Assignment 0

Problem 1

a. $f(x) = x_1^2 + 2x_1x_2 + 3x_2^2 + 2x_1 - 3x_2 + e^{x_1}$

$$\nabla f(x) = \begin{bmatrix} 2x_1 + 2x_2 + 2 + e^{x_1} \\ 2x_1 + 6x_2 - 3 \end{bmatrix}$$

$$Hf(x) = \begin{bmatrix} 2+e^{x} & 2 \\ 2 & 6 \end{bmatrix}$$

Since e^x is always positive and increasing Then $Hf(x) > 0$ resulting in $f(x)$ being strictly convex.

b. $g(x) = (\|x\|^2 - 4)^2$ Let $\|x\| = \sqrt{x_1^2 + x_2^2}$

$g(x) = (x_1^2 + x_2^2 - 4)^2$

$$\nabla g(x) = \begin{bmatrix} 2(x_1^2 + x_2^2 - 4)(2x_1) \\ 2(x_1^2 + x_2^2 - 4)(2x_2) \end{bmatrix} = \begin{bmatrix} 4x_1^3 + 4x_1x_2^2 - 16x_1 \\ 4x_1^2x_2 + 4x_1^3 - 16x_2 \end{bmatrix}$$

$$Hg(x) = \begin{bmatrix} 12x_1^2 + 4x_2^2 - 16 & 8x_1x_2 \\ 8x_1x_2 + 12x_1^2 & 4x_1^2 - 16 \end{bmatrix}$$ → Can be negative so NOT Convex.

c. $h(x) = -\log(x_1 x_2)$

$$\nabla h(x) = \begin{bmatrix} -\frac{1}{x_1x_2} x_2 \\ -\frac{1}{x_1x_2} x_1 \end{bmatrix} = \begin{bmatrix} -\frac{1}{x_1} \\ -\frac{1}{x_2} \end{bmatrix}$$

$$Hh(x) = \begin{bmatrix} \frac{1}{x_1^2} & 0 \\ 0 & \frac{1}{x_2^2} \end{bmatrix} \geq 0$$ So $h(x)$ is convex

Problem 2. $f(x) = (4x_1^2 - x_2)^2$ $\quad g(x) = x_1^2 + 4x_1x_2 + x_2^2 + x_1 - x_2$

a. $\nabla f(x) = \begin{bmatrix} 2(4x_1^2 - x_2)(8x_1) \\ -2(4x_1^2 - x_2) \end{bmatrix} = 0 = \begin{matrix} (16x_1)(4x_1^2 - x_2) \\ (-8x_1^2 + 2x_2) \end{matrix}$

$4x_1^2 - x_2 = 0 \longrightarrow \boxed{x_2 = 4x_1^2} \leftarrow$ all stationary points

$4x_1^2 - x_2 = 0$

$Hf(x) = \begin{bmatrix} 2(8x_1)^2 + 16(4x_1^2 - x_2) & -16x_1 \\ -2(8x_1) & 2 \end{bmatrix}$

~~So non-strict global minimizers.~~

Sub. $x_2 = 4x_1^2$

$= \begin{bmatrix} 128x_1^2 & -16x_1 \\ -16x_1 & 2 \end{bmatrix}$

The smallest value occurs when $x_1, x_2 = 0$. So

$Hf(x) = \begin{bmatrix} 0 & 0 \\ 0 & 2 \end{bmatrix}$ so So all points along line are non-strict - global min

b. $\nabla g(x) = \cancel{x_1^2 + 4x_1x_2 + x_2^2 + x_1 - x_2}$

$$\begin{bmatrix} 2x_1 + 4x_2 + 1 \\ 4x_1 + 2x_2 - 1 \end{bmatrix} = 0$$

$2x_1 + 4x_2 + 1 = 0$

$x_1 = -2x_2 - 1/2$

$4x_2 + 2(-2x_2 - 1/2) - 1 = 0$

$4x_2 - 4x_2 - 1 - 1 = 0$

$x_2 = -1/2$

$x_1 = 1/2$

$(1/2, -1/2)$

$$Hg(x) = \begin{bmatrix} 2 & 4 \\ 4 & 2 \end{bmatrix}$$

$\lambda = -2, 6$ Since $\pm$ it is a saddle point.

c. $\cancel{\lim_{\|x\|} (4x_1^2 - x_2)^2} =$

both g and f are not coercive, solve graphically.

$$\frac{(4x_1^2 - x_2)^2 \, x_1}{\sqrt{x_1^2 + x_2^2}}$$

$$\frac{(4x_1^2 - x_2)^2 \, x_2}{\sqrt{x_1^2 + x_2^2}}$$

$$\frac{(16x_1^4 - 8x_1^2x_2 + x_2^2)x_1}{(x_1^2 + x_2^2)^{1/2}}$$

$$\frac{16x_1^4x_2 - 8x_1^2x_2^2 + x_2^3}{\sqrt{x_1^2 + x_2^2}}$$

$$\frac{16x_1^5 - 8x_1^3x_2 + x_2^2x_1}{(x_1^2 + x_2^2)^{1/2}}$$

$$\cancel{\begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}} \qquad \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

3 $A \in \mathbb{R}^{n \times n}$ is a symmetric matrix $\quad R_A(x) = \dfrac{x^T A x}{\|x\|^2} \quad \forall x \neq 0$

Spectral decomposition Theorem:

1. All $\lambda \in \mathbb{R}$
2. There are n mutually orthogonal unit vectors $u_1 \ldots u_n$ corresponding to n eigenvalues $\lambda_1 \ldots \lambda_n$. Therefore the matrix A can be decomposed in the form $A = U \Lambda U^T \rightarrow U^T A U = \Lambda$
3. A can be represented as: $A = \sum_{i=1}^{n} \lambda_i u_i u_i^T$

$$R_A(x) = \frac{x^T A x}{\|x\|^2} = \frac{(Uy)^T A (Uy)}{\|x\|^2} \overset{\#2 \text{ S.D.T.}}{=} \frac{y^T \Lambda y}{(Uy)^T (Uy)} = \frac{y^T \Lambda y}{y^T y}$$

$$= \frac{\sum_{i=1}^{n} \lambda_i y_i^2}{\|y\|^2}$$

Let $n = 2$ for example Then the expression can be written as.

$$\frac{\lambda_1 y_1^2 + \lambda_2 y_2^2}{\|y\|^2}$$

If we assume λ_2 is max and λ_1 is minimum, Then the equation above is a weighted average that makes the $R_A(x)$ between λ_1 and λ_2 without loss in generality

$$\therefore \quad \lambda_{min}(A) \leq R_A(x) \leq \lambda_{max}(A) \qquad \forall x \neq 0.$$

$h(\alpha x + (1-\alpha)y) \leq \alpha h(x) + (1-\alpha) h(y)$

4. $f(x) = g(x) + h(x)$, $g(x) = \|y - Ax\|_2^2$ $h(x) = \|x\|_\infty$

Where $\|x\|_2^2 = x^T x$, $\|x\|_\infty = \max\{|x_i|\}$, A is orthogonal

a.

$\lim_{\|x\| \to \infty} g(x) = \infty$ b/c as $\|x\| \uparrow$ $g(x)$ will $\uparrow$ So it is coercive

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

$$y^T y - y^T A x - x^T A^T y + x^T A^T A x$$
$$= \|y\|_2^2 - 2y^T A x + \|x\|_2^2$$

So as $\lim_{\|x\| \to \infty} g(x) = \infty$ So coercive

b.

$h(x) = \|x\|_\infty$ ~~Coercive~~ Coercive. iff $\sqrt{x_1^2 + x_2^2} \to \infty$
$|x_i| \to$ either x_1 or x_2

So $f(x)$ is coercive.

b. $\sum_{i=1}^{k} f_i(x)$

Let $k=2$, Then assume f_1 is strongly convex and f_2 is convex.

Then $g(x) = f_1(x) + f_2(x)$ ~~$= g(\alpha x_1 + (1-\alpha)x_2) - f(\alpha x_1 + (1-\alpha)x_2)$~~

$h(\alpha x_1 + (1-\alpha)x_2) = f_1(\alpha x_1 + (1-\alpha)x_2) + f_2(\alpha x_1 + (1-\alpha)x_2)$

$\leq \alpha f_1(x_1) + (1-\alpha) f_2(x_2) - \frac{M}{2}\alpha(1-\alpha)\|x-y\|_2^2$

Theorem 6 b. Lecture 3.

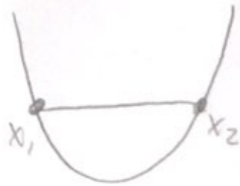
$+ \alpha f_1(x_1) + (1-\alpha) f_2(x_2)$

$= \alpha h(x) + (1-\alpha) h(x) - \frac{M}{2}\alpha(1-\alpha)\|x-y\|^2$

$\therefore$ $h(x)$ is strongly convex

c. all norms are convex

~~$g(x)$~~ $g(\alpha x_1 + (1-\alpha)x_2) =$



$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \end{bmatrix}$

4c. $g(x) = \|y - Ax\|_2^2$

$g(x) = (y-Ax)^T(y-Ax) = \|y\|_2^2 - 2y^TAx + \|x\|_2^2$

$\nabla g(x) = -2y^TA + 2x$

$Hg(x) = 2 \rightarrow$ Convex

$\begin{bmatrix} x_1 & x_2 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$

$x_1^2 + x_2^2$

$2x$

$h(x) = \|x\|_\infty = \max\limits_{i=\{1,\dots,n\}} \{|x_i|\}$

$\nabla h(x) = C$ $\qquad Hg(x) = 0 \quad \geq 0$

~~convex.~~

$$\|\alpha x_1 + (1-\alpha)x_2\|_\infty \leq \|\alpha x_1\| + \|(1-\alpha)x_2\|$$
$$= \alpha\|x_1\| + (1-\alpha)\|x_2\| \therefore \text{Convex}$$

So $f(x)$ is also convex

5. $f(x) = 4x_1^2 + 2x_1x_2 + 2x_2^2$

a. $\cancel{4x_1^3 + 2x_1^2 x_2 + 2x_2^2} \quad \cancel{4x_1^2 x_2 + 2x_1 x_2^2 + 2x_2^3}$

$\nabla f(x) = \begin{bmatrix} 8x_1 + 2x_2 \\ 2x_1 + 4x_2 \end{bmatrix} \qquad Hf(x) = \begin{pmatrix} 8 & 2 \\ 2 & 4 \end{pmatrix}$

$f(x)$ is coercive b/c $Hf(x)$ ↑ always

b. $8x_1 + 2x_2 = 0 \rightarrow x_2 = -4x_1$

$2x_1 + 4x_2 = 0 \rightarrow x_1 = -2x_2$

$x_1 = x_2 = 0$ - minimizer

c. plot / code attached

d.

$\gamma = 0.0025$ did not reach min, but converges

$\gamma = 0.025$ converges, to $\sim 10^{-5}$

$\gamma = 0.05$ Does not converge

e. $\gamma = 0.2265$, 0.025 is better because it got closer over the same iterations

The obj. has exponential decay vs. linear