SOLUTIONS MIDTERM

1 Problem 1

1. We choose $x_1 = x_2 = t$ and we obtain $f(t,t) = 5t^2 - 5t + 4$ and as a result $f(t,t) \to +\infty$ as $t \to +\infty$. We choose $x_1 = -x_2 = t$ and we obtain $f(t,-t) = -t^2 + 15t + 4$ and as a result $f(t,-t) \to -\infty$ as $t \to -\infty$.

2. The function f takes the form

$$f(x_1, x_2) = \frac{1}{2} x^T \begin{pmatrix} 2 & 3 \\ 3 & 2 \end{pmatrix} x + (-5, 10)x + 4 \tag{1}$$

   For the matrix Q we have $Q(1,1) = 2 > 0$ and $\det \begin{pmatrix} 2 & 3 \\ 3 & 2 \end{pmatrix} = -5 < 0$ which implies that Q is not PSD. Hence, since $\nabla^2 f = Q$ we conclude that f is not convex.

3. We have

$$\nabla f = (2x_1 + 3x_2 + 5, 2x_2 + 3x_1 - 10)^T \tag{2}$$

   Hence, $\nabla f = 0 \Rightarrow (x_1, x_2) = (8, -7)$, which means that $(8, -7)$ is a stationary point. Furthermore, the eigenvalues of Q is -1 and 5, which implies Q is neither PSD nor NSD. Hence, $(8, -7)$ is nor a local min neither a local max, but it is a saddle point.

2 Problem 2

1. We need to find the smallest m such that

$$f(x_k + \beta^m \alpha_0 \, d_k) \le f(x_k) + \sigma \, \beta^m \alpha_0 \, (\nabla f(x_k, y_k))^T d_k \tag{3}$$

   where $\nabla f(x_k, y_k) = (6x_k, 4y_k^3)^T$ and $d_k = -\nabla f(x_k, y_k)$. The minimum m that satisfies the inequality is $m = 2$. Hence, $a_0 = 1$, $a_1 = 0.25$, $(x_1, y_1) = (0.625, 0)$ and $f(x_1, y_1) = 1.171$.

2. By chain rule we have

$$\begin{aligned} \frac{\partial f}{\partial \omega^{(1)}} &= 2(y - \omega^{(2)} \sigma(\omega^{(1)} x))(-\omega^{(2)} \sigma'(\omega^{(1)} x))x \\ \frac{\partial f}{\partial \omega^{(2)}} &= 2(y - \omega^{(2)} \sigma(\omega^{(1)} x))(-\sigma(\omega^{(1)} x)) \end{aligned} \tag{4}$$

3 Problem 3

1. Note that f has a unique global minimum at $x^* = 0$, $\nabla f(x) = 6x^5$, and $\nabla^2 f(x) = 30x^4$. Then for $x_k \neq 0$:

$$x_{k+1} = x_k - \frac{\alpha(6x_k^5)}{30x_k^4} = (1 - \frac{\alpha}{5})x_k.$$

   Therefore, as long as $|1 - \frac{\alpha}{5}| < 1$, x_k converges to $x^* = 0$ as $k \to \infty$. The range of α can be found using $|1 - \frac{\alpha}{5}| < 1 \Rightarrow 0 < \alpha < 10$. For this range of α and any $x_0 \in \mathbb{R}$, we can show

$$x_k = (1 - \frac{\alpha}{5})^k x_0,$$

   hence x_k converges to 0 geometrically, i.e., the method converges "linearly".

2. Since $f(x) = -\cos(x)$, we have $\nabla f(x) = \sin(x)$ and $\nabla^2 f(x) = \cos(x)$. The Newton's method with $\alpha = 1$ becomes: $x_{k+1} = x_k - \frac{\sin(x_k)}{\cos(x_k)} = x_k - \tan(x_k)$. Therefore, we have:

$$\begin{aligned} x_1 &= x_0 - \tan(x_0) \\ x_2 &= x_1 - \tan(x_1) = x_0 - \tan(x_0) - \tan(x_0 - \tan(x_0)). \end{aligned}$$

$x_2 = x_0 \Rightarrow \tan(-x_0) = \tan(x_0 - \tan(x_0))$. Therefore it suffices to show that $\exists\, x_0 \in [\frac{\pi}{4}, \frac{\pi}{2})$ such that $-x_0 = x_0 - \tan(x_0) \Rightarrow 2x_0 = \tan(x_0)$. Let $h(x) = 2x - \tan(x), x \in [\frac{\pi}{4}, \frac{\pi}{2})$. Then $h(x)$ is continuous and $h(\frac{\pi}{4}) = \frac{\pi}{2} - 1 < 0$, $h(\frac{\pi}{4})$ tends to $+\infty$. By intermediate value theorem, there exists $x_0 \in [\frac{\pi}{4}, \frac{\pi}{2})$ such that $h(x_0) = 0$. For this x_0, the Newton's methods does not converge since the iterates are oscillating between x_0 and x_1.

4 Problem 4

1. False. $\nabla^2 f(x) = 30x^4$ implies $\nabla^2 f(0) = 0$, by definition, f is not strongly convex.
2. True. From lecture 7, if the step size is $\alpha = \frac{1}{L}$, condition number $\kappa = \frac{L}{m}$, then the convergence rate $\rho = 2(1 - \frac{1}{\kappa})$. Clearly the larger the condition number is, the slower the convergence is.
3. False. It is not necessary that the gradient at x_0 towards the exact solution. For example, let $f(x) = \frac{1}{2}x^\top Qx + x^\top b$ where $Q = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}$ and $b = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$. Clearly we have $x^* = \begin{pmatrix} -1/2 \\ 1 \end{pmatrix}$. If we start with $x_0 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$, by using exact line search, the step size $\alpha = \arg\min f(x_0 - \alpha \nabla f(x_0)) = 10/19$. Hence $x_1 = x_0 - \alpha \nabla f(x_0) = \begin{pmatrix} -11/19 \\ 28/19 \end{pmatrix} \neq x^*$.
4. True. If $x_1 - y_1, x_2 - y_2 \in C_3$, then $\lambda(x_1 - y_1) + (1-\lambda)(x_2 - y_2) = (\lambda x_1 + (1-\lambda)x_2) - (\lambda y_1 + (1-\lambda)y_2) \in C_3$.
5. False. It could be a saddle point.
6. False. Consider $f(x) = \max\{0, -x^2 + 1\}$, which is a non-convex function. However it does not admit any local minimum which is not global.