Optimization Method - Assignment 3

Q1.

1. For the Hessian matrix $H = \nabla^2 f(x^0) = \begin{bmatrix} 1 & 5 \\ 5 & 4 \end{bmatrix}$, $\det(H) = -21 < 0$. When $\beta_0 = 1$, $\det(\nabla^2 f(x^0) + \beta_0 I) = \det\left(\begin{bmatrix} 2 & 5 \\ 5 & 5 \end{bmatrix}\right) = -15 < 0$. So when $\beta_0 = 1$, it is still a nonpositive definite matrix. When $\beta_0 = 3$, $\det(\nabla^2 f(x^0) + \beta_0 I) = \det\left(\begin{bmatrix} 4 & 5 \\ 5 & 7 \end{bmatrix}\right) = 3 > 0$.
2. If $\beta_0 = 1$, for d_0, it should satisfy $d_0 = -[\nabla^2 f(x^0) + \beta_0 I]^{-1} \nabla f(x_0)$.
   So $d_0 = [-0.3333,\ 0.0667]^T$ but not $[3,\ 2]^T$.

Q2.

We know that $B_0 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, s_0 = \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix}, y_0 = \begin{bmatrix} 3 \\ 6 \\ 9 \end{bmatrix}$.

Then $B_1 = B_0 + \frac{(y_0 - B_0 s_0)(y_0 - B_0 s_0)^T}{(y_0 - B_0 s_0)^T s_0} s_0 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 2 \\ 0 & 2 & 5 \end{bmatrix}$.

Then we should verify that B_1 satisfies the quasi-Newton equation which is: $B_{k+1} s_k = y_k$. Since:

$$B_{k+1} s_k = B_k s_k + \frac{(y_k - B_k s_k)(y_k - B_k s_k)^T}{(y_k - B_k s_k)^T s_k} s_k = B_k s_k + (y_k - B_k s_k) = y_k$$

Here $B_1 s_0 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 2 \\ 0 & 2 & 5 \end{bmatrix} * \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \\ 6 \\ 9 \end{bmatrix}$ which means B_1 satisfies the quasi-Newton equation.

Q3.

We know that in the BFGS for, $s_k = x^{k+1} - x^k$, $y_k = \nabla f(x^{k+1}) - \nabla f(x^k)$. Thus:

$$y_k^T s_k = [\nabla f(x^{k+1}) - \nabla f(x^k)]^T (x^{k+1} - x^k)$$

We know that $x^{k+1} = x^k + \alpha_k p_k$, thus:

$$y_k^T s_k = [\nabla f(x^{k+1}) - \nabla f(x^k)]^T (\alpha_k p_k) = a_k [p_k^T \nabla f(x^{k+1}) - p_k^T f(x_k)]$$

And we know that $\nabla f(x^k)^T p_k = -\nabla f(x^k)^T (B_k^{-1}) \nabla f(x^k) < 0$ as all matrices B_k are the positive definite. So the $\nabla f(x^k)^T p_k$ must be negative.

According to the Wolfe condition in this question:

$$|p^T \nabla f(x_k + \alpha p)| \le \eta |p^T \nabla f(x_k)| = -\eta p^T \nabla f(x_k)$$

And if $\nabla f(x^{k+1}) \neq 0$, then:

$$a_k [p_k^T \nabla f(x^{k+1}) - p_k^T f(x_k)] \ge a_k \left[p_k^T \nabla f(x^{k+1}) - \frac{1}{\eta} |p^T \nabla f(x_k + \alpha p)| \right]$$

$$= a_k \left[p_k^T \nabla f(x^{k+1}) - \frac{1}{\eta} \left| p^T \nabla f(x^{k+1}) \right| \right] > 0$$

If $p_k^T \nabla f(x^{k+1}) = 0$, then

$$a_k [p_k^T \nabla f(x^{k+1}) - p_k^T f(x_k)] = -\alpha_k p_k^T f(x_k) > 0$$

To conclude, $y_k^T s_k$ must be larger than 0.

Q4.

(a) The Lagrange function of this problem is:

$$L(x,\lambda) = x_1^2 + x_1^2x_3^2 + 2x_1x_2 + x_2^4 + 8x_2 + \lambda(2x_1 + 5x_2 + x_3 - 3)$$

Thus:

$$\nabla L(x,\lambda) = \begin{bmatrix} (2+2x_3^2)x_1 + 2x_2 + 2\lambda \\ 2x_1 + 4x_2^3 + 8 + 5\lambda \\ 2x_1^2x_3 + \lambda \\ 2x_1 + 5x_2 + x_3 - 3 \end{bmatrix}$$

When $x = [0,0,2]^T$, $\nabla L(x,\lambda) = \begin{bmatrix} 2\lambda \\ 8+5\lambda \\ \lambda \\ -1 \end{bmatrix} \neq 0$, so this point is not a stationary point.

When $x = [0,0,3]^T$, $\nabla L(x,\lambda) = \begin{bmatrix} 2\lambda \\ 8+5\lambda \\ \lambda \\ 0 \end{bmatrix}$. In this case $\lambda = 0$ and $8 + 5\lambda$ cannot satisfy at the same time, so this point is not a stationary point.

When $x = [1,0,1]^T$, $\nabla L(x,\lambda) = \begin{bmatrix} 4+2\lambda \\ 10+5\lambda \\ 2+\lambda \\ 0 \end{bmatrix}$. In this case if $\lambda = -2$, then $\nabla L(x,\lambda) = 0$. So this point is not a stationary point.

(b) From question (a), we know that the stationary point is $x^* = [1,0,1]^T$. Then the Hessian matrix of the Lagrange function is:

$$\nabla^2 L(x,\lambda) = \begin{bmatrix} 2+2x_3 & 2 & 4x_1x_3 \\ 2 & 12x_2^2 & 0 \\ 4x_1x_3 & 0 & 2x_1^2 \end{bmatrix}, \nabla^2 L(x^*,\lambda^*) = \begin{bmatrix} 4 & 2 & 4 \\ 2 & 0 & 0 \\ 4 & 0 & 2 \end{bmatrix}$$

However, $\det\left(\nabla^2 L(x^*,\lambda^*)\right) < 0$ which means that it is not a positive definite matrix. So we cannot make conclusion immediately.

Since the constraint is active, we consider $y = (y_1, y_2, y_3) \neq 0$.

$$[y_1, y_2, y_3] \begin{bmatrix} 4 & 2 & 4 \\ 2 & 0 & 0 \\ 4 & 0 & 2 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = 4y_1^2 + 2y_3^2 + 4y_1y_2 + 8y_1y_3$$

So we cannot make sure that the result of $y^T\nabla^2 L(x,\lambda)y \geq 0$ which means that $[1,0,1]^T$ is a saddle point.

Q5.

The question is:

$$\min f(x) = x_1^3 - x_2^3 - 2x_1^2 - x_1 + x_2$$
$$s.t. -x_1 - 2x_2 \geq -2$$
$$x_1 \geq 0$$
$$x_2 \geq 0$$

And we should change the constrain to the form in $g_i(x) \leq 0$:

$$s.t.\ x_1 + 2x_2 - 2 \leq 0$$
$$-x_1 \leq 0$$
$$-x_2 \leq 0$$

Then the Lagrange function is:

$$L(x,\mu) = x_1^3 - x_2^3 - 2x_1^2 - x_1 + x_2 + \mu_1(x_1 + 2x_2 - 2) + \mu_2(-x_1) + \mu_3(-x_2)$$

We have:

$$\nabla L(x,\mu) = \begin{bmatrix} 3x_1^2 - 4x_1 - 1 + \mu_1 - \mu_2 \\ -3x_2^2 + 1 + 2\mu_1 - \mu_3 \end{bmatrix} = 0$$

$$\nabla^2 L(x,\mu) = \begin{bmatrix} 6x_1 - 4 & 0 \\ 0 & -6x_2 \end{bmatrix}$$

And the KKT conditions for this problem is:

1. $3x_1^2 - 4x_1 - 1 + \mu_1 - \mu_2 = 0$
2. $-3x_2^2 + 1 + 2\mu_1 - \mu_3 = 0$
3. $\mu_1(x_1 + 2x_2 - 2) = 0$
4. $\mu_2(-x_1) = 0$
5. $\mu_3(-x_2) = 0$
6. $\mu_1, \mu_2, \mu_3 \geq 0$

Then we should perform classification discussion:

Case1: μ_1 is active and others are not active. Then we have:

$$x_1 + 2x_2 - 2 = 0; \quad -x_1 < 0; \quad -x_2 < 0$$

Then we can get:

$$x_1 = 0\ (Rejected), x_2 = 1, \mu_1 = 1$$

$$Or\ x_2 = \frac{5}{27}, \mu_1 < 0\ (Rejected)$$

Case2: μ_2 is active and others are not active. Then we have:

$$x_1 + 2x_2 - 2 < 0; \quad -x_1 = 0; \quad -x_2 < 0$$

Then we can get:

$$x_1 = 0, x_2 = \frac{\sqrt{3}}{3}, \mu_2 = -1$$

μ_2 should be larger than 0, so this case is rejected.

Case3: μ_2 is active and others are not active. Then we have:

$$x_1 + 2x_2 - 2 < 0; \quad -x_1 < 0; \quad -x_2 = 0$$

Then we can get:

$$x_1 = \frac{2+\sqrt{7}}{3} (Accepted); \quad x_2 = 0; \quad \mu_3 = 1$$

So $[\frac{2+\sqrt{7}}{3}, 0]$ is a KKT point.

Q6.

We know that f, g are the convex function and $\bar{x}$ satisfies the first order necessary optimality conditions. Thus, we have:

$$\nabla L(x,\mu) = \nabla f(\bar{x}) + \mu \nabla g(\bar{x}) = 0$$

$$\mu g(\bar{x}) = 0$$

$$\mu \geq 0$$

Hence $f(x), g(x)$ are convex, for other points:

$$f(x) - f(\bar{x}) \geq \nabla f(\bar{x})^T (x - \bar{x})$$
$$g(x) - g(\bar{x}) \geq \nabla g(\bar{x})^T (x - \bar{x}) \Rightarrow -\mu\big(g(x) - g(\bar{x})\big) \leq -\mu \nabla g(\bar{x})^T (x - \bar{x})$$

Thus,

$$f(x) - f(\bar{x}) \geq \nabla f(\bar{x})^T (x - \bar{x}) = -\mu \nabla g(\bar{x})^T (x - \bar{x}) \geq -\mu\big(g(x) - g(\bar{x})\big)$$

And we know that $\mu g(\bar{x}) = 0$:

$$f(x) - f(\bar{x}) \geq -\mu g(x) \geq 0$$

Therefore, $\bar{x}$ is the global minimizer.

Q7.

We know that the problem is:

$$\min f(x) = x_1^2 + x_2^2$$
$$s.t.\ g(x) = x_1 - 1 \geq 0$$

And we apply the inverse barrier function:

$$\beta(x, \mu) = x_1^2 + x_2^2 + \mu(\frac{1}{x_1 - 1})$$

We have:

$$\nabla \beta(x, \mu) = \begin{bmatrix} 2x_1 + \mu\left(\frac{1}{(x_1 - 1)^2}\right) \\ 2x_2 \end{bmatrix} = 0$$

Then we can get:

$$2x_1(x_1 - 1)^2 = \mu;\ \ x_2 = 0$$

When $\mu \to 0$:

$$x_1 = 1, x_2 = 0\ because\ x_1 - 1 \geq 0$$

Thus, the minimizer is $[1, 0]^T$

Q8.

We know that the problem is

$$\min f(x) = x$$
$$s.t.\ h(x) = x - 1 = 0$$

Then we apply the penalty method:

$$\pi(x, \rho) = x + \frac{1}{2}\rho(x - 1)^2$$

We have:

$$\nabla \pi(x, \rho) = 1 + \rho(x - 1) = 0$$

So we can get:

$$x = \frac{\rho - 1}{\rho}$$
$$h(x) = \frac{\rho - 1}{\rho} - 1 = -\frac{1}{\rho}$$
$$\lambda = \rho\, h(x) = -1$$

When $\rho = 1, x^* = 0$ and $\lambda = -1$;

When $\rho = 10, x^* = 0.9$ and $\lambda = -1$;

When $\rho = 100, x^* = 0.99$ and $\lambda = -1$;

When $\rho = 10000, x^* = 0.999$ and $\lambda = -1$;

Then using the augmented Lagrangian method:

$$A(x, \lambda, \rho) = x_k + \lambda_k(x_k - 1) + \frac{1}{2}\rho_k(x_k - 1)^2$$

We have:

$$\nabla A(x, \lambda, \rho) = 1 + \lambda_k + \rho_k(x_k - 1) = 0$$

$$x_{k+1} = \frac{p_k - \lambda_k - 1}{\rho_k}$$

Then the updated multiplier is:

$$\lambda_{k+1} = \lambda_k + \rho\, h(x_{k+1}) = \lambda_k - \rho_k\left(\frac{p_k - \lambda_k - 1}{\rho_k} - 1\right) = -1$$

So the λ^* has a limit -1 and the $x^* = 1$.