n\mathbf{x}] + \nabla(x_1b_1 + x_2b_2 + \dots + x_nb_n) + 0$$

$$= \tfrac{1}{2}\cdot \nabla(x_1\mathbf{h}_1\mathbf{x} + x_2\mathbf{h}_2\mathbf{x} + \dots + x_n\mathbf{h}_n\mathbf{x}) + \mathbf{b}$$

$$= \tfrac{1}{2}\begin{bmatrix} d/dx_1(x_1\mathbf{h}_1\mathbf{x} + x_2\mathbf{h}_2\mathbf{x} + \dots + x_n\mathbf{h}_n\mathbf{x}) \\ d/dx_2(x_1\mathbf{h}_1\mathbf{x} + x_2\mathbf{h}_2\mathbf{x} + \dots + x_n\mathbf{h}_n\mathbf{x}) \\ \vdots \\ d/dx_n(x_1\mathbf{h}_1\mathbf{x} + x_2\mathbf{h}_2\mathbf{x} + \dots + x_n\mathbf{h}_n\mathbf{x}) \end{bmatrix} + \mathbf{b}$$

$$= \tfrac{1}{2}\begin{bmatrix} \mathbf{h}_1\mathbf{x} + (x_1h_{1,1} + x_2h_{2,1} + \dots + x_nh_{n,1}) \\ \mathbf{h}\,\mathbf{x} + (x_1h_{1,2} + x_2h_{2,2} + \dots + x_nh_{n,2}) \\ \mathbf{h}\,\mathbf{x} + (x_1h_{1,n} + x_2h_{2,n} + \dots + x_nh_{n,n}) \end{bmatrix} + \mathbf{b}$$

$$= \tfrac{1}{2}(\mathbf{H}x + \mathbf{H}^Tx) + \mathbf{b}$$

$$= \tfrac{1}{2}(\mathbf{H} + \mathbf{H}^T)x + \mathbf{b}$$

$$\nabla^2 f(x) = \nabla(\nabla^T f(x))$$

$$= \nabla(\tfrac{1}{2}(H+H^T)x + b)^T$$

$$= \nabla(\tfrac{1}{2}x^T(H+H^T) + b^T)$$

$$= \tfrac{1}{2}\nabla x^T(H+H^T) + 0$$

$$= \tfrac{1}{2}\nabla \begin{bmatrix} x_1h_{1,1} + x_2h_{2,1} + \dots + x_nh_{n,1} \\ x_1h_{1,2} + x_2h_{2,2} + \dots + x_nh_{n,2} \\ \\ x_1h_{1,n} + x_2h_{2,n} + \dots + x_nh_{n,n} \end{bmatrix}^T + \tfrac{1}{2}\nabla \begin{bmatrix} x_1h_{1,1} + x_2h_{1,2} + \dots + x_nh_{1,n} \\ x_1h_{2,1} + x_2h_{2,2} + \dots + x_nh_{2,n} \\ \\ x_1h_{n,1} + x_2h_{n,2} + \dots + x_nh_{n,n} \end{bmatrix}^T$$

$$= \tfrac{1}{2}\begin{bmatrix} h_{1,1} & h_{1,2} & \cdots & h_{1,n} \\ h_{2,1} & h_{2,2} & \cdots & h_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ h_{n,1} & h_{n,2} & \cdots & h_{n,n} \end{bmatrix} + \tfrac{1}{2}\begin{bmatrix} h_{1,1} & h_{2,1} & \cdots & h_{n,1} \\ h_{1,2} & h_{2,2} & \cdots & h_{n,2} \\ \vdots & \vdots & \ddots & \vdots \\ h_{n,1} & h_{n,2} & \cdots & h_{n,n} \end{bmatrix}$$

$$= \tfrac{1}{2}(H+H^T)$$

c) $f(\theta) = \tfrac{1}{P}\sum_{p=1}^{P}\log(1+e^{-\theta^T z_p})$, $\theta \in \mathbb{R}^{N\times 1}$, $\{z_p \text{ for } p=1,2,\dots,P\}$
$z_i \in \mathbb{R}^{N\times 1}$

$$\nabla f(\theta) = \nabla \tfrac{1}{P}\sum_{p=1}^{P}\log(1+e^{-\theta^T z_p})$$

$$= \tfrac{1}{P}\sum_{p=1}^{P}\nabla\log(1+e^{-\theta^T z_p})$$

$$= \tfrac{1}{P}\sum_{p=1}^{P}\begin{bmatrix} -(1+e^{-\theta^T z_p})^{-1} e^{-\theta^T z_p} z_{p,1} \\ -(1+e^{-\theta^T z_p})^{-1} e^{-\theta^T z_p} z_{p,2} \\ \\ -(1+e^{-\theta^T z_p})^{-1} e^{-\theta^T z_p} z_{p,P} \end{bmatrix}$$

$$\nabla f(\theta) = \frac{-1}{P}\sum_{p=1}^{P}\frac{e^{-\theta^T z_p}}{1+e^{-\theta^T z_p}} z_p = -\tfrac{1}{P}\sum_{p=1}^{P}\frac{1}{1+e^{\theta^T z_p}} z_p$$

$$\nabla^2 f(\theta) = \nabla\left(-\tfrac{1}{P}\sum_{p=1}^{P}\frac{1}{1+e^{\theta^T z_p}} z_p\right)^T = -\tfrac{1}{P}\sum_{p=1}^{P}\nabla\frac{1}{1+e^{\theta^T z_p}} z_p^T$$

$$= -\tfrac{1}{P}\sum_{p=1}^{P}\begin{bmatrix} -(1+e^{-\theta^T z_p})^{-2} e^{-\theta^T z_p} z_{p,1} \\ -(1+e^{-\theta^T z_p})^{-2} e^{-\theta^T z_p} z_{p,2} \\ \\ -(1+e^{-\theta^T z_p})^{-2} e^{-\theta^T z_p} z_{p,P} \end{bmatrix} \cdot z_p^T$$

$$\nabla^2 f(\theta) = \tfrac{1}{P}\sum_{p=1}^{P}\frac{e^{\theta^T z_p}}{(1+e^{\theta^T z_p})^2} z_p z_p^T$$

1.5 a) $f(x) = 2x_1^2 + x_2^2 - 2x_1x_2 + 2x_1^3 + x_1^4$

$$\nabla f(x) = \begin{bmatrix} 4x_1 - 2x_2 + 6x_1^2 + 4x_1^3 \\ 2x_2 - 2x_1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \rightarrow x_1^* = \begin{bmatrix} 0 \\ 0 \end{bmatrix},\ x_2^* = \begin{bmatrix} -0.5 \\ -0.5 \end{bmatrix},\ x_3^* = \begin{bmatrix} -1 \\ -1 \end{bmatrix}$$

$$\nabla^2 f(x) = \begin{bmatrix} 4 + 12x_1 + 12x_1^2 & -2 \\ -2 & 2 \end{bmatrix}$$

$$\nabla^2 f(x_1^*) = \begin{bmatrix} 4 & -2 \\ -2 & 2 \end{bmatrix} \rightarrow \lambda = \det\left(\begin{bmatrix} 4-\lambda & -2 \\ -2 & 2-\lambda \end{bmatrix}\right) = 0$$
$$= (4-\lambda)(2-\lambda) - 4 = 0$$
$$= \lambda^2 - 6\lambda + 4 = 0$$
$$\lambda = \begin{bmatrix} 5.24 \\ 0.76 \end{bmatrix}$$

Positive λ means $x_1^* = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ is a minimizer

$$\nabla^2 f(x_2^*) = \begin{bmatrix} 1 & -2 \\ -2 & 2 \end{bmatrix} \rightarrow \lambda = \det\left(\begin{bmatrix} 1-\lambda & -2 \\ -2 & 2-\lambda \end{bmatrix}\right) = 0$$
$$= (1-\lambda)(2-\lambda) - 4 = 0$$
$$= \lambda^2 - 3\lambda - 2 = 0$$
$$\lambda = \begin{bmatrix} 3.56 \\ -0.56 \end{bmatrix}$$

Mixed λ means $x_2^* = \begin{bmatrix} -0.5 \\ -0.5 \end{bmatrix}$ is a saddle point

$$\nabla^2 f(x_3^*) = \begin{bmatrix} 4 & -2 \\ -2 & 2 \end{bmatrix} \rightarrow$$ See λ calculations for $\nabla^2 f(x_1^*)$

Positive λ means $x_3^* = \begin{bmatrix} -1 \\ -1 \end{bmatrix}$ is a minimizer

b) $f(x) = x_1^2x_2^2 - 4x_1^2x_2 + 4x_1^2 + 2x_1x_2^2 + x_2^2 - 8x_1x_2 + 8x_1 - 4x_2$

$$\nabla f(x) = \begin{bmatrix} 2x_1x_2^2 - 8x_1x_2 + 8x_1 + 2x_2^2 - 8x_2 + 8 \\ 2x_1^2x_2 - 4x_1^2 + 4x_1x_2 + 2x_2 - 8x_1 - 4 \end{bmatrix}$$

$$= \begin{bmatrix} x_1(2x_2^2 - 8x_1 + 8) + (2x_2^2 - 8x_2 + 8) \\ 2(x_2(x_1 + 2x_1 + 1) - 2x_1^2 - 4x_1 - 2) \end{bmatrix} = \begin{bmatrix} 2(x_2^2 - 4x_1 + 4)(x_1 + 1) \\ 2(x_2 - 2)(x_1^2 - 2x_1 - 1) \end{bmatrix}$$

$$\nabla f(x) = \begin{bmatrix} 2(x_2 - 2)^2(x_1 + 1) \\ 2(x_2 - 2)(x_1 + 1)^2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \rightarrow$$

These eq's will both be zero at any pt where $x_1 = -1$ or $x_2 = 2$. Since they are lines with a continuious $\nabla f(x) = 0$, they will share a min/max behavior. All that to say we can test this using $x^* = \begin{bmatrix} -1 \\ 3 \end{bmatrix}$ (using $\begin{bmatrix} -1 \\ 2 \end{bmatrix}$ causes $\nabla^2 f(x^*)$ to zero out)

$$\nabla^2 f(x) = \begin{bmatrix} 2(x_2 - 2)^2 & 4(x_1+1)(x_2-2) \\ 4(x_1+1)(x_2-2) & 2(x_1 + 1)^2 \end{bmatrix}$$

$$\nabla^2 f(x^*) = \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix} \rightarrow \lambda = \det\left(\begin{bmatrix} 2-\lambda & 0 \\ 0 & -\lambda \end{bmatrix}\right) = 0$$
$$= (2-\lambda)(-\lambda)$$
$$= \lambda^2 - 2\lambda$$
$$\lambda = \begin{bmatrix} 2 \\ 0 \end{bmatrix} \rightarrow$$

Positive semi-definite λ means that the lines where $x_1 = -1$ or $x_2 = 2$ are minimizers

c) $f(x) = (x_1^2 - x_2)^2 + x_1^5$

$$\nabla f(x) = \begin{bmatrix} 4x_1 \cdot (x_1^2 - x_2) + 5x_1^4 \\ -2(x_1^2 - x_2) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \rightarrow x^* = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\nabla^2 f(x) = \begin{bmatrix} 4(3x_1^2 - x_2) + 20x_1^3 & -4x_1 \\ -4x_1 & 2 \end{bmatrix}$$

$$\nabla^2 f(x^*) = \begin{bmatrix} 0 & 0 \\ 0 & 2 \end{bmatrix} \rightarrow \lambda = \det\left(\begin{bmatrix} -\lambda & 0 \\ 0 & 2-\lambda \end{bmatrix}\right) = 0$$

$$= (-\lambda)(2-\lambda)$$

$$= \lambda^2 - 2\lambda$$

$$\lambda = \begin{bmatrix} 2 \\ 0 \end{bmatrix} \rightarrow$$ Positive semi-definite λ means that $x^* = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ is a minimizer

1.6 a) $f(x) = x_1^2 + \cosh(x_2)$

$$\nabla f(x) = \begin{bmatrix} 2x_1 \\ \sinh(x_2) \end{bmatrix} \qquad \nabla^2 f(x) = \begin{bmatrix} 2 & 0 \\ 0 & \cosh(x_2) \end{bmatrix}$$

$$\lambda = \det\left(\begin{bmatrix} 2-\lambda & 0 \\ 0 & \cosh(x_2)-\lambda \end{bmatrix}\right) = 0$$

$$= (2-\lambda)(\cosh(x_2)-\lambda)$$

$$= \lambda^2 - \lambda(\cosh(x_2)+2) + 2\cosh(x_2)$$

$$\lambda = \begin{bmatrix} 2 \\ \cosh(x_2) \end{bmatrix}$$

$\cosh(x_2)$ will always return a positive value $\therefore$ λ is positive definite and $f(x)$ is convex

b) $f(x) = x_1^2 + 2x_2^2 + 2x_3^2 + x_4^2 - x_1x_2 + x_1x_3 - 2x_2x_4 + x_1x_4$

$$\nabla f(x) = \begin{bmatrix} 2x_1 - x_2 + x_3 + x_4 \\ 4x_2 - x_1 - 2x_4 \\ 4x_3 + x_1 \\ 2x_4 - 2x_2 + x_1 \end{bmatrix} \qquad \nabla^2 f(x) = \begin{bmatrix} 2 & -1 & 1 & 1 \\ -1 & 4 & 0 & -2 \\ 1 & 0 & 4 & 0 \\ 1 & -2 & 0 & 2 \end{bmatrix}$$

From MATLAB, the eigenvalues of $\nabla^2 f(x)$ and $\nabla^2 -f(x)$ are $\lambda = \begin{bmatrix} 0.6099 \\ 1.3281 \\ 4.2427 \\ 5.8192 \end{bmatrix}$ and $\lambda = \begin{bmatrix} -5.8192 \\ -4.2427 \\ -1.3281 \\ -0.6099 \end{bmatrix}$ respectivley.

Since the set of eigenvalues corresponding to $\nabla^2 f(x)$ is positive definite, a subset of positive semi-definite, $f(x)$ is convex.

c) $f(x) = x_1^2 - 2x_2^2 - 2x_3^2 + x_4^2 - x_1x_2 + x_1x_3 - 2x_2x_4 + x_1x_4$

$$\nabla f(x) = \begin{bmatrix} 2x_1 - x_2 + x_3 + x_4 \\ -4x_2 - x_1 - 2x_4 \\ -4x_3 + x_1 \\ 2x_4 - 2x_2 + x_1 \end{bmatrix} \qquad \nabla^2 f(x) = \begin{bmatrix} 2 & -1 & 1 & 1 \\ -1 & -4 & 0 & -2 \\ 1 & 0 & -4 & 0 \\ 1 & -2 & 0 & 2 \end{bmatrix}$$

From MATLAB, the eigenvalues of $\nabla^2 f(x)$ and $\nabla^2 -f(x)$ are $\lambda = \begin{bmatrix} -4.6939 \\ -4.1453 \\ 1.1797 \\ 3.6595 \end{bmatrix}$ and $\lambda = \begin{bmatrix} -3.6595 \\ -1.1797 \\ 4.1453 \\ 4.6939 \end{bmatrix}$ respectivley.

Since neither set of eigenvalues is non-negative, which corresponds to a positive semi definite hessian, $f(x)$ is neither convex nor concave. //

1.7 $\quad f(x) = (x_1 + x_2)^2 + [2(x_1^2 + x_2^2 - 1) - \frac{1}{3}]^2$

a) $$\nabla f(x) = \begin{bmatrix} 2(x_1 + x_2) + 8x_1(2x_1^2 + 2x_2^2 - \frac{7}{3}) \\ 2(x_1 + x_2) + 8x_2(2x_1^2 + 2x_2^2 - \frac{7}{3}) \end{bmatrix}$$

$$2(x_1 + x_2) + 8x_1(2x_1^2 + 2x_2^2 - \tfrac{7}{3}) = 2(x_1 + x_2) + 8x_2(2x_1^2 + 2x_2^2 - \tfrac{7}{3})$$

$$8(x_1 - x_2)(2x_1^2 + 2x_2^2 - \tfrac{7}{3}) = 0$$

If $2x_1^2 + 2x_2^2 - \frac{7}{3} = 0$:

$$2(x_1 + x_2) = 0$$

so, $x_1 = x_2$

If $x_1 = x_2$:

$$4x_1 + 8x_1(4x_1^2 - \tfrac{7}{3}) = 0$$
$$x_1 + 2x_1(4x_1^2 - \tfrac{7}{3}) = 0$$
$$8x_1^3 - \tfrac{14}{3}x_1 + x_1 = 0$$
$$x_1(8x_1^2 - \tfrac{11}{3}x_1) = 0$$
$$x_1(x_1^2 - \tfrac{11}{24}x_1) = 0$$

so, $x_1 = x_2 = 0$ or $\pm\sqrt{11/24}$

$$x_1^* = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad x_2^* = \begin{bmatrix} \sqrt{11/24} \\ \sqrt{11/24} \end{bmatrix}, \quad x_3^* = \begin{bmatrix} -\sqrt{11/24} \\ -\sqrt{11/24} \end{bmatrix}$$

b) $\nabla^2 f(x) = \begin{bmatrix} 48x_1^2 + 16x_2^2 - 50/3 & 2 + 32x_1x_2 \\ 2 + 32x_1x_2 & 48x_2^2 + 16x_1^2 - 50/3 \end{bmatrix}$

$$\nabla^2 f(x_1^*) = \begin{bmatrix} -50/3 & 2 \\ 2 & -50/3 \end{bmatrix}$$

$$\lambda = \det\left(\begin{bmatrix} -50/3 - \lambda & 2 \\ 2 & -50/3 - \lambda \end{bmatrix}\right) = 0$$

$$= (-50/3 - \lambda)(-50/3 - \lambda) - 4$$

$$= \lambda^2 + 100/3\,\lambda + 2464/9$$

$$\lambda = \begin{bmatrix} -14.67 \\ -18.67 \end{bmatrix} \rightarrow x_1^* \text{ is a maximizer}$$

$$\nabla^2 f(x_2^*) = \begin{bmatrix} 38/3 & 50/3 \\ 50/3 & 38/3 \end{bmatrix}$$

$$\lambda = \det\left(\begin{bmatrix} 38/3 - \lambda & 50/3 \\ 50/3 & 38/3 - \lambda \end{bmatrix}\right) = 0$$

$$= (38/3 - \lambda)(38/3 - \lambda) - 2500/9$$

$$= \lambda^2 + 76/3\,\lambda - 352/3$$

$$\lambda = \begin{bmatrix} 4 \\ -29.3 \end{bmatrix} \rightarrow x_2^* \text{ is a saddle point}$$

$$\nabla^2 f(x_2^*) = \begin{bmatrix} 38/3 & 50/3 \\ 50/3 & 38/3 \end{bmatrix}$$

↳ see calculations for x_2^*, x_3^* is a saddle point