First homework assignment

Jurek Eisinger [*]

Stochastic optimization algorithms

September 19, 2023
Student ID (GU): 8508561538

Problem 1.1 - Penalty Method

Find the minimum of

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

subject to:

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

1

Define the function $f_P(\boldsymbol{x}, \mu)$.

Generally, we know that:

$$f_P(\boldsymbol{x}, \mu) = f(\boldsymbol{x}) + p(\boldsymbol{x}, \mu)$$

Where this function is to be minimized "without constraints". The constraining nature is captured in:

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

where $h = 0$ here. So we have for $x_1^2 + x_2^2 > 1$:

$$f_P(\boldsymbol{x}, \mu) = (x_1 - 1)^2 + 2(x_2 - 2)^2 + \mu(x_1^2 + x_2^2 - 1)^2$$

And else:

$$f_P(\boldsymbol{x}, \mu) = (x_1 - 1)^2 + 2(x_2 - 2)^2 = f(\boldsymbol{x})$$

[*] jurek.jonathan.eisinger@vub.be

2.

Compute the gradient.

If the constraints are not fulfilled:

$$\nabla f_P = \Big(2(x_1 - 1) + 4\mu x_1(x_1^2 + x_2^2 - 1) \quad , \quad 4(x_2 - 2) + 4\mu x_2(x_1^2 + x_2^2 - 1)\Big)^{\mathrm{T}}$$

If the constraints are fulfilled:

$$\nabla f_P = \Big(2(x_1 - 1) \quad , \quad 4(x_2 - 2)\Big)^{\mathrm{T}}$$

3.

Find the unconstrained minimum.

Unconstrained minimum is obtained by setting the gradient equal to zero and implying $\mu = 0$. So we find the minimum to be at

$$x_1 = 1$$

and

$$x_2 = 2.$$

We shall use the point

$$\boldsymbol{x} = \Big(1, 2\Big)^{\mathrm{T}}$$

as starting point for the gradient descent.

4.

Write a Matlab program for solving the unconstrained problem of finding the minimum of $f_P(\boldsymbol{x}, \mu)$ using the method of gradient descent.

We run the matlab code for the following values of μ:

```
mu = [0.1 1 10 100 1000]
```

and we obtain the following minima for the five different values of μ.

μ	x_1^*	x_2^*
0.1	0.7194	1.6389
1	0.4338	1.2102
10	0.3314	0.9955
100	0.3137	0.9553
1000	0.3118	0.9507

Table 1: Values of μ, x_1^*, and x_2^*

Given this table, we see a clear tendency of these values to converge, as is also visible in the plot shown in figure 1, where the x-axis is logarithmic for displaying purposes.

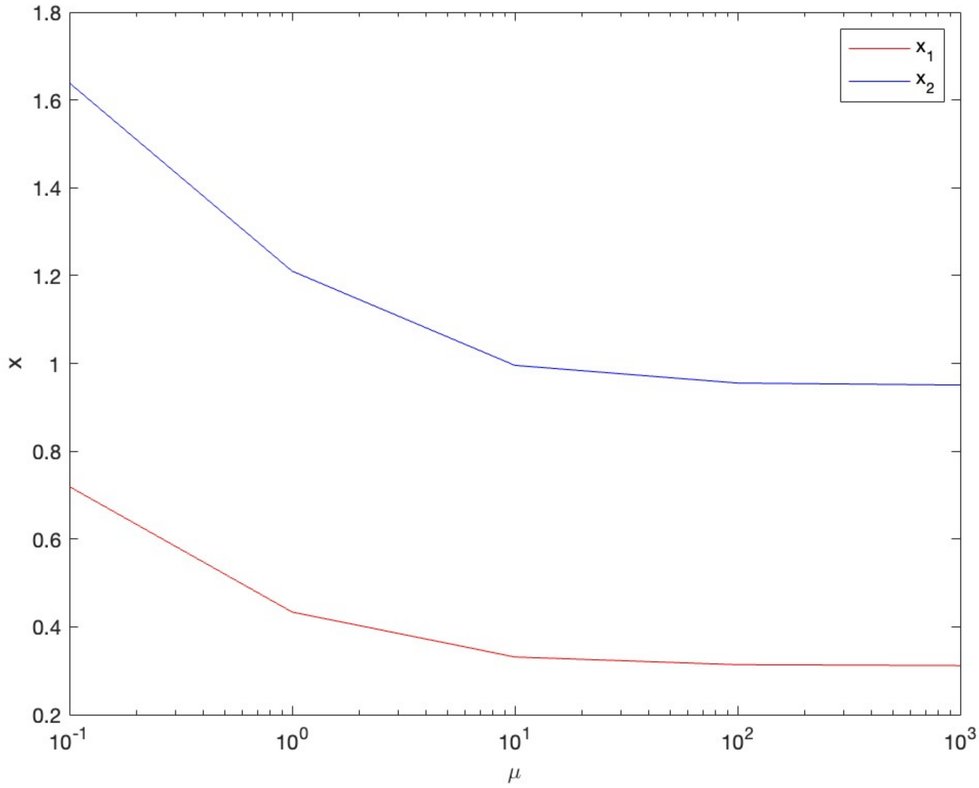

Figure 1: Convergence of variables as μ increases

Problem 1.2 - Constrained Optimization

Find global minimum of

$$f(\boldsymbol{x}) = 4x_1^2 - x_1x_2 + 4x_2^2 - 6x_2$$

this problem is very similar to the problem done in class. Find stationary point of f inside of the set S defined in the exercise. We find for the gradient:

$$\nabla f = \Big(8x_1 - x_2, \quad -x_1 + 8x_2 - 6\Big)^T$$

so that:

$$x_1 = \frac{1}{8}x_2$$

and

$$x_1 = 8x_2 - 6$$

So we find:

$$x_2 = \frac{16}{21}, \quad x_1 = \frac{2}{21}$$

or:

$$\boldsymbol{x}^* = (2/21, 16/21)$$

which lies inside the set S. Is this a global minimum? We will see. Let's take a look at the boundaries.
Firstly, the left boundary where $x_2 = 0$. Upon implying this, f looks like:

$$f = 4x_1^2$$

which gives the minimum:

$$\boldsymbol{x}^* = (0, 0)$$

Secondly, look at the upper boundary, where $x_1 = 0$. Here, we have:

$$f = 4x_2^2 - 6x_2$$

so that, because $\frac{\partial f}{\partial x_2} = 8x_2 - 6$, we find that:

$$\boldsymbol{x}^* = (0, 6/8)$$

Thirdly, look at the boundary line where $x_1 = x_2 = x$, so that:

$$f(\boldsymbol{x}) = 4x^2 - x^2 + 4x^2 - 6x = 7x^2 - 6x$$

so that, because $\frac{\partial f}{\partial x} = 14x - 6$, we have:

$$\boldsymbol{x}^* = (3/7, 3/7)$$

Moreover, there's three corners: $(0,0)$, $(0,1)$ and $(1,1)$, that are candidates for minima as well. What is left to do is evaluate the function at all the points that are potential candidates for minima. We find that the point

$$\boldsymbol{x}^* = (2/21, 16/21)$$

is the minimum, since $f(\boldsymbol{x}^*) = -2.286$ is minimal there.

b)

Find the minimum of

$$f(\boldsymbol{x}) = 15 + 2x_1 + 3x_2$$

subject to

$$h(\boldsymbol{x}) = x_1^2 + x_1x_2 + x_2^2 - 21 = 0$$

We minimize:

$$L = f(\boldsymbol{x}) + \lambda h(\boldsymbol{x}) = 15 + 2x_1 + 3x_2 + \lambda\Big(x_1^2 + x_1x_2 + x_2^2 - 21\Big)$$

Upon determining the partial derivatives and setting the three terms equal to zero, we find:

1. $$\frac{\partial L}{\partial x_1} = 2 + \lambda(2x_1 + x_2) = 0$$

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

3. $$\frac{\partial L}{\partial \lambda} = h(\boldsymbol{x}) = x_1^2 + x_1x_2 + x_2^2 - 21 = 0$$

We find, denoting the first two equations (1) and (2), that when solving (1) and (2) for λ and setting them equal to each other:

$$\begin{aligned}\frac{-2}{2x_1 + x_2} &= \frac{-3}{x1 + 2x_2} \\ &\implies 2x_1 + 4x_2 = 6x_1 + 3x_2 \\ &\implies x_2 = 4x_1\end{aligned} \tag{1}$$

so that, using $x_1 = \frac{-x_2}{2} \pm \sqrt{\frac{-3x_2^2}{4} + 21}$ (from constraint equation):

$$\begin{aligned} x_2 &= -2x_2 \pm 4\sqrt{\frac{-3x_2^2}{4} + 21} \\ \implies 3 &= \pm 4\sqrt{\frac{-3}{4} + 21/x_2^2} \end{aligned} \tag{2}$$

which yields:

$$x_2 = \pm 4 \tag{3}$$

We find that there's two possible candidates for minima;

$$\boldsymbol{x}^* = (-1, -4), \qquad \boldsymbol{x}^* = (1, 4).$$

Obviously, we then find the minimum to be at the point

$$\boldsymbol{x}^* = (-1, -4)$$

where $f(\boldsymbol{x}^*) = 1$.

Problem 1.3 - Basic GA programming

The task is finding the minimum of:

$$g(\boldsymbol{x}) = (1.5 - x_1 + x_1 x_2)^2 + (2.25 - x_1 + x_1 x_2^2)^2 + (2.625 - x_1 + x_1 x_2^3)^2$$

in the range $[-5, 5]$

a)

After trying out different values, to create the values in table 2, I used the default values:

- `tournamentSize` $= 2$
- `tournamentProbability` $= 0.75$
- `crossoverProbability` $= 0.8$
- `mutationProbability` $= 0.02$
- `numberOfGenerations` $= 2000$

Run	x_1	x_2	$g(x_1, x_2)$
1	3.0000116825	0.5000029653	2.195×10^{-11}
2	3.0000307560	0.5000077337	1.516×10^{-10}
3	2.9999875426	0.4999970049	2.504×10^{-11}
4	3.0000498295	0.5000125021	3.978×10^{-10}
5	2.9999717474	0.4999922365	1.410×10^{-10}
6	2.9999708533	0.4999922365	1.426×10^{-10}
7	3.0000307560	0.5000077337	1.516×10^{-10}
8	3.0000307560	0.5000077337	1.516×10^{-10}
9	2.9999613166	0.4999898523	2.466×10^{-10}
10	3.0000155568	0.5000038594	3.873×10^{-11}

Table 2: Table of Values

We see that these values are close to the values:

$$x_1 = 3.0 \qquad x_2 = 0.5 \qquad g(x_1, x_2) = 0 \tag{4}$$

which may turn out to be the minimum of the function.

b)

The different values for the median performance of the GA with varying mutation probability p_{mut} are shown in table 3. And the development of the performance with the mutation probability is depicted in figure 2. We clearly see that the performance is maximal for the value of $p = 0.02$, which is 1/ the number of genes (50). The rule of thumb we learned in class turns out to apply in this example very nicely. What is also intersting is that, after a minimum at around $\mu = 0.7$ the median performance over 100 runs increases again. Also interesting is that, when taking a look at table 3, we see that a probability of zero yields by

p_{mut}	Median Performance
0	0.992640
0.01	0.999838
0.02	0.999999
0.04	0.999999
0.07	0.999977
0.1	0.999960
0.3	0.999399
0.7	0.998703
0.8	0.998803
0.9	0.999587

Table 3: Table of p_{mut} and Median Performance

far the worst Median Performance. This shows that introducing the concept of mutation into a genetic algorithm can be a tool to increase performance.

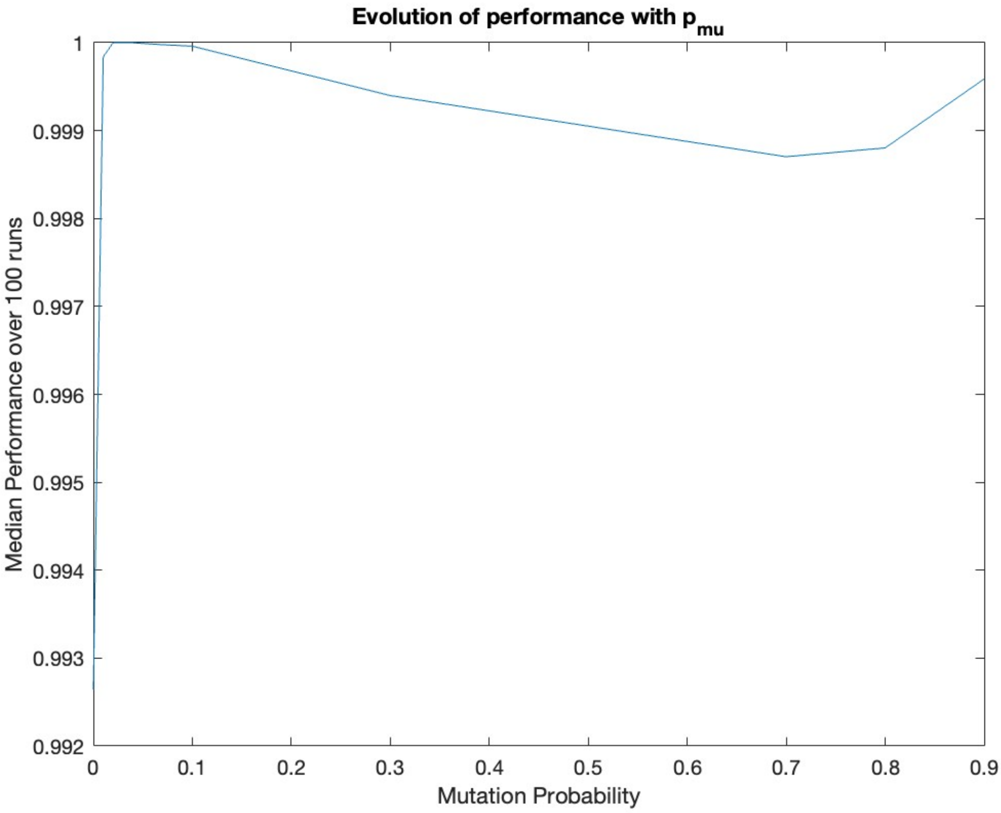

Figure 2: Development of the performance with increasing values of the mutation probability

c)

Educated guess from the results given in **a)**:

$$\boldsymbol{x}^* = (3, 0.5)$$

Actual calculation or proof that this is an actual stationary point found by:

$$\nabla g(\boldsymbol{x}) = \boldsymbol{0}$$

So:

$$\begin{aligned}\frac{\partial g(\boldsymbol{x})}{\partial x_1} &= 2(1.5 - x_1 + x_1 x_2) \cdot (-1 + x2) + 2(2.25 - x_1 + x_1 x_2^2)(-1 + x_2^2) \\ &+ 2(2.625 - x_1 + x_1 x_2^3)(-1 + x_2^3) = 0 \\ \frac{\partial g(\boldsymbol{x})}{\partial x_2} &= 2(1.5 - x_1 + x_1 x_2) x_1 + 4(2.25 - x_1 + x_1 x_2^2) x_1 x_2 \\ &+ 6(2.625 - x_1 + x_1 x_2^3) x_1 x_2^2 = 0\end{aligned} \tag{5}$$

upon plugging in the educated guess from above $\boldsymbol{x}^* = (3, 0.5)$, so

$$x_1^* = 3.0 \qquad x_2^* = 0.5$$

for x_1, x_2 in the two equations in 5, we find that $\nabla g(\boldsymbol{x})\Big|_{\boldsymbol{x}=\boldsymbol{x}^*} = 0$, so that this is indeed a stationary point. Furthermore, in fact

$$g(\boldsymbol{x}^*) = 0$$

as found in a).