1 Problem 1.1

1)

According to the question,

$$g(x_1, x_2) = x_1^2 + x_2^2 - 1.$$

Following the equation on the book, the penalty function takes the form

$$p(x_1, x_2) = \mu(max\{0, (x_1^2 + x_2^2 - 1)\}^2),$$

with

$$f_p(\mathbf{x}; \mu) = f(\mathbf{x}) + p(\mathbf{x}; \mu),$$

Formally written as one function, we have the following,

$$f_p(x_1, x_2; \mu) = \begin{cases} (x_1-1)^2 + 2(x_2-2)^2 + \mu(x_1^2 + x_2^2 - 1)^2 & \text{if } x_1^2 + x_2^2 \geq 1, \\ (x_1-1)^2 + 2(x_2-2)^2 & \text{otherwise.} \end{cases}$$

2)

The gradient of f_p is defined as follows,

$$\nabla f_p(\mathbf{x}; \mu) = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \frac{\partial f}{\partial \mu} \right).$$

When $\mathbf{x} = \{[x_1, x_2] \mid x_1^2 + x_2^2 \geq 1\}$,

$$\frac{\partial f}{\partial x_1} = 2(x_1 - 1) + 4\mu x_1(x_1^2 + x_2^2 - 1),$$
$$\frac{\partial f}{\partial x_2} = 4(x_2 - 2) + 4\mu x_2(x_1^2 + x_2^2 - 1),$$

otherwise

$$\frac{\partial f}{\partial x_1} = 2x_1 - 2,$$
$$\frac{\partial f}{\partial x_2} = 4x_2 - 8,$$

3)

For $\mu = 0$, $f_p(x_1, x_2) = (x_1 - 1)^2 + 2(x_2 - 2)^2 = x_1^2 + 2x_2^2 - 2x_1 - 8x_2 + 9$.
Investigate the Hessian as follows,

$$H = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_1} \\ \frac{\partial^2 f}{\partial x_1 \partial x_2} & \frac{\partial^2 f}{\partial x_2^2} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}.$$

The eigenvalues $\lambda_1 = 1, \lambda_2 = 2$. The function is positive definite. Thus f_p is a convex function, so the unconstrained minimum of the function will be at the point $\nabla f(\mathbf{x}) = 0$.

$$2x_1 - 2 = 0, \quad 4x_2 - 8 = 0; \quad x_1 = 1, \quad x_2 = 2.$$

5)

The parameter values for running the program are set as follows,

$$\eta = 0.0001, \quad \mu = [1\ 10\ 100\ 1000], \quad T = 10^{-6}, \quad x_0 = [1, 2].$$

The results of this program are shown in Table 1. The values of x_1 and x_2 is specified with 4 decimal precision.

μ	x_1	x_2
1	0.4341	1.2100
10	0.3316	0.9955
100	0.3140	0.9552
1000	0.3116	0.9508

Table 1: The Results

Figure 1 below shows the values of x_2 and x_1 plotted as a function of µ. It is clear that the values of **x** at $\mu = 100$ is nearly the same as that at $\mu = 1000$, which means **x** almost approach the actual minimum of the function.

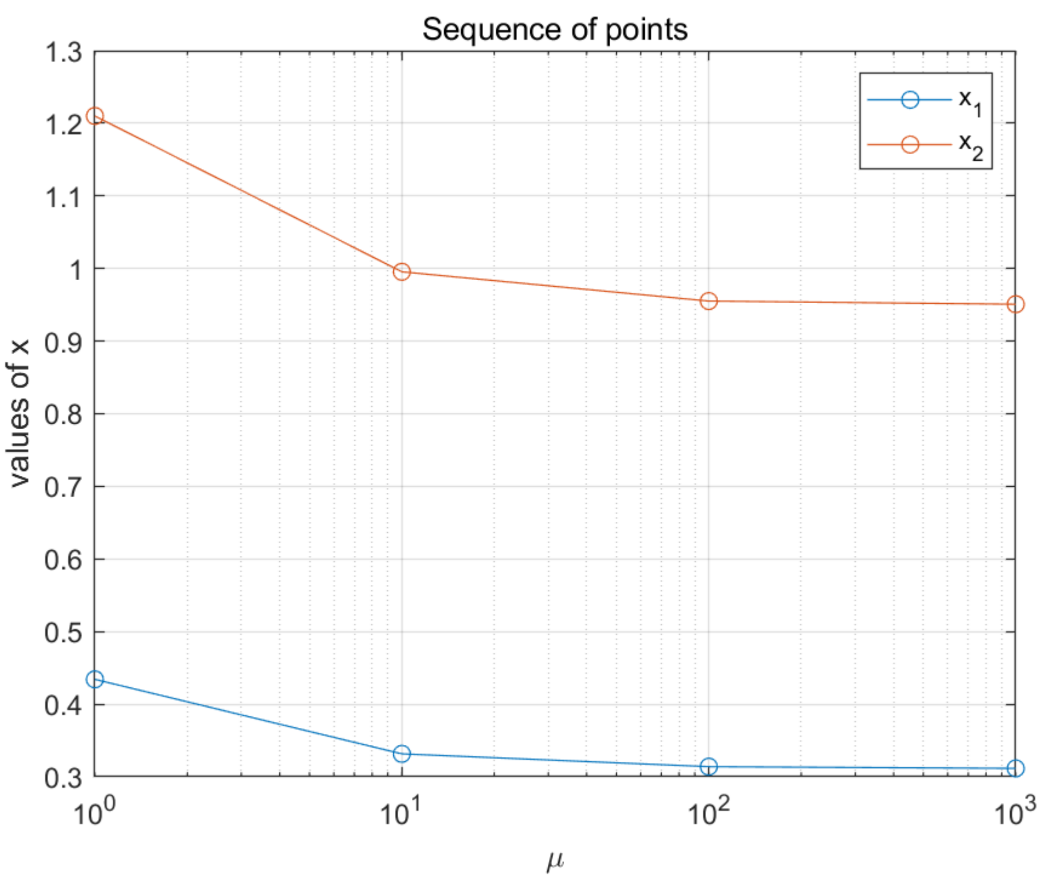

Figure 1: The convergent results.

2 Problem 1.2

For the points in the inner of set S,

$$\begin{aligned} f(x_1, x_2) &= 4x_1^2 + 2x_2^3; \\ \nabla f &= \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}\right) = (8x_1, 6x_2^2). \end{aligned}$$

When $\nabla f = 0$, there is a stationary point$(0, 0)$.
For the points on the boundary of set S, define the function L as follows,

$$\begin{aligned} L(x_1, x_2, \lambda) &= f(x_1, x_2) + \lambda h(x_1, x_2) \\ &= 4x_1^2 + 2x_2^3 + \lambda(x_1^2 + x_2^2 - 4). \end{aligned}$$

Setting the gradient of L to 0,

$$\begin{aligned} \frac{\partial L}{\partial x_1} &= 8x_1 + 2\lambda x_1 = 0, \\ \frac{\partial L}{\partial x_2} &= 6x_2^2 + 2\lambda x_2 = 0, \\ \frac{\partial L}{\partial \lambda} &= x_1^2 + x_2^2 - 4 = 0. \end{aligned}$$

Do the calculation,

$$\begin{aligned} x_1(8 + 2\lambda) = 0, &\quad x_1 = 0 \text{ or } \lambda = -4, \\ 2x_2(3x_2 + \lambda) = 0, &\quad x_2 = 0 \text{ or } \lambda = -3x_2, \\ x_1^2 + x_2^2 = 4, & \end{aligned}$$

If $x_1 = 0, \quad x_2 = \pm 2$.
If $\lambda = -4, \quad x_2 = -\frac{\lambda}{3} = \frac{4}{3},\ x_1 = \pm\frac{2\sqrt{5}}{3}$.
If $x_2 = 0,\ x_1 = \pm 2$.

So there are seven critical points: $(0, 0),\ (0, 2),\ (0, -2),\ (\frac{2\sqrt{5}}{3}, \frac{4}{3}),\ (-\frac{2\sqrt{5}}{3}, \frac{4}{3}),\ (2, 0),\ (-2, 0)$.
Do the calculation,

$$f(0, 0) = 0,\ f(0, 2) = 16,\ f(0, -2) = -16,\ f(\pm\frac{2\sqrt{5}}{3}, \frac{4}{3}) = \frac{368}{27}.\ f(\pm 2, 0) = 16.$$

So the max value is 16 at $(0, -2),\ (2, 0),\ (-2, 0)$ and the min value is -16 at $(0, -2)$.

3 Problem 1.3

a)

The table 2 below contains 10 sets of selected parameters and the results.(In this table Size means the size of tournament).

Size	$p_{tournament}$	p_{cross}	$p_{mutation}$	nGenerations	Fitness	x_1	x_2	$g(x_1, x_2)$
2	0.75	0.8	0.02	2000	1.0000	3.0000	0.5000	0.0
2	0.3	0,8	0.02	2000	0.9976	2.8905	0.4687	0.0024
2	0.4	0.8	0.02	2000	0.9999	3.0017	0.5005	0.0001
2	0.5	0.8	0.02	2000	1.0000	2.9995	0.4998	0.0
2	0.5	0.8	0.02	1000	0.9983	2.9687	0.4922	0.0017
3	0.5	0.8	0.02	1000	1.0000	2.9999	0.4999	0.0
2	0.5	0.8	0.02	1500	0.9997	3.0105	0.5029	0.0003
2	0.5	0.7	0.02	2000	0.9999	3.0017	0.5004	0.0001
2	0.5	0.5	0.02	2000	1.0000	2.9998	0.4999	0.0
10	0.5	0.5	0.03	2000	0.9976	2.8906	0.4687	0.0024

Table 2: Selected parameters and the results.

b)

Table 3 and table 4 show median fitness values for different mutation probabilities. And Figure 2 in Page 5 is plotted the datas in these tables. From Figure 2, the median fitness values increases dramatically at points betweem $p_{mut} = 0$ and $p_{mut} = 0.02$. After the point at $p_{mut} = 0.1$, the decrease of median fitness values appears. This results just satisfy the estimate discussed on the lecture: the fitness value always performs well at the point of $\frac{1}{\text{Number of Genes}}$, which equals to 0.02 at this program.

p_{mut}	0	0.005	0.01	0.02
Median	0.9904	0.9978	0.9998	0.9999

Table 3: Median performance($p_{mut} \leq 0,02$)

p_{mut}	0.03	0.05	0.1	0.2	0.25	0.5
Median	0.9999	0.9999	0.9999	0.9996	0.9994	0.9985

Table 4: Median performance ($p_{mut} > 0,02$)

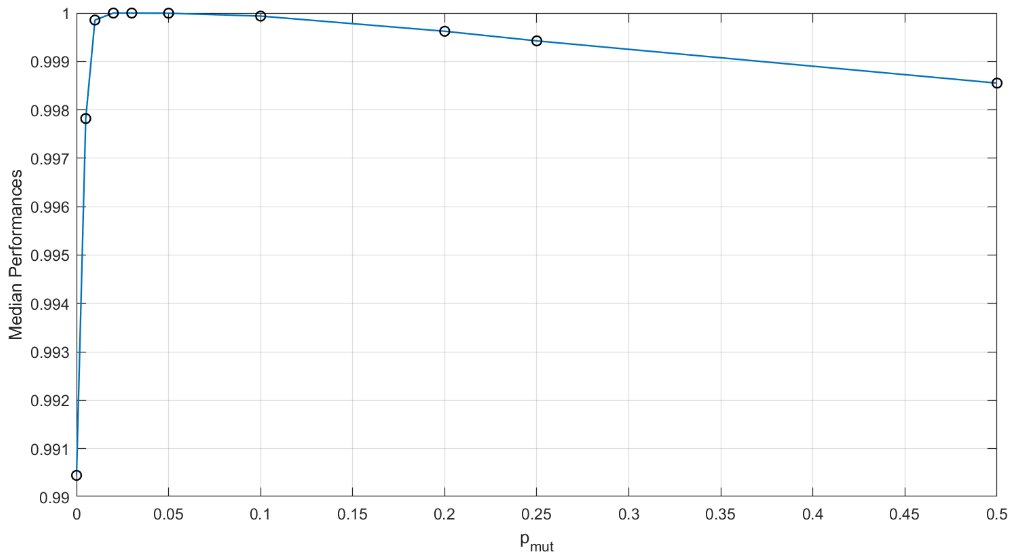

Figure 2: Median performance.

c)

According to the results of part a), the value of x_1 is close to 3, and the value of x_2 is close to 0.5. So the actual minimum point $(x_1^*, x_2^*)^T$ may equal to (3.0 , 0.5). That point (3.0 , 0.5) is a stationary point of the function g is proved as follows:
The gradient of function g is defined as follows,

$$\nabla g(x_1, x_2) = \left(\frac{\partial g}{\partial x_1}, \frac{\partial g}{\partial x_2}\right).$$

Calculate the gradient,

$$\frac{\partial g}{\partial x_1} = 2(x_2 - 1)(1.5 - x_1 + x_1x_2) + 2(x_2^2 - 1)(2.25 - x_1 + x_1x_2^2) + 2(x_2^3 - 1)(2.625 - x_1 + x_1x_2^3),$$
$$\frac{\partial g}{\partial x_2} = 2x_1(1.5 - x_1 + x_1x_2) + 4x_1x_2(2.25 - x_1 + x_1x_2^2) + 6x_1x_2^2(2.625 - x_1 + x_1x_2^3).$$

Then, put the point $(x_1 = 3, x_2 = 0.5)$ into these there parts:

$$1.5 - x_1 + x_1x_2, \quad 2.25 - x_1 + x_1x_2^2, \quad 2.625 - x_1 + x_1x_2^3.$$

Calculate the values of these three same parts of partial derivatives,

$$1.5 - x_1 + x_1x_2 = 1.5 - 3 + 3 \times 0.5 = 0;$$
$$2.25 - x_1 + x_1x_2^2 = 2.25 - 3 + 3 \times 0.25 = 0;$$
$$2.625 - x_1 + x_1x_2^3 = 2.625 - 3 + 3 \times 0.125 = 0.$$

So the value of partial derivetives is

$$\frac{\partial g}{\partial x_1} = 2(x_2 - 1) \times 0 + 2(x_2^2 - 1) \times 0 + 2(x_2^3 - 1) \times 0 = 0,$$
$$\frac{\partial g}{\partial x_2} = 2x_1 \times 0 + 4x_1x_2 \times 0 + 6x_1x_2^2 \times 0 = 0.$$

So, $\nabla g(3.0\,,\,0.5) = (0, 0)$. According to the definition on the book, points$(x_1^*, x_2^*)^T =$ (3.0 , 0.5) is a stationary point.