FFR105 - HP1

Nicholas Granlund

2023-09-17

Problem 1.1 - Penalty method

1. Define (and specify clearly, in your report, as a function of x_1, x_2, and μ) the function $f_p(x;\mu)$, consisting of the sum of $f(x_1, x_2)$ and the penalty term.

The penalty function is defined as:

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

However, since we will only consider one inequality constraint in this problem, we can simplify the penalty function to obtain a new expression as:

$$p(\mathbf{x};\mu) = \mu\big(max[0, g(\mathbf{x})]\big)^2$$

With this we can now express the objective function, which is the function we want to minimize with respect to $\mathbf{x}$. This is the sum of the function f and the penalty p:

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

$$f_{penalty}(\mathbf{x};\mu) = (x_1 - 1)^2 + 2(x_2 - 2)^2 + \mu(max[0, x_1^2 + x_2^2 - 1])^2$$

The objective function $f_{penalty}$ can be expressed differently depending if the constraint is fullfilled or not. We thereby obtain the final expression for the objective function as:

$$f_{penalty}(\mathbf{x};\mu) = \begin{cases} (x_1 - 1)^2 + 2(x_2 - 2)^2 + \mu(x_1^2 + x_2^2 - 1)^2 & \text{if } (x_1^2 + x_2^2 - 1) > 0 \\ (x_1 - 1)^2 + 2(x_2 - 2)^2 & \text{Otherwise} \end{cases}$$

2. Next, compute (analytically) the gradient $\nabla f_p(x;\mu)$, and include it in your report. Make sure to include both the case where the constraints are fulfilled and the case where they are not.

To obtain the gradient of the objective function $f_{penalty}(\mathbf{x};\mu)$ we just have to compute the gradient separately for each scenario when the constrain is fullfilled and not fullfilled. We obtain the gradient:

$$\nabla_{\mathbf{x}} f_{penalty}(\mathbf{x};\mu) = \begin{cases} \begin{bmatrix} 2(x_1-1)+4\mu x_1(x_1^2+x_2^2-1) \\ 4(x_2-2)+4\mu x_2(x_1^2+x_2^2-1) \end{bmatrix} & \text{if } (x_1^2+x_2^2-1) > 0 \\ \\ \begin{bmatrix} 2(x_1-1) \\ 4(x_2-2) \end{bmatrix} & \text{Otherwise} \end{cases}$$

3. Find (analytically) the unconstrained minimum (i.e., for $\mu = 0$) of the function, and include it in your report. This point will be used as the starting point for gradient descent.

We find the unconstrained minimum of the objective function by finding the point of the objective function which has the slope 0. We solve the system of equations:

$$2(x_1 - 1) = 0$$
$$4(x_2 - 2) = 0$$

Which has the solution:

$$x_1 = 1$$
$$x_2 = 2$$

We can also verify that this point does not fullfill the constraint as it is not located inside the unit circle.

$$(x_1^2 + x_2^2 - 1) = (1^2 + 2^2 - 1) = 4$$

4. Write a Matlab program for solving the unconstrained problem of finding the minimum of $f_p(x;\mu)$ using the method of gradient descent. You should make use of the skeleton files available on Canvas.

See appended Matlab files.

5. Run the program for a suitable sequence of μ values (see `RunPenaltyMethod.m`). Select a suitable (small) value for the step length η, and specify it clearly, along with the sequence of μ values, in your report. Example of suitable parameter values: $\eta = 0.0001$, $T = 10^{-6}$, sequence of μ values: 1, 10, 100, 1000. As output, the program gives the components of the vector x for the different μ values. You should include the table in your report; see below. Specify the values of x_1^* and x_2^* with 4 decimal precision. Do not just print the raw Matlab output (with many decimals, for example) in your report! You should also check that your results are reasonable, i.e., that the sequence of points appears to be convergent, for example by plotting the values of x_1 and x_2 as functions of μ.

The Matlab program was run with $\eta = 0.0001$, $T = 10^{-8}$ and the sequence of μ was 0.1, 1, 10, 100, 1000 and 10000. The starting point was, as instructed, set to be the unconstrained solution $\mathbf{x}_0 = [1 \quad 2]^\top$

Table 1: Solution found with GD, μ, x_1^*, and x_2^*

μ	x_1^*	x_2^*
0.1	0.6963	1.6419
1	0.4338	1.2102
10	0.3341	0.9955
100	0.3137	0.9553
1000	0.3118	0.9507
10000	NaN	NaN

However, we can observe that not all solutions found satisfy the constraint. e.g., We can see that the solutions found for $\mu = 0.1$ and $\mu = 1$ is not inside the unit circle. By curiosity we plot the GD trajectory along with the contour of the objective function and the constraint. These plots are shown in Figure 1-6 below.

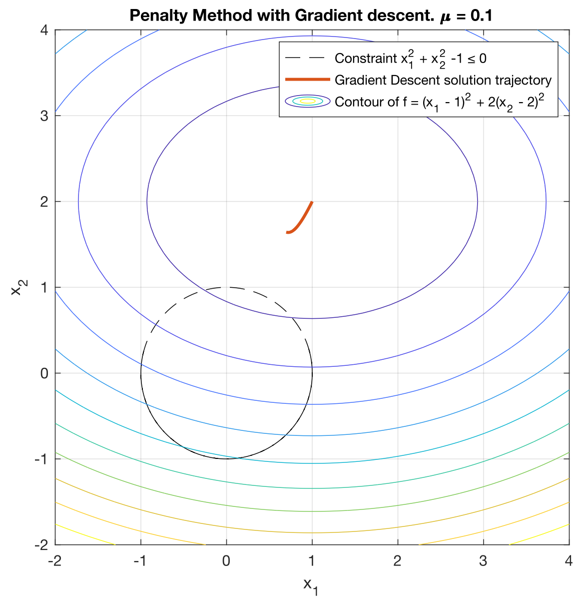

Figure 1: GD trajectory for $\mu = 0.1$

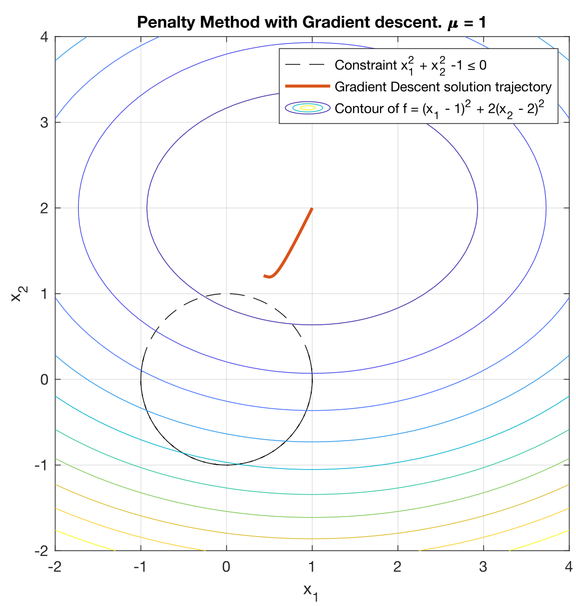

Figure 2: GD trajectory for $\mu = 1$

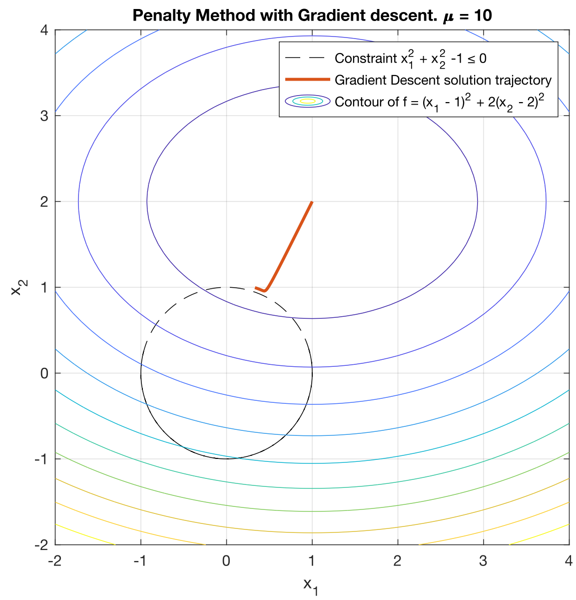

Figure 3: GD trajectory for $\mu = 10$

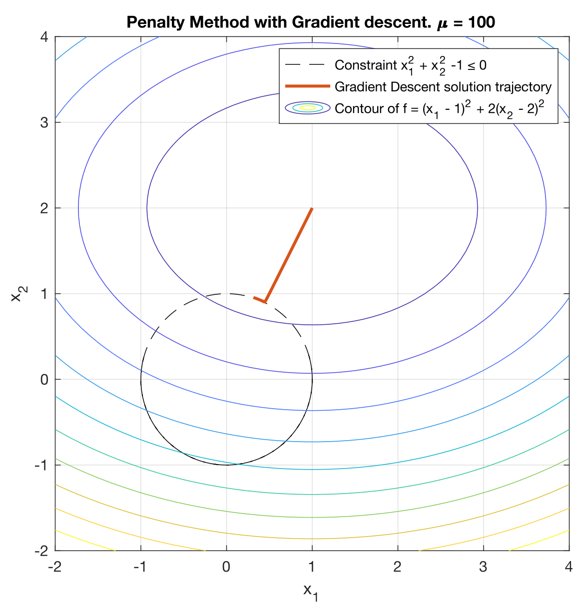

Figure 4: GD trajectory for $\mu = 100$

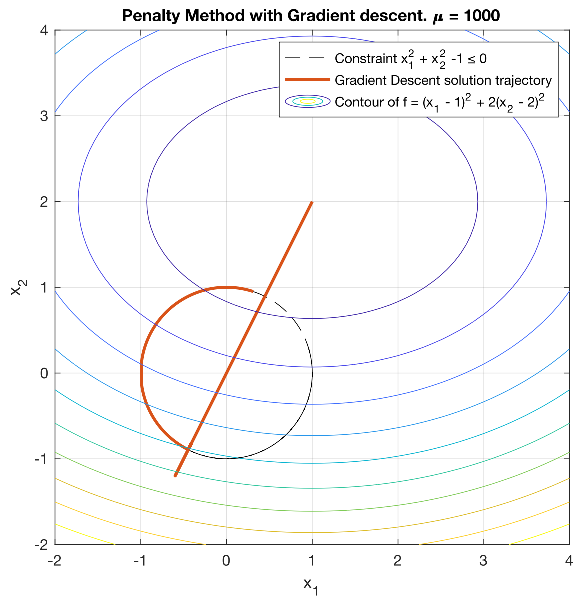

Figure 5: GD trajectory for $\mu = 1000$

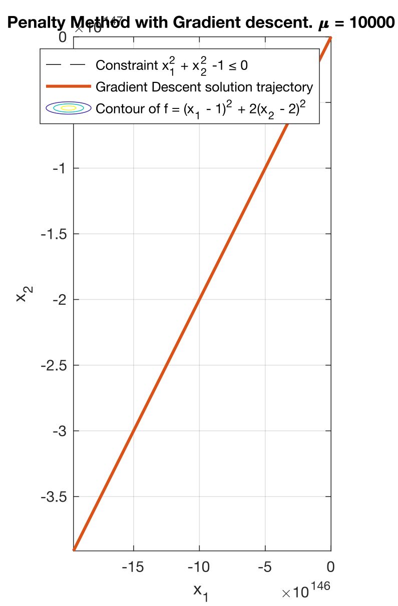

Figure 6: GD trajectory for $\mu = 10000$

Sure enough, not all solutions found do satisfy the constraint. If μ is set too small, then the penalty for not satisfying the constraint is not large enough to point the GD solution towards the constraint. This is apparent in Figure 1-2 when μ is indeed to small to reach the constraint boundary. On the other hand, if μ is too large, there is higher penalty for not satisfying the constraint and the GD may overshoot the boundary of the constraint, which can in some cases make the solution diverge.

For instance, when $\mu = 1000$ the penalty for not satisfying the constraint is fairly large. According to Figure 5, we can see that the first iteration of GD takes us from our initial guess $\mathbf{x}_0$ towards the boundary of our constraint. However since the penalty is so large, the GD overshoots in the first step, making the next approximation still be outside the boundary of the constraint (below to left). Nevertheless, it still manages to converge towards the solution. This is not the case when $\mu = 10000$. As seen in Figure 6, the solution never manages to settle and breaks only when the GD has reached a NaN solution. This is because the GD iterations continuously overshoots the boundary and the penalty grows larger and larger. By observing the trajectory for the GD iterations, we can determine that $\mu = 100$ is the most suitable value to find a minimum of our objective function.

Furthermore, as requested by the assignment - a plot of the values of x_1 and x_2 as a function of μ is shown in Figure 7 below. There it can be seen that the solution converges as μ grows. To conclude this part of the assignment, I would claim that the material presented here does indeed show that the results are reasonable.

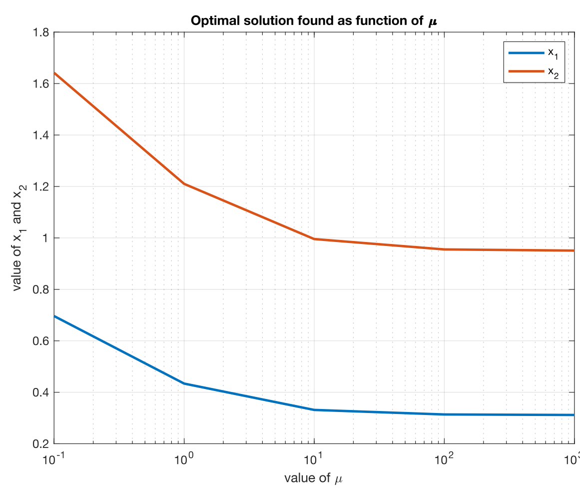

Figure 7: GD solution which minimizes the objective function. Plotted over μ in semilogx plot. The solution converges as μ grows

Problem 1.2 - Constrained optimization

a) Use the analyticall method to determine the global minimum (x_1^*, x_2^*) of the function $f(x_1, x_2)$

1. Interior: The first step is to consider the interior of the function. We have the function:

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

The gradient of this is computed as:

$$\nabla f(x_1, x_2) = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \end{bmatrix}$$
$$\nabla f(x_1, x_2) = \begin{bmatrix} 8x_1 - x_2 \\ 8x_2 - x_1 - 6 \end{bmatrix}$$

Solving this for when $\nabla f(x_1, x_2) = 0$ we get a point of the interior of the set at:

$$P_1 = \left(\frac{2}{21}, \frac{16}{21}\right)$$

2. Boundary: The second step is to find stationary points on the boundary. For this triangular boundary, we have three boundaries and three corners to consider.

Case 1: Rightmost boundary $x_1 = 0, 0 \leq x_2 \leq 1$

$$f(0, x_2) = 4x_2^2 - 6x_2$$
$$\frac{\partial}{\partial x_2} f(0, x_2) = 8x_2 - 6$$
$$8x_2 - 6 = 0 \quad \rightarrow \quad P_2 = \left(0, \frac{3}{4}\right)$$

Case 2: Topmost boundary $0 \leq x_1 \leq 1, x_2 = 1$

$$f(x_1, 1) = 4x_1^2 - x_1 - 2$$
$$\frac{\partial}{\partial x_1} f(x_1, 1) = 8x_1 - 1$$
$$8x_1 - 1 = 0 \quad \rightarrow \quad P_3 = \left(\frac{1}{8}, 1\right)$$

Case 3: Bottom boundary $x_1 = x_2 (= x)$

$$\begin{aligned} f(x_1, x_2) = 4x_1^2 - x_1 x_2 + 4x_2^2 - 6x_2 = 7x^2 - 6x \\ \frac{\partial}{\partial x} f(x_1, x_2) = 14x - 6 \\ 14x - 6 = 0 \quad \rightarrow \quad P_4 = \left(\frac{3}{7}, \frac{3}{7}\right) \end{aligned}$$

Case 4: Corners

$$P_5 = \left(0, 0\right)$$

$$P_6 = \left(0, 1\right)$$

$$P_7 = \left(1, 1\right)$$

3. Investigate points and find minimum: Finally, we will compute the function value at these found points to determine where the minimum is located. We get:

$$P_1 = \left(\frac{2}{21}, \frac{16}{21}\right) \quad \rightarrow \quad f\left(\frac{2}{21}, \frac{16}{21}\right) = -2.285$$

$$P_2 = \left(0, \frac{3}{4}\right) \quad \rightarrow \quad f\left(0, \frac{3}{4}\right) = -2.250$$

$$P_3 = \left(\frac{1}{8}, 1\right) \quad \rightarrow \quad f\left(\frac{1}{8}, 1\right) = -2.062$$

$$P_4 = \left(\frac{3}{7}, \frac{3}{7}\right) \quad \rightarrow \quad f\left(\frac{3}{7}, \frac{3}{7}\right) = -1.286$$

$$P_5 = \left(0, 0\right) \quad \rightarrow \quad f(0, 0) = 0$$

$$P_6 = \left(0, 1\right) \quad \rightarrow \quad f(0, 1) = -2$$

$$P_7 = \left(1, 1\right) \quad \rightarrow \quad f(1, 1) = 1$$

We can thereby conclude that the function has a minimum in the interior of the constraint at $P_1 = \left(\frac{2}{21}, \frac{16}{21}\right)$ with the value $f\left(\frac{2}{21}, \frac{16}{21}\right) = -2.285$

Problem 1.3 - Basic GA program

a) Carry out 10 runs of the GA algorithm.Note the values found, both for x_1 and x_2 and for the function $g(x_1, x_2)$

The algorithm was run 10 times with the parameters listed below. We obtained the solutions presented in Table 2, which is consistently around $x_1 = 3$, $x_2 = 0.5$ with a function value of $g(x_1, x_2) = 0$:

1. `populationSize` = 100
2. `maximumVariableValue` = 5
3. `numberOfGenes` = 50
4. `numberOfVariables` = 2
5. `tournamentSize` = 2
6. `tournamentProbability` = 0.75
7. `crossoverProbability` = 0.8
8. `mutationProbability` = 0.02
9. `numberOfGenerations` = 2000

Table 2: Solution found with GA on 10 different runs, $g(x_1^*, x_2^*)$, x_1^*, and x_2^*

$g(x_1^*, x_2^*)$	x_1^*	x_2^*
0.0000	3.0000	0.5000
0.0000	2.9999	0.4999
0.0000	3.0000	0.5000
0.0000	2.9999	0.4999
0.0000	2.9999	0.4999
0.0000	3.0000	0.5000
0.0000	2.9998	0.4999
0.0000	3.0000	0.4999
0.0000	2.9999	0.4999
0.0000	2.9999	0.4999

b) Make a parameter search for the mutation rate. Tabulate median fitness value over 100 runs for different mutation probabilities.

By running the GA 100 times for a given mutation probability p_{mut}, we obtain the median fitness that is presented in Table 3 below. This data is also plotted in Figure 8.

As we can see, the optimal mutation probability occurs at around $p_{mut} = 0.020$, which is equal to the inverse of the amount of genes (1/50). By setting the mutation probability as $1/m$ for a given problem, we ensure that the probability scales with the amount of genes. Furthermore this probability ensures that, on average, atleast one gene will mutate in each generation, preventing excessive exploration (high mutation rates) or excessive exploitation (low mutation rates).

Moreover, as the mutation probability increases, the median fitness becomes increasingly influenced by randomness and mutation, indicating that the genetic algorithm is favoring excessive exploration.

Finally, having a zero mutation probability means that the GA will resort to only exploitation, which also negatively impacts the median fitness. In the absence of mutation the GA can become stagnant at sub-optimal solutions due to low diversity in the genes.

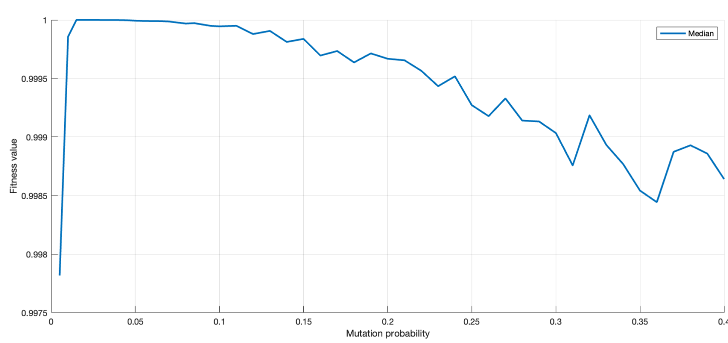

Figure 8: Mediam performance of GA as a function of mutation probability p_{mut}.

Table 3: Table of median fitness value obtained for a mutation probability

Mutation probability	Median fitness of 100 runs
0.000	0.992539197021038
0.005	0.997817448043847
0.010	0.999857697963962
0.015	0.999999927795171
0.020	0.999999988610861
0.025	0.999999934664160
0.030	0.999999421152109
0.035	0.999999418091428
0.040	0.999998726662775
0.045	0.999997285312284
0.050	0.999993521300850
0.055	0.999990925916507
0.060	0.999990052802438
0.065	0.999989373562712
0.070	0.999986407655067
0.075	0.999977616821504
0.080	0.999968923914241
0.085	0.999972508944636
0.090	0.999960523552190
0.095	0.999948865344294
0.100	0.999945112510761
0.110	0.999949780123482
0.120	0.999879510488538
0.130	0.999906290571212
0.140	0.999812036428136
0.150	0.999838148537463
0.160	0.999695844551200
0.170	0.999734426859633
0.180	0.999637153894994
0.190	0.999714041557952
0.200	0.999668294525260
0.210	0.999655362095071
0.220	0.999565660587857
0.230	0.999434203114296
0.240	0.999518539632532
0.260	0.999178332477606
0.280	0.999140594023956
0.300	0.999034162430002
0.320	0.999185316518929
0.340	0.998768910214929
0.360	0.998443599597400
0.380	0.998929200287116
0.400	0.998641031787893

c) Prove analytically that the point $(x_2^*, x_2^*)^\top$ indeed is a stationary point of the function g.

To prove that the point is a stationary point of the function. We need to compute the gradient of the function

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

We need to compute the partial derivatives of g with respect to x_1 and x_2. The gradient will have two rows corresponding to these partial derivatives. We compute the gradient as:

$$\nabla g(x_1, x_2) = \begin{bmatrix} \frac{\partial g}{\partial x_1} \\ \frac{\partial g}{\partial x_2} \end{bmatrix}$$

$$\frac{\partial g}{\partial x_1} = 2(1.5 - x_1 + x_1 \cdot x_2)(-1 + x_2) + 2(2.25 - x_1 + x_1 \cdot x_2^2)(-1 + x_2^2) + 2(2.625 - x_1 + x_1 \cdot x_2^3)(-1 + x_2^3)$$

$$\frac{\partial g}{\partial x_2} = 2(1.5 - x_1 + x_1 \cdot x_2)(x_1) + 2(2.25 - x_1 + x_1 \cdot x_2^2)(2x_1 \cdot x_2) + 2(2.625 - x_1 + x_1 \cdot x_2^3)(3x_1 \cdot x_2^2)$$

Now we evaluate the gradient at the point the GA found at $(x_1^*, x_2^*) = (3, 0.5)$. We insert the point in the expressions above and get:

$$\left.\frac{\partial g}{\partial x_1}\right|_{(3,0.5)} = 2(\underbrace{1.5 - 3 + 3 \cdot 0.5}_{=0})(-1+0.5) + 2(\underbrace{2.25 - 3 + 3 \cdot (0.5)^2}_{=0})(-1+(0.5)^2) + 2(\underbrace{2.625 - 3 + 3 \cdot (0.5)^3}_{=0})(-1+(0$$

$$\left.\frac{\partial g}{\partial x_2}\right|_{(3,0.5)} = 2(\underbrace{1.5 - 3 + 3 \cdot 0.5}_{=0})(3) + 2(\underbrace{2.25 - 3 + 3 \cdot (0.5)^2}_{=0})(2 \cdot 3 \cdot 0.5) + 2(\underbrace{2.625 - 3 + 3 \cdot (0.5)^3}_{=0})(3 \cdot 3 \cdot (0.5)^2)$$

Which indeed proves that:

$$\left.\frac{\partial g}{\partial x_1}\right|_{(3,0.5)} = 0$$

$$\left.\frac{\partial g}{\partial x_2}\right|_{(3,0.5)} = 0$$

The GA has hence found a stationary point of the function $g(x_1, x_2)$ $(x_1^*, x_2^*) = (3, 0.5)$.