Home Problem 1

Erik Norlin

September, 2023

Problem 1.1

Find the minimum of the function $f(x_1, x_2) = (x_1 - 1)^2 + 2(x_2 - 2)^2$, subject to the constraint $g(x_1, x_2) = x_1^2 + x_2^2 - 1 \leq 0$.

1. **Define the function $f_p(\mathbf{x}; \mu)$ consisting of the sum of $f(x_1, x_2)$ and the penalty term.**

   The penalty function follows as:

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

   Where the penalty term is:

   $$p(\mathbf{x}; \mu) = \mu \left(\sum_{i=1}^{m} (max\{g_i(\mathbf{x}), 0\})^2 + \sum_{i=1}^{k} (h_i(\mathbf{x}))^2 \right)$$

   Inserting the given function along with the inequality constraint into the penalty function we get:

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

   From this, we have two different conditions based on whether the constraint is fulfilled or not:

   $$f_p(x_1, x_2; \mu) = \begin{cases} (x_1 - 1)^2 + 2(x_2 - 2)^2 + \mu(x_1^2 + x_2^2 - 1)^2, \text{ if } x_1^2 + x_2^2 \geq 1. \\ (x_1 - 1)^2 + 2(x_2 - 2)^2, \text{ otherwise, since } \mu(0) = 0. \end{cases}$$

2. **Compute (analytically) the gradient $f_p(\mathbf{x}; \mu)$. Including both cases where the constraints are fulfilled and where they are not.**

   Where the constraints are fulfilled:

   $$x_1^2 + x_2^2 \leq 1, (\mu(0) = 0) \Rightarrow f_p(x_1, x_2; \mu) = (x_1 - 1)^2 + 2(x_2 - 2)^2$$

   We derive $f_p(x_1, x_2; \mu)$ with respect to x_1 respectively x_2 to get the gradient:

   $$\nabla f_p = \begin{cases} f'_{px_1} = 2 * 1 * (x_1 - 1) + 0 = 2(x_1 - 1) \\ f'_{px_2} = 2 * 1 * 2(x_2 - 2) + 0 = 4(x_2 - 2) \end{cases}$$

Where the constraints are *not* fulfilled:

$$x_1^2 + x_2^2 > 1 \Rightarrow f_p(x_1, x_2; \mu) = (x_1 - 1)^2 + 2(x_2 - 2)^2 + \mu(x_1^2 + x_2^2 - 1)^2$$

We derive in the same way as before to get the gradient:

$$\nabla f_p = \begin{cases} f'_{px_1} = 2(x_1 - 1) + 2 * \mu(x_1^2 + x_2^2 - 1) * 2x_1 = 2(x_1 - 1) + 4x_1\mu(x_1^2 + x_2^2 - 1) \\ f'_{px_2} = 4(x_2 - 2) + 2 * \mu(x_1^2 + x_2^2 - 1) * 2x_2 = 4(x_2 - 2) + 4x_2\mu(x_1^2 + x_2^2 - 1) \end{cases}$$

3. **Find (analytically) the unconstrained minimum (i.e. for $\mu = 0$) of the function. This point will be used as the starting point for gradient descent.**

   Unconstrained function where $\mu = 0$:

   $f_p(x_1, x_2; \mu) = (x_1-1)^2+2(x_2-2)^2+0*4x_2(x_1^2+x_2^2-1) \Rightarrow f_p(x_1, x_2) = (x_1-1)^2+2(x_2-2)^2$

   In order to find the minimum, we need to find potential stationary points of the function. To do this we need to get the gradient and then set it to zero. From there we can calculate the values of x_1 and x_2 of the stationary points.

$$\nabla f_p = \begin{cases} f'_{px_1} = 2(x_1 - 1) = 0 \\ f'_{px_2} = 4(x_2 - 2) = 0 \end{cases}$$

   We can easily see that x_1 can only be 1 and x_2 can only be 2 in order to fulfill the requirements. We can also see that this is the only solution. Thus, the minimum of the function is at the point $(x_1^*, x_2^*)^T = (1, 2)^T$. This point will be the starting point of the gradual descent.

4. **Write a Matlab program for solving the unconstrained problem of finding the minimum of $f_p(\mathbf{x}; \mu)$ using the method of gradient descent.**

   See Matlab files.

5. **Run the program for a suitable sequence of μ values.**

   The Matlab program was run with the recommended values of the input parameters step length, $\eta = 0.0001$, and gradient tolerance, $T = 10^{-6}$. Some extra values of μ were added, one was set to zero which neutralizes the penalty term to zero. This was done to be certain that the program would recognize the minimum of the function to be at the point calculated in step 3, $(x_1^*, x_2^*)^T = (1, 2)^T$. The result shows that the minimum does occur at this point (see table 1).

Table 1: Matlab results for different μ-values in problem 1.1

μ	x_1^*	x_2^*
0	1.0000	2.0000
1	0.4338	1.2102
10	0.3314	0.9955
50	0.3159	0.9602
100	0.3137	0.9553
500	0.3120	0.9512
1000	0.3118	0.9507

From table 1 we can also see that as μ increases x_1 and x_2 converges to about 0.31 respectively 0.95 when μ is large. To illustrate this we can plot x_1 and x_2 as a function of μ (see figure 1).

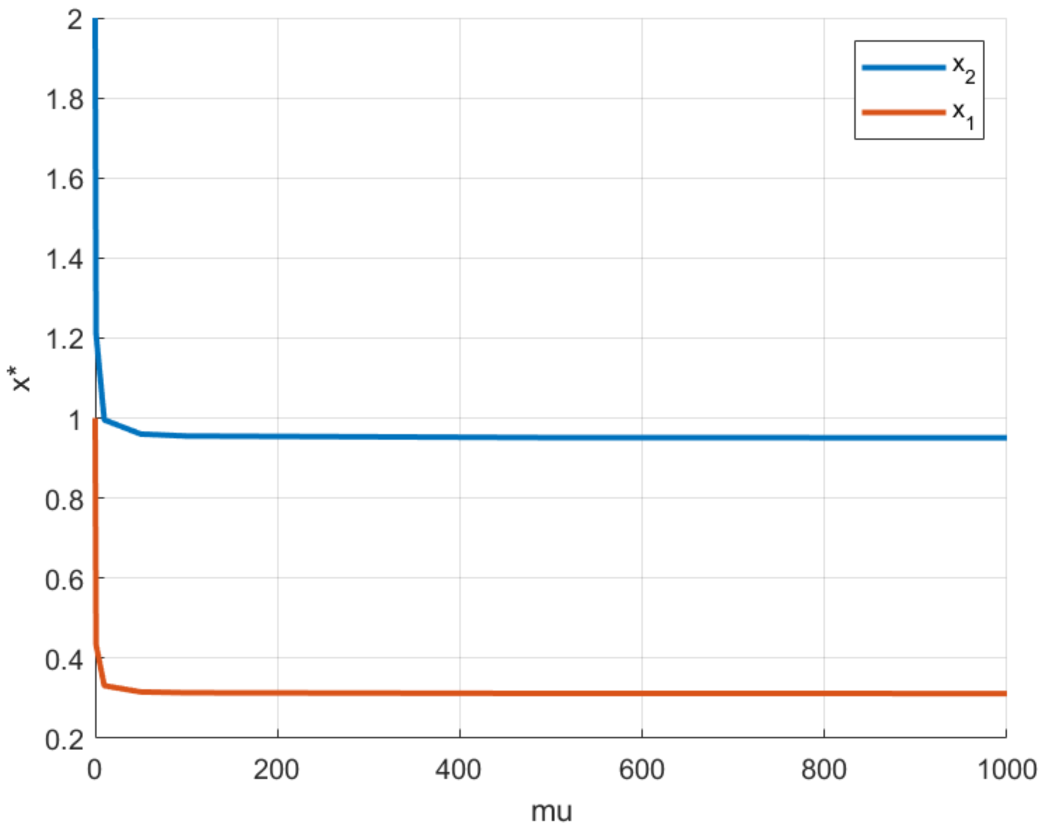

Figure 1: Matlab results problem 1.1

From figure 1 we can see that when μ is small a weak penalty is enforced, the penalty term have therefore little effect on constraining the objective function. However, the larger the value of μ is, i.e. the penalty, the clearer it becomes where the minimum of the objective function is (approximately) when it's constrained by the given constraint. Thus, in order for the penalty function to be efficient and useful, μ must be large.

Problem 1.2

a) Use the analytical method to determine the global minimum $(x_1^*, x_2^*)^T$ (as well as the corresponding function value) of the function

$$f(x_1, x2_) = 4x_1^2 - x_1x_2 + 4x_2^2 - 6x_2,$$

on the (closed) set S, shown in figure 2. The corners of the triangle are located at (0,0), (0,1) and (1,1).

In order to find the global minimum potential stationary points must be found. This is done by investigating potential stationary points in the interior of S, on the boundaries

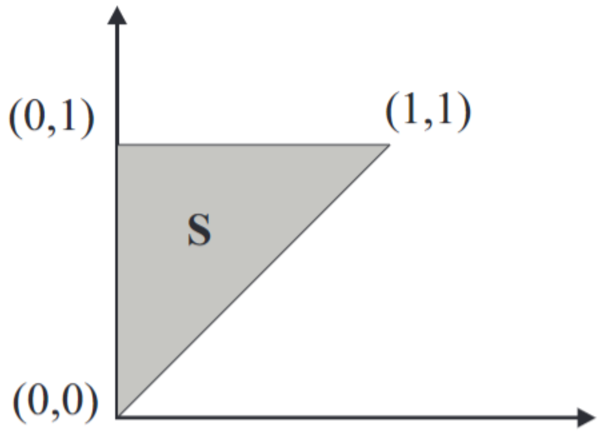

Figure 2: The set S used in Problem 1.2a.

of S and the corners of S. Every stationary point found is then evaluated in the objective function to find the lowest function value. To find potential stationary points in the interior of S we must get the partial derivative of the objective function and then set it to zero which gives us an equation system to solve for x_1 and x_2.

$$\nabla f = \begin{cases} f'_{x_1} = 2 * 4x_1^2 - 1 * x_2 + 0 - 0 = 8x_1 - x_2 \\ f'_{x_2} = 0 - 1 * x_1 + 2 * 4x_2 - 1 * 6 = 8x_2 - x_1 - 6 \end{cases}$$

$$\Rightarrow \nabla f = 0 \begin{cases} f'_{x_1} = 8x_1 - x_2 = 0 & (1) \\ f'_{x_2} = 8x_2 - x_1 - 6 = 0 & (2) \end{cases}$$

From (1) we get that $x_2 = 8x_1$ which we then put into (2) to solve for x_1:

$$8x_2 - x_1 - 6 = 0 \Rightarrow 8(8x_1) - x_1 - 6 = 0 \Rightarrow 63x_1 - 6 = 0 \Rightarrow x_1 = \frac{6}{63} = \frac{2}{21}$$

We put x_1 into (1) to solve for x_2:

$$8x_2 - x_1 - 6 = 0 \Rightarrow 8\frac{2}{21} - x_2 = 0 \Rightarrow x_2 = \frac{16}{21} \Rightarrow p_1 = \left(\frac{2}{21}, \frac{16}{21}\right)^T$$

To find potential stationary points on the boundaries of S we put one boundary into the objective function at a time, derive it, set the derivative to zero and then solve for x_1 and x_2, like before.
Boundary $x = x_1 = x_2$:

$$f(x, x) = 4x^2 - x * x + 4x^2 - 6x \Rightarrow 7x^2 - 6x$$

$$f'_x = 14x - 6 = 0 \Rightarrow x = \frac{6}{14} = \frac{3}{7} \Rightarrow p_2 = \left(\frac{3}{7}, \frac{3}{7}\right)^T$$

Boundary $x_2 = 1,\ 0 < x_1 < 1$:

$$f(x_1, 1) = 4x_1^2 - x_1 * 1 + 4 * 1^2 - 6 * 1 \Rightarrow 4x_1^2 - x_1 - 2$$

$$f'_{x_1} = 8x_1 - 1 = 0 \Rightarrow x_1 = \frac{1}{8} \Rightarrow p_3 = \left(\frac{1}{8}, 1\right)^T$$

Boundary $x_1 = 0,\ 0 < x_2 < 1$:

$$f(0, x_2) = 4 * 0^2 - 0 * x_2 + 4x_2^2 - 6x_2 \Rightarrow 4x_2^2 - 6x_2$$
$$f'_{x_2} = 8x_2 - 6 = 0 \Rightarrow x_2 = \frac{3}{4} \Rightarrow p_4 = \left(0, \frac{3}{4}\right)^T$$

Corners of S:

$$p_5 = (0, 0)^T,\ p_6 = (0, 1)^T,\ p_7 = (1, 1)^T$$

Evaluating all stationary points:

$$f(p_1) = f\left(\frac{2}{21}, \frac{16}{21}\right) = 4\left(\frac{2}{21}\right) - \left(\frac{2}{21}\right)\left(\frac{16}{21}\right) + 4\left(\frac{16}{21}\right)^2 - 6\left(\frac{16}{21}\right) =$$
$$\frac{16}{441} - \frac{32}{441} + \frac{1024}{441} - \frac{2016}{441} = -\frac{1008}{441} \approx -2.29$$

$$f(p_2) = f\left(\frac{3}{7}, \frac{3}{7}\right) = 4\left(\frac{3}{7}\right)^2 - \left(\frac{3}{7}\right)\left(\frac{3}{7}\right) + 4\left(\frac{3}{7}\right)^2 - 6\left(\frac{3}{7}\right)$$
$$7\left(\frac{3}{7}\right)^2 - 6\left(\frac{3}{7}\right) = \frac{63}{49} - \frac{126}{49} = -\frac{63}{49} \approx -1.29$$

$$f(p_3) = f\left(\frac{1}{8}, 1\right) = 4\left(\frac{1}{8}\right)^2 - 1\left(\frac{1}{8}\right) + 4 * 1^2 - 6 * 1 = \frac{4}{64} - \frac{1}{8} - 2 = \frac{1}{16} - \frac{2}{16} - \frac{32}{16} \approx -2.06$$

$$f(p_4) = f\left(0, \frac{3}{4}\right) = 4 * 0^2 - 0\left(\frac{3}{4}\right) + 4\left(\frac{3}{4}\right)^2 - 6\left(\frac{3}{4}\right) = \frac{9}{4} - \frac{18}{4} = -2.25$$

$$f(p_4) = f(0, 0) = 4 * 0^2 - 0 * 0 + 4 * 0^2 - 6 * 0 = 0$$

$$f(p_5) = f(0, 1) = 4 * 0^2 - 0 * 1 + 4 * 1^2 - 6 * 1 = -4 - 6 = 2$$

$$f(p_6) = f(1, 1) = 4 * 1^2 - 1 * 1 + 4 * 1^2 - 6 * 1 = 4 - 1 + 4 - 6 = 1$$

We can conclude that the global minimum of $f(x_1, x_2)$ is -2.29 at $p_1 = \left(\frac{2}{21}, \frac{16}{21}\right)^T$

b) Use the Lagrange multiplier method to determine the minimum $(x_1^*, x_2^*)^T$ (as well as the corresponding function value) of the function $f(x_1, x_2) = 15+2x_1+3x_2$ subject the constraint $h(x_1, x_2) = x_1^2 + x_1x_2 + x_2^2 - 21 = 0$.

To use the Lagrange multiplier method we have to define and derive $L(x_1, x_2, \lambda)$, set the gradient to zero and then solve for x_1 and x_2:

$$L(x_1, x_2, \lambda) = f(x_1, x_2) + \lambda h(x_1, x_2) = 15 + 2x_1 + 3x_2 + \lambda(x_1^2 + x_1x_2 + x_2^2 - 21) =$$
$$15 + 2x_1 + 3x_2 + \lambda x_1^2 + \lambda x_1x_2 + \lambda x_2^2 - \lambda 21$$
$$\nabla L = \begin{cases} L'_{x_1} = 0 + 2 * 1 + 0 + 2 * \lambda x_1 + 1 * \lambda x_2 + 0 + 0 = 2 + 2\lambda x_1 + \lambda x_2 \\ L'_{x_2} = 0 + 0 + 3 * 1 + 0 + 1 * \lambda x_1 + 2 * \lambda x_2 - 0 = 3 + \lambda x_1 + 2\lambda x_2 \\ L'_{\lambda} = 0 + 0 + 0 + 1 * x_1^2 + 1 * x_1x_2 + 1 * x_2^2 - 1 * 21 = x_1^2 + x_1x_2 + x_2^2 - 21 \end{cases}$$

$$\Rightarrow \nabla L = 0 \begin{cases} L'_{x_1} = 2 + 2\lambda x_1 + \lambda x_2 = 0 & (3) \\ L'_{x_2} = 3 + \lambda x_1 + 2\lambda x_2 = 0 & (4) \\ L'_{\lambda} = x_1^2 + x_1 x_2 + x_2^2 - 21 = 0 & (5) \end{cases}$$

To solve this equation system we start by multiplying (4) with 2 and subtracting with (3):

$$2(3 + \lambda x_1 + 2\lambda x_2) - (2 + 2\lambda x_1 + \lambda x_2) = 6 + 2\lambda x_1 + 4\lambda x_2 - 2 - 2\lambda x_1 - \lambda x_2$$
$$\Rightarrow 4 + 3\lambda x_2 = 0 \tag{6}$$

We then multiply (3) with 2 and subtract with (4):

$$2(2 + 2\lambda x_1 + \lambda x_2) - (3 + \lambda x_1 + 2\lambda x_2) = 4 + 4\lambda x_1 + 2\lambda x_2 - 3 - \lambda x_1 - 2\lambda x_2 = 1 + 3\lambda x_1 = 0$$
$$\Rightarrow \lambda = -\frac{1}{3x_1} \tag{7}$$

We put (7) into (6) and get:

$$4 + 3x_2\left(-\frac{1}{3x_1}\right) = 0 \Rightarrow 4 - \frac{x_2}{x_1} = 0 \Rightarrow x_2 = 4x_1 \tag{8}$$

We solve for x_1 by putting (8) into (5):

$$x_1^2 + x_1(4x_1) + (4x_1)^2 - 21 = x_1^2 + 4x_1^2 + 16x_1^2 - 21 = 21x_1^2 - 21 = 0$$
$$\Rightarrow \frac{21x_1^2 - 21}{21} = \frac{0}{21} \Rightarrow x_1^2 = 1 \Rightarrow x_1 = \pm 1$$

We then put each value of x_1 into (8) and solve for x_2:

$$x_2 = 4 * \pm 1 \Rightarrow x_2 = \pm 4$$

The stationary points we get from Lagrange multiplier method is $(1, 4)^T$ and $(-1, -4)^T$. These points are then evaluated in the objective function to determine the minimum:

$$f(1, 4) = 15 + 2 * 1 + 3 * 4 = 15 + 2 + 12 = 29$$
$$f(-1, -4) = 15 + 2 * -1 + 3 * -4 = 15 - 2 - 12 = 1$$

Thus, the minimum of the objective function is 1 at the point $(x_1^*, x_2^*)^T = (-1, -4)^T$.

Problem 1.3

a) List the selected parameters (used in RunSingle.m) and also include a table of the values of x1, x2, and $g(x_1, x_2)$ found in your 10 runs.

The selected parameters for the GA follow table 2. Population size, maximum variable range, number of genes, number of variables were set and not allowed to change. The other parameters were selected in a way to reach as close to the minimum as possible.

Table 2: Parameters in RunSingle.m for problem 1.3a

Parameters	
populationSize	100
maximumVariableValue	5
numberOfGenes	50
numberOfVariables	2
tournamentSize	3
tournamentProbability	0.782
crossoverProbability	0.8
mutationProbability	0.02
numberOfGenerations	2000

Table 3: Matlab results for 10 single runs in problem 1.3a

$g(x_1, x_2)$	x_1^*	x_2^*
1.618613e-04	2.9687499395	0.4921503512
2.087219e-14	3.0000000596	0.4999999851
2.680078e-13	3.0000012517	0.5000002831
2.087219e-14	3.0000000596	0.4999999851
2.183973e-03	3.1250002421	0.5294348457
2.087219e-14	3.0000000596	0.4999999851
2.087219e-14	3.0000000596	0.4999999851
2.087219e-14	3.0000000596	0.4999999851
7.924772e-13	3.0000021458	0.5000005811
2.087219e-14	3.0000000596	0.4999999851

b)

Table 4: Matlab results from batch runs for different mutation probabilities in problem 1.3b

p_{mut}	Median performance
0.00	0.9942931649
0.01	0.9998576980
0.02	0.9999999919
0.03	0.9999994689
0.04	0.9999985323
0.05	0.9999902985
0.06	0.9999859600
0.07	0.9999787790
0.08	0.9999742872
0.09	0.9999583999
0.10	0.9999536186
0.20	0.9996746959
0.30	0.9989615735
0.40	0.9988601474
0.50	0.9986373886

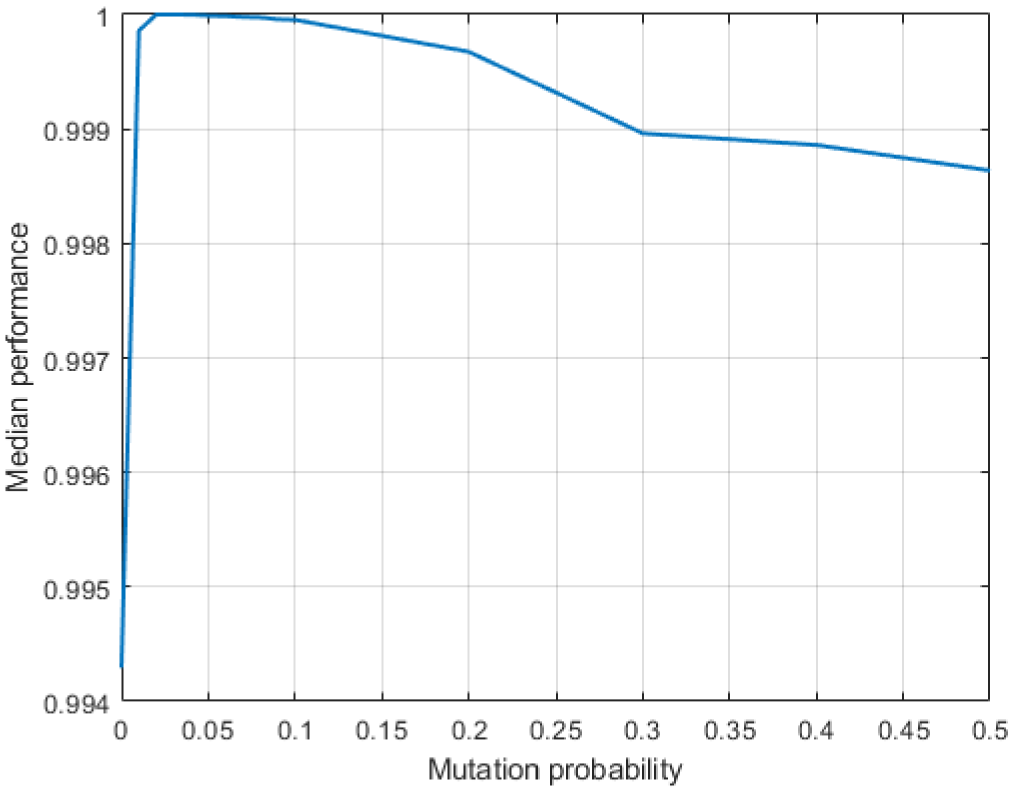

Figure 3: Plotted results from table 4

It's generally said that the optimal value for p_{mut} in a GA is around 1/N (N = number of genes). We can confirm that this is the case with the results from the batch runs since the number of genes used in the algorithm were 50 genes per chromosome and the mutation probability that gave highest performance was 0.02 (1/50) (see table 4 and figure 3). In this particular problem the fitness values are very close to the optimum no matter what mutation probability used. This could be that this is a relative simple GA. In more complex problem it could be that the difference in median performance for different mutation probabilities would be greater. Therefore it's always important to choose a mutation probability that is 1/N.

c)

From table 3 we make an educated guess based on the 10 runs from (a) that the minimum of the objective function is at the point $(x_1, x_2)^T = (3, 0.5)^T$. To determine whether this is the true minimum analytically we have to derive the objective function, insert the point in the gradient and then set the gradient to zero. If the equations holds true, the point is indeed a stationary point, and assumed to be a minimum (since it's not necessary to prove that it's a minimum).

$$g(x_1, x_2) = (1.5 - x_1 + x_1 x_2)^2 + (2.25 - x_1 + x_1 x_1^2)^2 + (2.625 - x_1 + x_1 x_2^3)^2$$

$$g'_{x_1} = 2(1.5 - x_1 + x_1 x_2)(x_2 - 1) + 2(2.25 - x_1 + x_1 x_2^2)(x_2^2 - 1) + (2.625 - x_1 + x_1 x_2^3)(x_2^3 - 1)$$

$$g'_{x_2} = 2*(1.5 - x_1 + x_1 x_2)*x_1 + 2*(2.25 - x_1 + x_1 x_2^2)*2x_1 x_2 + 2*(2.625 - x_1 + x_1 x_2^3)*3x_1 x_2 =$$

$$2x_1(1.5 - x_1 + x_1 x_2) + 4x_1 x_2(2.25 - x_1 + x_1 x_2^2) + 6x_1 x_2(2.625 - x_1 + x_1 x_2^3)$$

$$\Rightarrow$$

$$\nabla g \begin{cases} g'_{x_1} = 2(1.5 - x_1 + x_1 x_2)(x_2 - 1) + 2(2.25 - x_1 + x_1 x_2^2)(x_2^2 - 1) + (2.625 - x_1 + x_1 x_2^3)(x_2^3 - 1) \\ g'_{x_2} = 2x_1(1.5 - x_1 + x_1 x_2) + 4x_1 x_2(2.25 - x_1 + x_1 x_2^2) + 6x_1 x_2(2.625 - x_1 + x_1 x_2^3) \end{cases}$$

$$\Rightarrow$$

$$\nabla g\left(3, \frac{1}{2}\right) = 0 \begin{cases} g'_{x_1}(3, 0.5) = 0 & (9) \\ g'_{x_2}(3, 0.5) = 0 & (10) \end{cases}$$

$$g'_{x_1}(3, 0.5) = 2(1.5-3+3*0.5)(0.5-1) + 2(2.25-3+3*0.5^2)(0.5^2-1) + (2.625-3+3*0.5^3)(0.5^3-1) =$$

$$2(1.5 - 3 + 1.5)(-0.5) + 2(2.25 - 3 + 0.75)(-0.75) + (2.625 - 3 + 0.375)(-0.875) =$$

$$2(0)(-0.5) + 2(0)(-0.75) + (0)(-0.875) = 0$$

$$g'_{x_2}(3, 0.5) = 2*0.5*(1.5-3+3*0.5) + 4*3*0.5*(2.25-3+3*0.5^2) + 6*3*0.5*(2.625-3+3*0.5^3) =$$

$$(1.5 - 3 + 1.5) + 6(2.25 - 3 + 0.75) + 9(2.625 - 3 + 0.375) =$$

$$0 + 6(0) + 9(0) = 0$$

We can see that the estimated point $(x_1, x_2)^T = (3, 0.5)^T$ from (a) is truly a stationary point, and assumed to be a minimum, since the equations (9) and (10) holds true. The educated guess was therefore liable.