FFR105, Stochastic Optimization Algorithms
Home Problem 1

September 24, 2019

Ella Guiladi
930509-0822
guiladi@student.chalmers.se

1 Problem 1.1

1. The assignment is to find the minimum of the function

$$f(x_1, x_2) = (x_1 - 1)^2 + 2(x_2 - 2)^2, \tag{1}$$

subject to the constraint

$$g(x_1, x_2) = x_1^2 + x_2^2 - 1 \leq 0, \tag{2}$$

by using the penalty method. The definition of $f_p(\boldsymbol{x}; \mu)$ from the course literature is as follows

$$f_p(\boldsymbol{x}; \mu) = f(\boldsymbol{x}) + p(\boldsymbol{x}; \mu). \tag{3}$$

Where the penalty term $p(\boldsymbol{x}; \mu)$ is defined as

$$p(\boldsymbol{x}; \mu) = \mu \left(\sum_{i=1}^{m} (max\{g_i(\boldsymbol{x}), 0\})^2 + \sum_{i=1}^{k} (h_i(\boldsymbol{x}))^2 \right). \tag{4}$$

Which in our case becomes

$$p(\boldsymbol{x}; \mu) = \mu(max\{x_1^2 + x_2^2 - 1, 0\})^2, \tag{5}$$

since there's only one constraint, the summation symbols from (4) disappears. The final expression for $f_p(\boldsymbol{x}; \mu)$ in our case thereby becomes

$$f_p(\boldsymbol{x}; \mu) = \begin{cases} (x_1 - 1)^2 + 2(x_2 - 2)^2 + \mu(x_1^2 + x_2^2 - 1)^2, & if \quad x_1^2 + x_2^2 - 1 > 0 \\ (x_1 - 1)^2 + 2(x_2 - 2)^2, & otherwise. \end{cases} \tag{6}$$

2. The gradient of $f_p(\boldsymbol{x}; \mu)$ is computed by taking the partial derivative of the function, including both the case with the constraint, as well as without.

$$\frac{\partial f_p}{\partial x_1} = \begin{cases} 2(x_1 - 1) + 4\mu \cdot x_1(x_1^2 + x_2^2 - 1), & if \quad x_1^2 + x_2^2 - 1 > 0 \\ 2(x_1 - 1), & otherwise, \end{cases} \tag{7}$$

$$\frac{\partial f_p}{\partial x_2} = \begin{cases} 4(x_2 - 2) + 4\mu \cdot x_2(x_1^2 + x_2^2 - 1), & if \quad x_1^2 + x_2^2 - 1 > 0 \\ 4(x_2 - 2), & otherwise. \end{cases} \tag{8}$$

3. By choosing $\mu = 0$ the unconstrained minimum can be calculated by solving the function $\nabla f_p(x_1, x_2)$, which results in the following equations

$$\nabla f_p(\boldsymbol{x}; \mu = 0) = \begin{cases} 2(x_1 - 1), \\ 4(x_2 - 2). \end{cases} \tag{9}$$

By setting these equations equal to zero, following point $(x_1, x_2)^T = (1, 2)^T$ can be presented. One can also see that this is a global minimum, since the unconstrained function

$$f_p(\boldsymbol{x}; \mu = 0) = (x_1 - 1)^2 + 2(x_2 - 2)^2, \tag{10}$$

never can be < 0 since $(x_1 - 1)^2$ and $2(x_2 - 2)^2$ always will be ≥ 0. This point will be used as the starting point for the gradient descent method.

4. The Matlab code for the main file RunPenaltyMethod.m and the function files RunGradientDescent.m and ComputeGradient.m are used with the penalty method and the gradient descent method, for finding the minimum of the function in attached file 1.1.The function file ComputeGradient.m is used to compute the gradient without the constraints. RunGradientDescent.m calculates the gradient with ComputeGradient.m and finding new points with gradient descent. The main file RunPenaltyMethod.m uses RunGradientDescent.m to run the penalty method. By running the main file one gets the minimum of the function as an output.

5. The resulted table from running the main file (RunPenaltyMethod.m) is presented below, with the parameter values chosen as: $\eta = 0.0001$, $T = 10^{-6}$, $\mu \in \{0, 1, 10, 100, 1000\}$

Table 1: Values of x_1^*,x_2^* with corresponding values of μ

μ	x_1^*	x_2^*
0	1	2
1	0.434	1.210
10	0.331	0.995
100	0.314	0.955
1000	0.312	0.951

From analysing the result in table 1, it is clear that the sequence of points are convergent, i.e the result converge to a set point.

2 Problem 1.2

(a) For this problem, the analytical method described on pp. 29-30 in the course book is used in order to determine the global minimum $(x_1^*,x_2^*)^T$ and the function value of the following function

$$f(x_1, x_2) = 4x_1^2 - x_1x_2 + 4x_2^2 - 6x_2. \tag{11}$$

The function (11) is on the closed set S with corners located at (0,0), (0,1) and (1,1). First of all, the interior points of the set is investigated by calculating the gradient of the function from equation (11)

$$\nabla f(x_1, x_2) = (8x_1 - x_2, -x_1 + 8x_2 - 6). \tag{12}$$

Setting the equation (12) equal to zero, one obtains

$$0 = \begin{cases} 8x_1 - x_2, \\ -x_1 + 8x_2 - 6 \end{cases} \tag{13}$$

To solve the values for x_1, x_2, following calculation were performed by expressing x_2 as x_1 and substituting it in equation (13).

$$\begin{cases} x_2 = 8x_1 \\ -x_1 + 64x_1 - 6 = 0 \\ x_1 = \frac{2}{21}. \end{cases} \tag{14}$$

Now x_2 can be calculated as

$$x_2 = 8 \cdot \frac{2}{21} = \frac{16}{21}. \tag{15}$$

which results in following stationary point

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

where $P_1 \in S$.

The next step is to investigate the boundary of the set, δS, which give rise to four different cases.
1. For $x_1 = 0$ and $0 < x_2 < 1$, one get the function

$$f(0, x_2) = 4x_2^2 - 6x_2 \tag{17}$$

with the derivative

$$f'(0, x_2) = 8x_2 - 6. \tag{18}$$

By setting $f'(0, x_2)$ equal to zero, the stationary point

$$P_2 = \left(0, \frac{3}{4}\right)^T, \tag{19}$$

is obtained where $P_2 \in S$.

2. It is possible to use the equation for the straight line: $y = kx + m$, where $k = 1$ and $m = 0$, which gives $x_2 = x_1 = x$. This results in following equations

$$\begin{cases} f(x) = 4x - x^2 + 4x^2 - 6x, \\ f'(x) = 14x - 6, \end{cases} \tag{20}$$

By setting $f'(x)$ equal to zero, one can solve the following stationary point

$$P_3 = \left(\frac{3}{7}, \frac{3}{7}\right)^T, \tag{21}$$

where $P_3 \in S$.

3. For $x_2 = 1$ and $0 < x_1 < 1$ one obtains the following equations

$$\begin{cases} f(x_1, 1) = 4x_1^2 - x_1 - 2, \\ f'(x_1, 1) = 8x_1 - 1. \end{cases} \tag{22}$$

By setting $f'(x_1, 1)$ equal to zero, the stationary point

$$P_4 = \left(\frac{1}{8}, 1\right)^T, \tag{23}$$

is found, where $P_4 \in S$.

4. The last stationary points are the corners

$$\begin{cases} P_5 = (0, 0)^T, \\ P_6 = (0, 1)^T \\ P_7 = (1, 1)^T \end{cases} \tag{24}$$

Lastly, we need to insert all the stationary points in the function from equation (11) to find the global minimum.

$$\begin{cases} P_1 = \left(\frac{2}{21}, \frac{16}{21}\right)^T, & f(P_1) = -\frac{16}{7} \\ P_2 = \left(0, \frac{3}{4}\right)^T, & f(P_2) = -\frac{9}{4} \\ P_3 = \left(\frac{3}{7}, \frac{3}{7}\right)^T, & f(P_3) = -\frac{9}{7} \\ P_4 = \left(\frac{1}{8}, 1\right)^T, & f(P_4) = -\frac{33}{16} \\ P_5 = (0, 0)^T, & f(P_5) = 0 \\ P_6 = (0, 1)^T, & f(P_6) = -2 \\ P_7 = (1, 1)^T, & f(P_7) = 1 \end{cases} \tag{25}$$

By comparing all the function values in the different stationary points, the conclusion can be drawn that the function from equation (11) has its global minimum in the stationary point $P_1 = \left(\frac{2}{21}, \frac{16}{21}\right)^T$, with the corresponding function value $f(P_1) = -\frac{16}{7}$.

(b) For this problem, the Lagrange multiplier method is used in order to determine the global minimum and function value of the following function

$$f(x_1, x_2) = 15 + 2x_1 + 3x_2, \tag{26}$$

subject to the constraint

$$h(x_1, x_2) = x_1^2 + x_1x_2 + x_2^2 - 21 = 0 \tag{27}$$

The Lagrange function $L(x_1, x_2, \lambda)$ is defined as

$$L(x_1, x_2, \lambda) = f(x_1, x_2) + \lambda h(x_1, x_2) = 15 + 2x_1 + 3x_2 + \lambda(x_1^2 + x_1x_2 + x_2^2 - 21). \tag{28}$$

The next step is to calculate the gradient of L and setting in to zero, i.e calculating the partial derivatives of $L(x_1, x_2, \lambda)$ equal to zero,

$$\frac{\partial L}{\partial x_1} = 2 + \lambda(2x_1 + x_2) = 0, \tag{29}$$

$$\frac{\partial L}{\partial x_2} = 3 + \lambda(x_1 + 2x_2) = 0, \tag{30}$$

$$\frac{\partial L}{\partial \lambda} = x_1^2 + x_1x_2 + x_2^2 - 21 = 0. \tag{31}$$

The result is an equation system that is solved as following:

$$2 \cdot (30) - (29) = 4 + 3\lambda x_2 = 0, \tag{32}$$

resulting in the following expression for $x_2 = -\frac{4}{3\lambda}$. The next step is to find an expression for x_1. This is done by inserting the expression for x_2 in equation (29), which gives the expression for x_1 as

$$x_1 = \frac{2}{3\lambda} - \frac{3}{3\lambda} = -\frac{1}{3\lambda}. \tag{33}$$

In order to get the value of λ, the expressions of x_1 and x_2 are inserted into equation (31)

$$(\frac{-1}{3\lambda})^2 + (\frac{-1}{3\lambda})(\frac{-4}{3\lambda}) + (\frac{-4}{3\lambda})^2 - 21 = 0, \tag{34}$$

from which the value of λ can be calculated to $\lambda = \pm\frac{1}{3}$. Lastly, we can insert the value of λ in the obtained expression for x_1 and x_2, which result in the following points

$$\begin{cases} P_1 = (x_1, x_2)^T = (-1, -4)^T, \\ P_2 = (x_1, x_2)^T = (1, 4)^T \end{cases} \tag{35}$$

By inserting each point in the function from equation (26), the corresponding function values are calculated as

$$\begin{cases} f(P_1) = 1, \\ f(P_2) = 29. \end{cases} \tag{36}$$

It is clear that the minimum value of the function $f(x_1, x_2)$ corresponds to the point $P_1 = (-1, -4)^T$ with the function value $f(P_1) = 1$.

3 Problem 1.3

(a) For this problem the standard genetic algorithm (GA) that was written during the Matlab introduction was used, with some changes. The Matlab code in attached file 1.3 contains the main file FunctionOptimization.m as well as the function files InitializePopulation.m, DecodeChromosome.m, EvaluateIndividual.m, InsertBestIndividual.m, TournamentSelect.m., Cross.m. and Mutate.m. All the parameters are hardcoded in the mail file. By running the main file, the global minimum of the function with its corresponding x_1, x_2 values will be found, with the help of the functions in the other function files. The function file Mutate.m, for example, carries out mutation by looping through all the genes in a chromosome with probability p_{mut}, specified by a mutation probability variable in the main file.

By running the GA in the interval $x_1, x_2 \in [-10, 10]$ the location of the minimum $(x_1^*, x_2^*)^T = (0, -1)^T$ was found, which gives the global minimum value of the function $g(x_1^*, x_2^*) = 3$.

(b) By running the GA a 100 times, each lasting a 100 generations, with following mutation rates, $p_{mut} \in \{0.00, 0.02, 0.05, 0.10\}$, one get the median fitness values shown in the table below. The median fitness values are lowest when $p_{mut} = 0.00$,

Table 2: Mutation rates and median fitness values

p_{mut}	Median Fitness Value
0.00	0.029
0.02	0.333
0.05	0.332
0.10	0.327

which corresponds to the case when the GA performs poorly and no mutations occur. In the case where no mutations occur, the GA probably gets stuck in a local optimum, which inhibits the algorithm to continue and explains the low median fitness value. It is known from the course book that a typical optimal mutation rate value for a binary chromosome is $1/m$, where m is the chromosome length. In our case the chromosome length is 50, giving us the optimal mutation rate: $p_{mut} = 0.02$. From the result in table 2, it is clear that the GA performs best at the optimal mutation rate, where the algorithm succeeds to find the highest median fitness value. The results also show that for increasing mutation rates, the median fitness value is still high, however decreasing with increased mutation rate.

(c) To prove that the point $(x_1^*, x_2^*)^T = (0, -1)^T$ is a stationary point of the function $g(x_1^*, x_2^*)$, it's necessary to prove that $\nabla g(x_1^*, x_2^*) = (0, 0)$, which is done by calculating $\frac{\partial g}{\partial x_1} = 0$ and $\frac{\partial g}{\partial x_2} = 0$ in the point $(0, -1)^T$.

$$
\begin{aligned}
g(x_1,x_2) =&(1+(x_1+x_2+1)^2(19-14x_1+3x_1^2-14x_2+6x_1x_2+3x_2^2))\times \\
&(30+(2x_1-3x_2)^2(18-32x_1+12x_1^2+48x_2-36x_1x_2+27x_2^2)).
\end{aligned} \tag{37}
$$

To calculate the partial derivative of $g(x_1, x_2)$ the product rule can be used, by defining the function as

$$
g(x_1,x_2) = h(x_1,x_2)f(x_1,x_2), \tag{38}
$$

where $h(x_1, x_2)$ and $f(x_1, x_2)$ can be defined as

$$
\begin{cases}
h_1(x_1,x_2) = (x_1+x_2+1)^2 \\
h_2(x_1,x_2) = (19-14x_1+3x_1^2-14x_2+6x_1x_2+3x_2^2) \\
h(x_1,x_2) = 1+h_1(x_1,x_2)h_2(x_1,x_2),
\end{cases} \tag{39}
$$

$$
\begin{cases}
f_1(x_1,x_2) = (2x_1-3x_2)^2 \\
f_2(x_1,x_2) = 18-32x_1+12x_1^2+48x_2-36x_1x_2+27x_2^2 \\
f(x_1,x_2) = 30+f_1(x_1,x_2)f_2(x_1,x_2).
\end{cases} \tag{40}
$$

From the definition of the product rule, one get

$$
\nabla g(x_1,x_2) = \nabla h(x_1,x_2)f(x_1,x_2) + h(x_1,x_2)\nabla f(x_1,x_2), \tag{41}
$$

where

$$
\nabla h(x_1,x_2) = \nabla h_1(x_1,x_2)h_2(x_1,x_2) + h_1(x_1,x_2)\nabla h_2(x_1,x_2), \tag{42}
$$

$$
\nabla f(x_1,x_2) = \nabla f_1(x_1,x_2)f_2(x_1,x_2) + f_1(x_1,x_2)\nabla f_2(x_1,x_2). \tag{43}
$$

To solve equation (41), one needs to find the partial derivatives of the functions from equations (42) and (43).

$$
\begin{aligned}
\frac{\partial h_1}{\partial x_1} &= 2(x_1+x_2+1) & \qquad \frac{\partial h_1}{\partial x_2} &= 2(x_1+x_2+1) \\
\frac{\partial f_1}{\partial x_1} &= 4(2x_1-3x_2) & \qquad \frac{\partial f_1}{\partial x_2} &= -6(2x_1-3x_2) \\
\frac{\partial f_2}{\partial x_1} &= -32+24x_1-36x_2) & \qquad \frac{\partial f_2}{\partial x_2} &= 48-36x_1+54x_2
\end{aligned}
$$

.

Since $h_1(0,-1) = 0$ it's not necessary to compute $\nabla h_2(x_1, x_2)$, since it will be equal to zero according to equation (42). When inserting the point $(0,-1)^T$, one get the following values on the derivatives

$$
\begin{aligned}
\frac{\partial h_1}{\partial x_1} &= 0 & \qquad \frac{\partial h_1}{\partial x_2} &= 0 \\
\frac{\partial f_1}{\partial x_1} &= 12 & \qquad \frac{\partial f_1}{\partial x_2} &= -18 \\
\frac{\partial f_2}{\partial x_1} &= 4 & \qquad \frac{\partial f_2}{\partial x_2} &= -6
\end{aligned}
$$

.

Thereby, the values for (42) and (43) can be calculated to formulate the complete equation (41).

$$\frac{\partial h(0,-1)}{\partial x_1} = 0 \cdot h_2(0,-1) = 0, \qquad \frac{\partial h(0,-1)}{\partial x_2} = 0 \cdot h_2(0,-1) = 0,$$

it's also know that

$$\nabla h_2(0,-1) = 0. \tag{44}$$

The same thing is done for $f(x_1, x_2)$, resulting in

$$\frac{\partial f(0,-1)}{\partial x_1} = 12 \cdot (-3) + 9 \cdot 4 = 0, \quad \frac{\partial f(0,-1)}{\partial x_2} = -18 \cdot (-3) + 9 \cdot (-6) = 0.$$

This gives rise to the final expression of equation (41), as

$$\nabla g(x_1, x_2) = (0,0)^T \cdot f(x_1, x_2) + h(x_1, x_2) \cdot (0,0)^T, \tag{45}$$

resulting in the conclusion that $(x_1^*, x_2^*)^T = (0,-1)^T$ is a stationary point of $g(x_1^*, x_2^*)$.