FFR105 - Homework Problem 1

Stochastic optimization algorithms

Felix Waldschock
000420-T398

Chalmers University of Technology
September 2023

Contents

1 Problem 1.1 - Penalty method **2**
- 1.1 Defining the function $(f_p(\underline{x};\mu)$ . . . 2
- 1.2 Calculating the gradient of $(f_p(\underline{x};\mu)$ . . . 2
  - 1.2.1 Case 1: constraint fulfilled (max term = 0) . . . 2
  - 1.2.2 Case 2: constraint unfulfilled (max term $\neq 0$) . . . 2
- 1.3 Find minimum for $f_p(\underline{x},\mu)$ unconstrained . . . 2
- 1.4 Execution of program and results . . . 3

2 Problem 1.2 - Constrained optimization **4**
- 2.1 Part a) . . . 4
- 2.2 Part b) . . . 6

3 Problem 1.3 - Basic GA program **7**
- 3.1 Part a) Varying parameters . . . 7
- 3.2 Part b) Varying mutation rate . . . 8
- 3.3 Part c) Estimate of stationary point . . . 9

1 Problem 1.1 - Penalty method

1.1 Defining the function $(f_p(\underline{x};\mu)$

As in Formula (2.67) in the course book:

$$f_p(\underline{x};\mu) = f(\underline{x}) + p(\underline{x};\mu) \tag{1}$$

where

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

and the penalty term

$$p(\underline{x}, \mu) = \mu(\sum_{i=1}^{m}(max\{g_i(\underline{x}, 0)\})^2 + \sum_{i=1}^{k}(h_i(\underline{x}))^2) \tag{3}$$

since no equality constrains are give ($k = 0$) the second sum term drops out and the following Equation (4) derives

$$f_p(\underline{x}, \mu) \ = (x_1 - 1)^2 + 2(x_2 - 2)^2 + \mu(\sum_{i=1}^{m}(max\{g_i(\underline{x}, 0)\})^2 \tag{4}$$

Furthermore, since g_i is of dimension 1 the sum cancels out as well and

$$g(x_1, x_2) = x_1^2 + x_2^2 - 1 \leq 0 \tag{5}$$

$$f_p(\underline{x}, \mu) \ = (x_1 - 1)^2 + 2(x_2 - 2)^2 + \mu(max\{x_1^2 + x_2^2 - 1, 0\})^2 \tag{6}$$

remains.

1.2 Calculating the gradient of $(f_p(\underline{x};\mu)$

Here a separation for the two cases of the max function need to be considered.

1.2.1 Case 1: constraint fulfilled (max term = 0)

$$\nabla f_p(x_1, x_2) = \big[2(x_1 - 1), 4(x_2 - 2)\big] \tag{7}$$

1.2.2 Case 2: constraint unfulfilled (max term $\neq 0$)

$$\nabla f_p(x_1, x_2) = \big[2(x_1 - 1) + 2x_1\mu(x_1^2 + x_2^2 - 1), 4(x_2 - 2) + 4x_2\mu(x_1^2 + x_2^2 - 1)\big] \tag{8}$$

1.3 Find minimum for $f_p(\underline{x}, \mu)$ unconstrained

Here the components of the gradient from Formula (7) are set to zero. This will return the vector $\underline{x}$ with two components.

$$\begin{aligned} \frac{\partial f}{\partial x_1} &= 2(x_1 - 1) \\ 2(x_1 - 1) &= 0 \\ x_1 &= 1 \end{aligned} \tag{9}$$

$$\begin{aligned} \frac{\partial f}{\partial x_2} &= 4(x_2 - 2) \\ 4(x_2 - 2) &= 0 \\ x_2 &= 2 \end{aligned} \tag{10}$$

And therefore $\underline{x} = (1, 2)$. This point will be used as the starting point for the gradient decent algorithm.

1.4 Execution of program and results

To run the program the suggested parameters where chosen

- $\mu = [1, 10, 100, 1000]$
- $\eta = 0.0001$
- $gradientTolerance = 1e - 6$

With these settings, solutions for the constrained function are found at $\underline{x}^*$ as shown in the Table (1).

μ	x_1^*	x_2^*
1	0.4337	1.2101
10	0.3313	0.9955
100	0.3137	0.9552
1000	0.3117	0.9507

Table 1: Solutions of x_1^* and x_2^* for different iteration steps μ

When plotting the values against the iteration steps μ as in Figure (1) it can be seen that after the first 100 iterations, the found solutions is already very close to the found one at 1000 iterations. Therefore one can say, that in this case it is converging well to a solution.

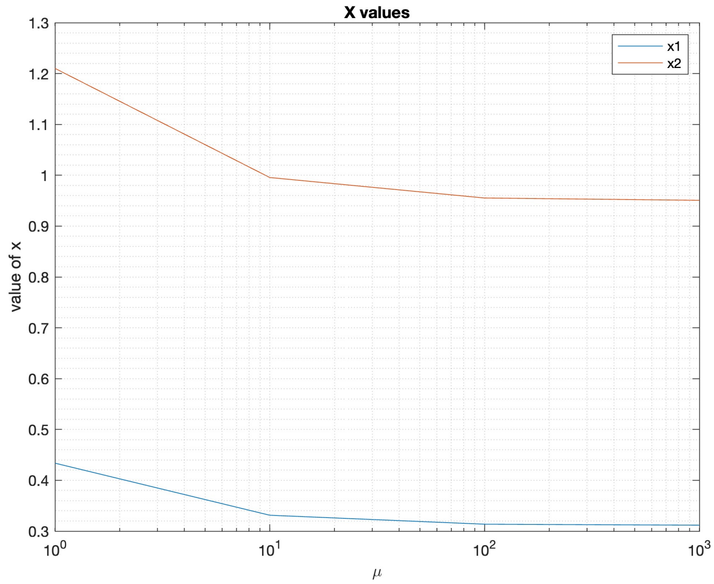


Figure 1: Solutions of x_1 and x_2 plotted against the iteration steps (X-axis logarithmic scaling) μ

2 Problem 1.2 - Constrained optimization

2.1 Part a)

To find the stationary point of the function $f(x_1, x_2)$ for

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

the partial derivatives are determined

$$\frac{\partial f}{\partial x_1} = 8x_1 - x_2 \tag{12}$$

$$\frac{\partial f}{\partial x_x} = -x_1 + 8x_2 - 6 \tag{13}$$

setting Formula (12) and (13) zero

$$8x_1 = x_2$$

$$x_1 = 8x_2 - 6$$

the point of a stationary point lays at $(\frac{2}{21}, \frac{16}{21})$, Formula(14).

$$\begin{aligned} x_2 &= \frac{16}{21} \\ x_1 &= \frac{2}{21} \end{aligned} \tag{14}$$

Furthermore the set S needs to be investigated. This for the three segments

1. Line 1 $(0,0) -> (0,1)$
2. Line 2 $(0,1) -> (1,1)$
3. Line 3 $(0,0) -> (1,1)$

For Line 1

$$\begin{aligned} &x_1 = 0, 0 \leq x_2 \leq 1 \\ &\hat{f}_1(0, x_2) = 4x_2^2 - 6x_2 \\ &\frac{\partial \hat{f}_1}{\partial x_2} = 8x_2 - 6 \\ &x_2 = \frac{3}{4} \end{aligned} \tag{15}$$

For Line 2

$$\begin{aligned} &0 \leq x_1 \leq 1, x_2 = 1 \\ &\hat{f}_2(x_1, 1) = 4x_1^2 - x_1 + 4 - 6 \\ &\frac{\partial \hat{f}_2}{\partial x_1} = 8x_1 - 1 \\ &x_1 = \frac{1}{8} \end{aligned} \tag{16}$$

For Line 3

$$\begin{aligned} &x_1 = x_2 \\ &\hat{f}_3(x) = 4x^2 - x^2 + 4x^2 - 6x = 7x^2 - 6x \\ &\frac{\partial \hat{f}_3}{\partial x_2} = 14x - 6 \\ &x = \frac{6}{14} \end{aligned} \tag{17}$$

Evaluating the functions $\hat{f}_n$ with the found point of interest the following points for minima are found

$$\begin{aligned} &\hat{f}_1(0, \frac{3}{4}) = -2.25 \\ &\hat{f}_2(\frac{1}{8}, 1) = -2.06 \\ &\hat{f}_3(\frac{6}{14}) = -1.28 \end{aligned} \tag{18}$$

When comparing the found stationary points in (18) with (14) it can be seen that the minimum lays in the solutions of Formula (14) at position $(\frac{2}{21}, \frac{16}{21})$.

2.2 Part b)

To solve the problem with the Lagrange multiplier method the Formula (2.50) of the course book is used:

$$L(x_1, x_2, \lambda) = f(x_1, x_2) + \lambda h(x_1, x_2) \tag{19}$$

Since the gradient shall be zero, it can be said:

$$\nabla f(x_1, x_2) = \lambda \nabla h(x_1, x_2) \tag{20}$$

With the two functions:

$$f(x_1, x_2) = 15 + 2x_1 + 3x_2 \tag{21}$$

$$h(x_1, x_2) = x_1^2 + x_1 x_2 + x_2^2 - 21 = 0 \tag{22}$$

and their gradients

$$\nabla f(x_1, x_2) = [2, 3] \tag{23}$$

$$\nabla h(x_1, x_2) = [2x_1 + x_2, x_1 + 2x_2] \tag{24}$$

This system of equations derives:

$$\begin{aligned} 2 &= (2x_1 + x_2)\lambda \\ 3 &= (x_1 + 2x_2)\lambda \\ x_1^2 &+ x_1 x_2 + x_2^2 - 21 = 0 \end{aligned} \tag{25}$$

Solving the system of equations (25) two candidates for a minimum are found:

$$[\{\lambda = -\frac{1}{3}, x_1 = -1, x_2 = -4\}, \{\lambda = \frac{1}{3}, x_1 = 1, x_2 = 4\}] \tag{26}$$

Now the function $f(x_1, x_2)$ (21) is evaluated at both points

$$\begin{aligned} f(1, 4) &= 29 \\ f(-1, -4) &= 1 \end{aligned} \tag{27}$$

It can be said that the point $(-1, -4)$ is the one with a minimum.

3 Problem 1.3 - Basic GA program

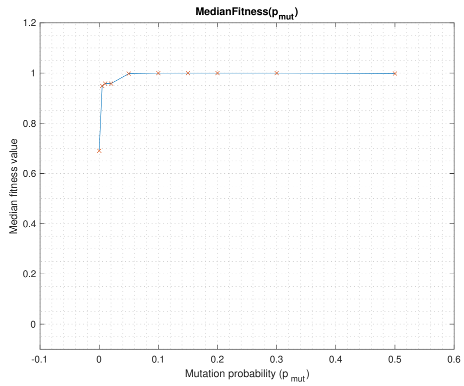
3.1 Part a) Varying parameters

This subsection contains the 10 different parameter settings which were used in the *Run-Single.m* file. This experiment shall help understanding the impact of the parameters.

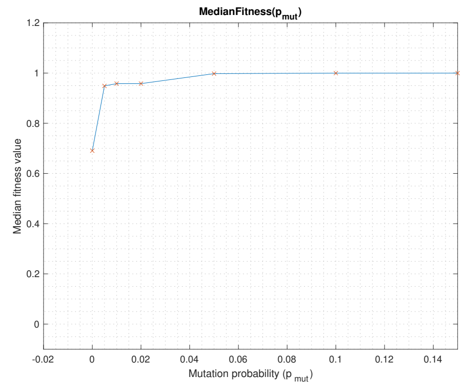
tourSize	p_{Tour}	p_{Cross}	p_{Mut}	numGeneration	Fitness	x_1	x_2	$g(x_1, x_2)$
2	0.75	0.8	0.02	2000	0.9997	2.9585	0.4882	0.0003
4	0.75	0.8	0.02	2000	0.9581	3.75	0.6384	0.0438
10	0.75	0.8	0.02	2000	0.958	3.75	0.6385	0.0438
4	0.5	0.8	0.02	2000	0.958	3.75	0.6385	0.0438
4	0.1	0.8	0.02	2000	0.9322	2.4999	0.35	0.0727
4	0.75	0.3	0.02	2000	0.9976	2.8906	0.4687	0.0023
4	0.75	0.3	0.4	2000	0.9999	2.987	0.4972	0.0000
4	0.75	0.3	0.4	200	0.9991	3.0281	0.5011	0.0009
10	0.9	0.5	0.4	1000	0.9999	3.0013	0.4992	0.0000
10	0.9	0.5	1	1000	0.589	1.9625	0.2788	0.6979

Table 2: Parameters and corresponding outputs for a single run, tourSize = size of tournament, pTour = tournament probability, pCross = probability of crossing, pMut = probability of mutation, numGeneration = number of generations

From the found values for x_1 and x_2 in the Table (2) it can be assumed that the solution lays at $[3, 1]$. When comparing the first three lines, it can be seen, that only an increase of the tournament size results in worse found solutions. Furthermore, comparing the second with the 6th line, one can see that a reduction of p_{Cross} reaches far better results. In regard of computation efforts the line 7 and 8 show remarkable information. The number of generations was reduced by the factor 10, with it the computation amount, and the results are relatively close to each other. In the last line, p_{mut} was set to 1. Therefore, every gene was mutated, as expected the GA performs very bad. One could say it is not able to learn from the good individuals, and the reach result is just random.



(a) Full range of tested p_{mut} displayed



(b) Zoomed into area $p_{mut} < 0.15$

Figure 2: Median fitness values as function of the mutation probability p_{mut}

3.2 Part b) Varying mutation rate

As discussed in the lecture, mutation helps the algorithm to "escape" from local minima and not get stuck. In this subsection the results for the set of mutation rates $p_{mut} = [0, 0.005, 0.01, 0.02, 0.05, 0.1, 0.15, 0.2, 0.3, 0.5]$ get compared. As seen in Table 3 the first run, executed with $p_{mut} = 0$, shows that if the GA is not supported by any mutation, a relatively low median is reached. Already with a mutation probability of 0.005 a median value of 95% is reached.

p_{mut}	median value
0	0.6906
0.005	0.9488
0.01	0.9579
0.02	0.9580
0.05	0.9977
0.1	0.9998
0.15	0.9999
0.2	0.9999
0.3	0.9999
0.5	0.9984

Table 3: Median value for the 10 iterations with mutation probability $= [0, 0.005, 0.01, 0.02, 0.05, 0.1, 0.15, 0.2, 0.3, 0.5]$

The median fitness values can also be seen as a function of the mutation probability in the Figure (2). It is clear to see, that to the point $p_{mut} = 0.02$ the median fitness values are increasing. After the point $p_{mut} > 0.02$ the median values converge around 0.999. A slight decrease can be observed at $p_{mut} = 0.5$. As discussed in the course FAQ, a peak at around $\frac{1}{numberOfGenes}$ could be expected. Since finding an optimum for the problem is rather simple, the GA performs still very well at way higher mutation rates than $\frac{1}{numberOfGenes} = 0.02$.

3.3 Part c) Estimate of stationary point

From the evaluation in 3.2 the best found variables are $[2.9805, 0.4992]$, therefore my guess for the stationary point is at $[3, 0.5]$. To prove that this is a stationary point, the gradient of the function $g(x_1, x_2)$ is derived and evaluated at the point $[3, 0.5]$, the Equation (28) needs to be fulfilled.

$$\nabla g(x_1, x_2) = [\frac{\partial g}{\partial x_1}, \frac{\partial g}{\partial x_1}] = [0, 0] \tag{28}$$

First the gradient is derived:

$$\frac{\partial g}{\partial x_1} = 2(1.5 - x_1 + x_1 x_2)(x_2 - 1) + 2(2.25 - x_1 + x_1 x_2^2)(x_2^2 - 1) + 2(2.625 - x_1 + x_1 x_2^3)(x_2^3 - 1) \tag{29}$$

$$\frac{\partial g}{\partial x_2} = 2(1.5 - x_1 + x_1 x_2)(x_1) + 2(2.25 - x_1 + x_1 x_2^2)(2 x_1 x_2) + 2(2.625 - x_1 + x_1 x_2^3)(3 x_1 x_2^2) \tag{30}$$

When evaluating the terms in the brackets, one can see they result in zero for the point $[3, 0.5]$

$$\begin{gathered} \underbrace{(1.5 - 3 + 3 * 0.5)}_{=0} \\ (2.25 - 3 + 3 * 0.25) = \underbrace{(3 - 3)}_{=0} \\ \underbrace{(2.625 - 3 + 3 * 0.125)}_{=0} \end{gathered} \tag{31}$$

Since these bracket terms show up in both partials, both of them cancel out to zero. Therefore it can be said, when evaluating Equation (28) at the point $[3, 0.5]$ the solution is $[0, 0]$ and that the point is a stationary one.

List of Figures

1 Solutions of x_1 and x_2 for different iteration steps 4
2 Median fitness values as function of the mutation probability p_{mut} 8

List of Tables

1 Solutions of x_1^* and x_2^* for different iteration steps μ 3
2 Parameters and corresponding outputs for a single run 7
3 Median value for 10 iterations 8