SOLUTIONS HW 2

1 Problem 1

1. Let us consider $f(x) = -x^2$ and $\alpha = -1$, then $\mathcal{S} = (-\infty, -1] \cup [1, +\infty)$. The $\mathcal{S}$ is not convex, because although $-1, 1 \in \mathcal{S}$, we have $(-1+1)/2 = 0 \notin \mathcal{S}$.

2. By triangle inequality, we have $\|tx + (1-t)y\|_p \leq \|tx\|_p + \|(1-t)y\|_p = t\|x\|_p + (1-t)\|y\|_p$. Hence the ℓ_p norm $\|\cdot\|_p$ is a convex function, and its sublevel set is convex.

3. We observe that f can be rewritten as $f(x) = \frac{1}{2}x^T Q x + p^T x + r$ with Q being given as

$$Q = \begin{bmatrix} 4 & 0 & 2 \\ 0 & 4 & 3 \\ 2 & 3 & 1 \end{bmatrix}$$

Therefore, we can directly get

$$\nabla^2 f = \begin{bmatrix} 4 & 0 & 2 \\ 0 & 4 & 3 \\ 2 & 3 & 1 \end{bmatrix}, \ \forall x$$

Since $\det([4]) = 4 > 0$, $\det \begin{pmatrix} 4 & 0 \\ 0 & 4 \end{pmatrix} = 16 > 0$ and $\det(\nabla^2 f) = -36 < 0$, we know $\nabla^2 f$ is not PSD. Since $\det(-[4]) = -4 < 0$, $\det \begin{pmatrix} -4 & 0 \\ 0 & -4 \end{pmatrix} = 16 > 0$ and $\det(-\nabla^2 f) = 36 > 0$, we know $\nabla^2 f$ is not NSD. Therefore, $\nabla^2 f$ is indefinite. We can conclude that f is neither convex nor concave.

4. Yes. The function $f(x) = x\log(x)$, $x > 0$ is convex since $f''(x) = \frac{1}{x} > 0$. We also prove that the function $g(x, y) = x\log(x) + y\log(y)$ is convex. Indeed, let us fix x_1, x_2, y_1, $y_2 \in \mathbb{R}^+$ and $a \in [0, 1]$, then

$$\begin{aligned} g(ax_1 + (1-a)x_2, ay_1 + (1-a)y_2) &= f(ax_1 + (1-a)x_2) + f(ay_1 + (1-a)y_2) \\ &\leq af(x_1) + (1-a)f(x_2) + af(y_1) + (1-a)f(y_2) \\ &= ag(x_1, y_1) + (1-a)g(x_2, y_2) \end{aligned} \tag{1}$$

As a result the set $\mathcal{S} \equiv \{(x_1, x_2) : x_1,\ x_2 > 0, \quad g(x_1, x_2) \leq 4\}$ is convex.

5. Let us fix x_1, $x_2 \in \mathbb{R}^n$ and $a \in [0, 1]$. By concavity of g it holds $g(ax_1+(1-a)x_2) \geq ag(x_1)+(1-a)g(x_2)$. In order to prove that $f \circ g$ is concave, we proceed as follows

$$\begin{aligned} f(g(ax_1 + (1-a)x_2)) &\geq f(ag(x_1) + (1-a)g(x_2)), \quad \text{by concavity of g, \& the fact f is increasing} \\ &\geq af(g(x_1)) + (1-a)f(g(x_2)), \quad \text{by concavity of f} \end{aligned} \tag{2}$$

Hence, $f \circ g$ is concave.

2 Problem 2

1. The function f does not have maximum over $\mathbb{R}^3$ because $f(x_1, 0, 0) = 2x_1^2 - 2x_1 + 5$ is not bounded. The function f has a unique minimum. Indeed,

$$\nabla f = [4x_1 - 2, 4x_2 + 2x_3 - 2, 2x_3 + 2x_2 - 2]^T \tag{3}$$

and $\nabla f(x) = 0 \Rightarrow (x_1, x_2, x_3) = (0.5, 0, 1)$. Since,

$$\nabla^2 f = \begin{pmatrix} 4 & 0 & 0 \\ 0 & 4 & 2 \\ 0 & 2 & 2 \end{pmatrix} = \begin{pmatrix} 4 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 \\ 0 & 2 & 2 \\ 0 & 2 & 2 \end{pmatrix} \tag{4}$$

is PD, we conclude the result.

2. Since Q is PD we have $\nabla f(x) = Qx$. We consider $g(a_k) = f(x_k - a_k\, Q\, x_k)$ and we minimize g

$$g(a_k) = f((I - a_k\, Q)\, x_k) = \frac{1}{2} x_k^T Q x_k - (x_k^T Q^2 x_k) a_k + \frac{1}{2}(x_k^T Q^3 x_k) a_k^2 \tag{5}$$

Hence $g(a_k)$ is minimized when $a_k = \frac{x_k^T Q^2 x_k}{x_k^T Q^3 x_k}$

3. Let us consider matrix A whose i^{th} row is a_i and the column vector $b = (b_1, .., b_n)^T$, then

$$\begin{aligned} f(x) &= \frac{1}{n}(Ax - b)^T (Ax - b) + \frac{\lambda}{2} x^T I x \\ &= \frac{1}{n}(x^T A^T A x + b^T b - b^T A x - x^T A^T b) + \frac{\lambda}{2} x^T I x \end{aligned} \tag{6}$$

We have

$$\nabla f = \frac{1}{n}(2A^T A x - 2A^T b) + \lambda x = \left(\frac{2}{n} A^T A + \lambda I\right) x - \frac{2}{n} A^T b \tag{7}$$

and $\nabla f = 0 \Rightarrow x^* = \left(A^T A + \frac{n}{2}\lambda I\right)^{-1} A^T b$. Also, $\nabla^2 f = \frac{2}{n} A^T A + \lambda I$ is PD because $x^T \nabla^2 x = \frac{2}{n}(Ax)^T (Ax) + \lambda x^T x > 0$ for all $x \neq 0$. Hence, the optimal solution x^* is unique. It is worth mentioning that $A^T A = \sum_{i=1}^n a_i a_i^T$ and $A^T b = \sum_{i=1}^n a_i^T b_i$.

3 Problem 3

We must find the minimum m such that

$$f(x_k + \beta^m \tilde{\alpha} d_k) \leq f(x_k) + \sigma \beta^m \tilde{\alpha} \nabla f^T d_k \tag{8}$$

where $\nabla f = [4x_1, 8x_2^3]^T$, and since we apply steepest decent we choose $d_k = -\nabla f$. Hence, by substitution we obtain

$$f(1 - 0.5^m 4, 0) = 2(1 - 0.5^m 4)^2 \leq 2 - 0.80.5^m \tag{9}$$

and the minimum m that satisfies the inequality is $m = 2$, which implies that $a_k = \tilde{\alpha}\beta^m = 1 \cdot 0.5^2 = 0.25$.

4 Problem 4

We have

$$\begin{aligned} f(x_k) - f(x_{k+1}) &\geq (\nabla f(x_k))^T\, \alpha\, D\, \nabla f(x_k) - \frac{L}{2} \|\alpha\, D\, \nabla f(x_k)\|_2^2 \\ &\geq \alpha \left(\lambda_{min} - \frac{L}{2} \alpha \lambda_{max}^2\right) \|\nabla f(x_k)\|^2 \end{aligned} \tag{10}$$

We know $\lambda_{min} - \frac{L}{2}\alpha\lambda_{max}^2 > 0$. We observe that

$$\alpha \left(\lambda_{min} - \frac{L}{2} \alpha \lambda_{max}^2\right) \sum_{k=0}^{n} \|\nabla f(x_k)\|^2 \leq f(x_0) - f(x_{n+1}) \leq f(x_0) - f_{min} \tag{11}$$

As a result for all $n \in \mathbb{N}$

$$\sum_{k=0}^{n} \|\nabla f(x_k)\|^2 \leq \frac{f(x_0) - f_{min}}{\alpha \left(\lambda_{min} - \frac{L}{2} \alpha \lambda_{max}^2\right)} \tag{12}$$

which implies that as $n \to \infty$ the series converges and as a result $\lim_{n \to \infty} \nabla f(x_n) = 0$.