Stochastic optimization algorithms 2022
Home problems, set 1

Name: Nicole Adamah
Personal id: 9905300662
Date: 2022/09/20

Problem 1.1

In this problem, we shall use the penalty method (see pp. 30-33 in the course book) to find the minimum of the function

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

subject to the constraint

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

1. Define (and specify clearly, in your report, as a function of x1, x2, and μ) the function fp(x; μ), consisting of the sum of f(x1, x2) and the penalty term.

Solution:

- **Starting from the objective function f(x), form the function:**

$$f_p\left(x;\mu\right) = f\left(x_1, x_2\right) + p\left(x_1, x_2;\mu\right) = \left(x_1 - 1\right)^2 + 2\left(x_2 - 2\right)^2 + \mu \cdot max\left(x_1^2 + x_2^2 - 1\right)^2 \tag{3}$$

- **Thus we obtain:**

$$f_p\left(x;\mu\right) = \begin{cases} \left(x_1 - 1\right)^2 + 2\left(x_2 - 2\right)^2 + \mu\left(x_1^2 + x_2^2 - 1^2\right) & x_1^2 + x_2^2 \geq 1 \\ \left(x_1 - 1\right)^2 + 2\left(x_2 - 2\right)^2 & \textbf{,Otherwise} \end{cases} \tag{4}$$

- **Now we can compute the gradient by taking the partial derivatives :**

$$\frac{\partial f_p}{\partial x_1} = \begin{cases} 2\left(x_1 - 1\right) + 4\mu x_1\left(x_1^2 + x_2 - 1^2\right) & x_1^2 + x_2^2 \geq 1 \\ 2\left(x_1 - 1\right) & \textbf{,Otherwise} \end{cases} \tag{5}$$

$$\frac{\partial f_p}{\partial x_2} = \begin{cases} 4\left(x_2 - 2\right) + 4\mu x_2\left(x_1^2 + x_2 - 1^2\right) & x_1^2 + x_2^2 \geq 1 \\ 4\left(x_2 - 2\right) & \textbf{,Otherwise} \end{cases} \tag{6}$$

2. *Next, compute (analytically) the gradient ∇fp(x; μ), 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.*

- **Now we can compute the gradient analytically :**

$$\nabla f_p(x;\mu) = \left(\frac{\partial f_p}{\partial x_1}, \frac{\partial f_p}{\partial x_2}\right)^T \tag{7}$$

• **Case: When the constraint is not satisfied:** $x_1^2 + x_2^2 \geq 1$

$$\frac{\partial f_p}{\partial x_1} = 2\,(x_1 - 1) + 4\mu x_1 \left(x_1^2 + x_2 - 1^2\right) \tag{8}$$

$$\frac{\partial f_p}{\partial x_2} = 4\,(x_2 - 2) + 4\mu x_2 \left(x_1^2 + x_2 - 1^2\right) \tag{9}$$

• **Case: When the constraint is satisfied:**

$$\frac{\partial f_p}{\partial x_1} = 2\,(x_1 - 1) \tag{10}$$

$$\frac{\partial f_p}{\partial x_2} = 4\,(x_2 - 2) \tag{11}$$

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

• By putting *μ = 0* we obtain the unconstrained minimum: $\Delta f_p(x;\mu = 0) = 0$ (12)

$$\frac{\partial f}{\partial x_1} = 2x_1 - 2 = 0 \longrightarrow x_1 = 1 \tag{13}$$

$$\frac{\partial f}{\partial x_2} = 4x_2 - 8 = 0 \longrightarrow x_2 = 2 \tag{14}$$

4. Write a Matlab program for solving the unconstrained problem of finding the minimum of fp(x; μ) using the method of gradient descent. You should make use of the skeleton files available on Canvas, namely

• **See attached code in map 1.1**

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: η = 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 x1 and x2 as functions of μ.

We can see from table 1 that x1 converges to 0,31 and x2 to 0,95 near μ=100 and μ=1000 and this can also be proven by looking at figure 1 and figure 2, where x1 and x2 are plotted as a function of μ.

Table 1: Finding the minimum

μ	x_1	x_2
1	0,4338	1,2102
10	0,3314	0,9955
100	0,3137	0,9553
1000	0,3118	0,9507

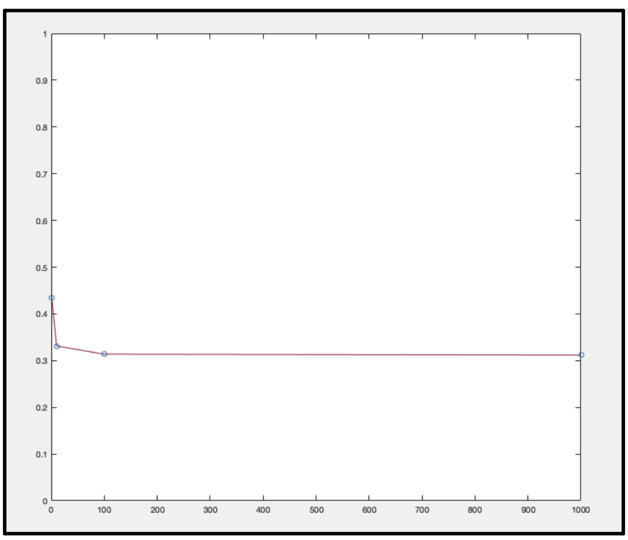

Figure 1.X1 plotted as a function of μ.

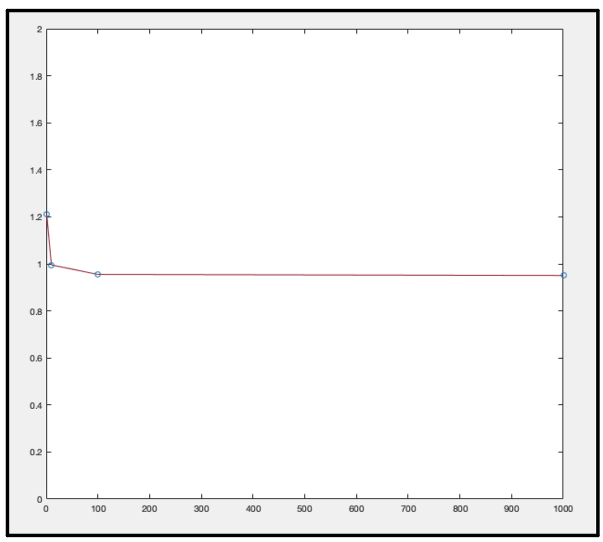

Figure 2. X2 plotted as a function of μ.

Problem 1.2

a) (2p) Use the analytical method described on pp. 29-30 in the course book to determine the global minimum

$(X_1^*, X_2^*)^T$ *(as well as the corresponding function value) of the function,*

$$f\left(x_1, x_2\right) = 4x_1^2 - x_1x_2 + 4x_2^2 - 6x_2 \tag{15}$$

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

b) (1p) Use the Lagrange multiplier method described on pp. 25-28 in the course book to determine the minimum

$(X_1^*, X_2^*)^T$ *(as well as the corresponding function value) of the function* $f\left(x_1x_2\right) = 15 + 2x_1 + 3x_2$ (16)

subject to the constraint. $h\left(x_1x_2\right) = x_1^2 + x_1x_2 + x_2^2 - 21 = 0$ (17)

Solution:

a)

- **Calculating the stationary points in the interior of S**

$$\nabla f = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}\right)^T \longrightarrow \begin{cases} \dfrac{\partial f}{\partial x_1} = 8x_1 - x_2 = 0 & (18) \\[2ex] \dfrac{\partial f}{\partial x_2} = -x_1 + 8x_2 - 6 = 0 & (19) \end{cases}$$

- **By solving eq. (18) & (19) we get the value of x1 & x2:**

$$x_2 = 8x_1 \longrightarrow -x_1 + 64x_1 - 6 = 0 \longrightarrow 63x_1 = 6$$

$$\longrightarrow \begin{cases} x_1 = \dfrac{6}{63} = \dfrac{2}{21} & (20) \\[2ex] x_2 = \dfrac{16}{21} & (21) \end{cases}$$

- **We then get the stationary point, that satisfies the boundary S:**

$$P_1 = \left(\frac{2}{21}, \frac{16}{21}\right) \tag{22}$$

$$P_1 \in S \quad 0 < \frac{2}{21}, \frac{16}{21} < 1 \ \& \frac{2}{21} < \frac{16}{21}$$

- **We then consider the boundary ∂S of S, which can be divided into four parts:**
- **Case 1, where:** $x_1 = 0, 0 < x_2 < 1$

$$f\left(0, x_2\right) = 4x_2 - 6 = 0 \tag{23}$$

$$f'\left(0, x_2\right) = 8x_2 - 6 = 0 \tag{24}$$

$$x_2 = \frac{6}{8} \rightarrow P_2 = \left(0, \frac{3}{4}\right) \tag{25}$$

- **Case 2, where:** $0 < x_1 < 1, x_2 = 1$

$$f\left(x_1, 1\right) = 4x_1^2 - x_1 - 2 \tag{26}$$

$$f'\left(x_1, 1\right) = 8x_1 - 1 = 0 \tag{27}$$

$$x_1 = \frac{1}{8} \rightarrow P_3 = \left(\frac{1}{8}, 1\right) \tag{28}$$

- **Case 3, where:** $x_1 = x_2 = x$

$$f\left(x, x\right) = 4x^2 = x^2 + 4x^2 - 6x = 7x^2 - 6x \tag{29}$$

$$f'\left(x, x\right) = 14x - 6 = 0 \tag{30}$$

$$x = \frac{6}{14} \rightarrow P_4 = \left(\frac{3}{7}, \frac{3}{7}\right) \tag{31}$$

- **Case 4, where evaluate the corners:**

$$P_5 = (0, 0) \tag{32}$$

$$P_6 = (0, 1) \tag{33}$$

$$P_7 = (1, 1) \tag{34}$$

- **We can now compute the function values to find the minimum of the seven points:**

$$P_1 = \left(\frac{2}{21}, \frac{16}{21}\right) \longrightarrow f\left(P_1\right) = -2,285$$

$$P_2 = \left(0, \frac{3}{4}\right) \longrightarrow f\left(P_2\right) = -2,250$$

$$P_3 = \left(\frac{1}{8}, 1\right) \longrightarrow f\left(P_3\right) = -2,062$$

$$P_4 = \left(\frac{3}{7}, \frac{3}{7}\right) \longrightarrow f\left(P_4\right) = -1,286$$

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

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

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

- **We can now see that the global minimum is -2,285 with the points** $P_1 = \left(\frac{2}{21}, \frac{16}{21}\right)$

b) solution: $f\left(x_1 x_2\right)=15+2x_1+3x_2$ (16) $h\left(x_1 x_2\right)=x_1^2+x_1x_2+x_2^2-21=0$ (17)

- **We start the calculations to find the minimum by defining Lagrange multiplier function:**

$$L\left(x_1, x_2 \lambda\right)=f\left(x_1, x_2\right)+\lambda h\left(x_1, x_2\right) \tag{35}$$

$$L\left(x_1, x_2, \lambda\right)=15+2x_1+3x_2+\lambda(x_1^2+x_1x_2+x_2^2-21=0) \tag{36}$$

$$\lambda = Lagrange\ multiplier$$

- **Now we can compute the gradient of L by taking the partial derivatives and setting the equations equal to zero:**

$$\frac{\partial L}{\partial x_1}=2+\lambda\left(2x_1+x_2\right)=0 \tag{37}$$

$$\frac{\partial L}{\partial x_2}=3+\lambda\left(x_1+2x_2\right)=0 \tag{38}$$

$$\frac{\partial L}{\partial \lambda}=x_1^2+x_1x_2+x_2^2-21=0 \tag{39}$$

- **The next step is to solve the equation system:**

$$eq.\,(37) \longrightarrow x_2=-\frac{2}{\lambda}-2x_1 \longrightarrow eq.\,(38)$$

$$3+\lambda(x_1+2\left(-\frac{2}{\lambda}-2x_1\right)=0 \longrightarrow \begin{cases} x_1=-\dfrac{1}{3\lambda} & (40) \\[2ex] x_2=-\dfrac{4}{3\lambda} & (41) \end{cases}$$

$eq.\,(40)$ & $eq.\,(41)$ In $eq.\,(39)$

$$\left(-\frac{1}{3\lambda}\right)^2+\left(-\frac{1}{3\lambda}\cdot-\frac{4}{3\lambda}\right)+\left(-\frac{4}{3\lambda}\right)^2-21=0 \longrightarrow \lambda=\pm\frac{1}{3} \tag{42}$$

- **We can now use the value on lambda to get the points and corresponding f(x):**

$$\lambda=\frac{1}{3} \longrightarrow \begin{cases} x_1=-1 \\ x_2=-4 \end{cases} \qquad \lambda=-\frac{1}{3} \longrightarrow \begin{cases} x_1=1 \\ x_2=4 \end{cases}$$

In $eq.\,(39) \longrightarrow L\left(x_1, x_2, \lambda\right)=1$ | In $eq.\,(39) \longrightarrow L\left(x_1, x_2, \lambda\right)=29$

- **We can conclude the minimum is obtained from the first case:**

$$\left(x_1^*, x_2^*\right)^T=(-1,-4)^T \quad \text{with} \quad f\left(x_1^*, x_2^*\right)=1$$

Problem 1.3

a) In this problem, you will implement a genetic algorithm (GA) for finding the minimum of the function

$$g(x_1, x_2) = (1,5 - x_1 + x_1 x_2)^2 + \left(2,25 - x_1 + x_1 x_2^2\right)^2 + \left(2,625 - x_i + x_1 x_2^3\right)^2 \quad (43)$$

in the range [−5, 5] (for both x1 and x2). Write a section Problem1.3 where, for part(a) above, you list the selected parameters (used in RunSingle.m) and also include a (neatly formatted, not a Matlab screen dump) table of the values of x1, x2, and (note!) g(x1, x2) found in your 10 runs. Then, for part (b), include another neatly formatted table showing the values of pmut and the median performance (of the GA) obtained over the corresponding batch runs. Also, include a plot of the median performance as a function of pmut. Then, include a brief discussion of your findings related to the optimal value of pmut for this problem. Finally, for part (c), write down the true minimum (x∗1, x∗2)T, as well as the entire derivation, with all relevant intermediate steps, showing that this point is indeed a stationary point. Omitting intermediate steps may result in the solution being returned, with resulting point deductions; see below

Solution: a)

- **The listed parameters are presented in table 2.**
- **The obtained values from the GA are presented in table 3.**

Table 2. Selected parameters

Population size	100
Max variable value	5
Number of genes	50
Number of variables	2
Tournament size	3
Tournament probability	0, 75
Crossover probability	0, 75
Mutation probability	0, 02
Number of generations	4000

Table 3. $x_1, x_2, g(x_1, x_2)$

	$g(x_1, x_2)$	x_1	x_2
1	0	$2,9999$	$0,4999$
2	$1 \cdot 10^{-4}$	$3,0291$	$0,5079$
3	$1 \cdot 10^{-5}$	$3,0066$	$0,5018$
4	0	$3,0005$	$0,5001$
5	$1 \cdot 10^{-5}$	$3,0021$	$0,5006$
6	0	$2,9998$	$0,4999$
7	0	$2,9998$	$0,4999$
8	$1 \cdot 10^{-5}$	$2,9980$	$0,4995$
9	$1 \cdot 10^{-5}$	$2,9927$	$0,4980$
10	0	$2,9998$	$0,4999$

b)

We can see from table 4 and figure 3 that the median performance is is better when the mutation probability is between 0.01 and 0.8. The conclusion that can be drawn is that higher mutation probability doesn’t necessarily correspond to a better performing GA. In most cases the mutation rate is set to constant and around 1/m where m is the chromosome length.

Table 4. Pmut and median performance

	p_{mut}	$Median$
1	0	$0,8116$
2	$0,01$	$0,9666$
3	$0,02$	$0,9977$
4	$0,04$	$0,9997$
5	$0,08$	$1,000$
6	$0,1$	$1,000$
7	$0,3$	$1,000$
8	$0,5$	$0,9984$
9	$0,7$	$0,9988$
10	1	$0,9445$

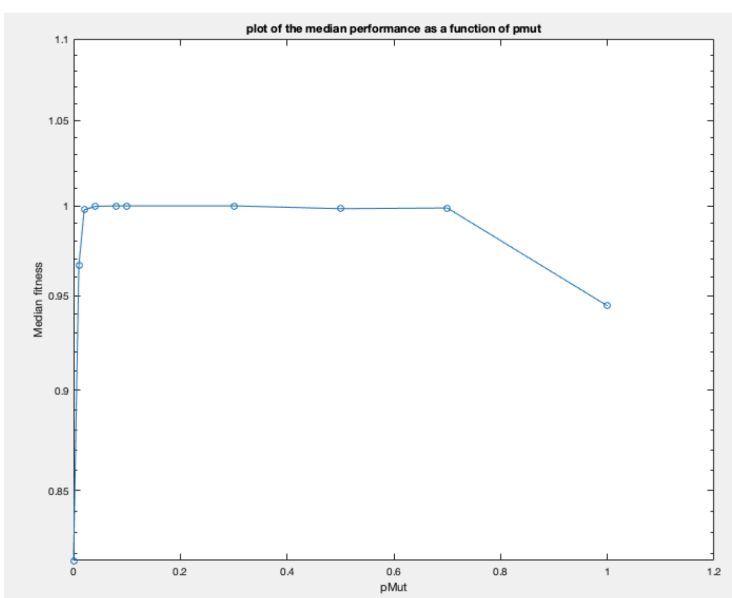

Figure 3. Median fitness plotted as a function of pMut.

c)

- **Given the result from the GA the minimum is most likely :**

$$(x_1^*, x_2^*)^T = (3, 0.5)^T$$

- **The first step to prove this is to take the partial derivatives of the function:**

$$\Delta g\left(x_1, x_2\right) = \left(\frac{\partial g}{\partial x_1}, \frac{\partial g}{\partial x_2}\right)^T \tag{44}$$

$$g\left(x_1, x_2\right) = \left(1,5 - x_1 + x_1 x_2\right)^2 + \left(2,25 - x_1 + x_1 x_2^2\right)^2 + \left(2,625 - x_1 + x_1 x_2^3\right)^2$$

- **The derivations are done with the sum/difference rule:** $(f \pm g)' = f' \pm g'$ (45)

$$\frac{\partial g}{\partial x_1} = 2x_2^6 x_1 + 2x_2^4 x_1 + 6x_1 - 4x_2^3 x_1 - 2x_2^2 x_1 - 4x_2 x_1 + 5,25x_2^3 + 4,5x_2^2 + 3x_2 - 12,75 + 0 \tag{46}$$

$$\frac{\partial g}{\partial x_1} = 2x_2^6 x_1 + 2x_2^4 x_1 + 6x_1 - 4x_2^3 x_1 - 2x_2^2 x_1 - 4x_2 x_1 + 5,25x_2^3 + 4,5x_2^2 + 3x_2 - 12,75 \tag{47}$$

$$\frac{\partial g}{\partial x_2} = 6x_1^2 x_2^5 + 4x_1^2 x_2^3 + 0 - 6{x_1}^2{x_2}^2 - 2{x_1}^2 x_2 - 2x_1 + 15,75x_1 x_2^2 + 9x_1 x_2 + 3x_1 - 0 + 0 \tag{48}$$

$$\frac{\partial g}{\partial x_2} = 6x_1^2 x_2^5 + 4x_1^2 x_2^3 - 6{x_1}^2{x_2}^2 - 2{x_1}^2 x_2 - 2x_1 + 15,75x_1 x_2^2 + 9x_1 x_2 + 3x_1 \tag{49}$$

- **We can now insert the points and if the result is 0 for the partial derivatives then the points are proven to be stationary:**

$$\frac{\partial g}{\partial x_1} = 2(0,5)^6(3) + 2(0,5)^4 3 + 6(3) - 4(0,5)^3(3) - 2(0,5)^2(3) - 4(0,5)(3) + 5,25(0,5)^3 + 4,5(0,5)^2 + 3(0,5) - 12,75 \tag{50}$$

$$\longrightarrow \quad \frac{\partial g}{\partial x_1} = 0$$

$$\frac{\partial g}{\partial x_2} = 6(3)^2(0,5)^5 + 4(3)^2(0,5)^3 - 6(3)^2(0,5)^2 - 2(3)^2(0,5) - 2(3) + 15,75(3)(0,5)^2 + 9(3)(0,5) + 3(3) \tag{51}$$

$$\longrightarrow \quad \frac{\partial g}{\partial x_2} = 0$$

- **We can now confirm that the assumed points are the true minimum:** $(x_1^*, x_2^*)^T = (3, 0.5)^T$