CHALMERS UNIVERSITY OF TECHNOLOGY

FFR105 - Stochastic optimization algorithms

David Kastö

950207-8554

Home problem 1

Tuesday 25th September, 2018

Problem 1.1 - Penalty method

The aim was 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

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

To that end, the penalty method discussed in Section 2.4.2.3 of Biologically Inspired Optimization Methods: An Introduction by Wahde[1] was used to find a proper starting point for gradient descent. The penalty method applies the penalty function

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

Here μ is a positive parameter, g_i the i:th out of m inequality constraints (like the one in (2)) and h_i the i:th out of k equality constraints.

The initially described problem is then equivalent to minimizing (with respect to $\boldsymbol{x}$)

$$f_p(\boldsymbol{x};\mu) = f(\boldsymbol{x}) + p(\boldsymbol{x};\mu) = (x_1-1)^2 + 2(x_2-2)^2 + \mu\left(\max\{(x_1^2+x_2^2-1),0\}\right)^2. \quad (3)$$

Calculating the gradient of (3) analytically, one obtains

$$\nabla f_p(\boldsymbol{x};\mu) = \begin{cases} \left(\frac{\partial f_p(x_1,x_2;\mu)}{\partial x_1},\ \frac{\partial f_p(x_1,x_2;\mu)}{\partial x_2}\right), & x_1^2+x_2^2-1>0 \\ (2(x_1-1),\ 4(x_2-2)), & \text{otherwise,} \end{cases}$$

where

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

Setting $\mu = 0$, the unconstrained minimum can be found by setting $\nabla f_p(\boldsymbol{x};0) = \mathbf{0}$ and solve for x_1 and x_2, i.e.

$$\frac{\partial f_p(x_1,x_2;0)}{\partial x_1} = 2(x_1-1) = 0 \implies x_1 = 1 \quad (4)$$
$$\frac{\partial f_p(x_1,x_2;0)}{\partial x_2} = 4(x_2-2) = 0 \implies x_2 = 2. \quad (5)$$

By considering (1), it can easily be seen that $f(x_1,x_2)$ cannot be negative since it consists of a sum of two squared expressions. The only case where it assumes the value zero is described by (4) and (5), and thus we can safely conclude that the obtained stationary point is a minimum.

To solve the unconstrained problem of finding the minimum of $f_p(\boldsymbol{x},\mu)$, a program implementing the gradient descent method was written in Matlab. The program consists of three scripts; the main function PenaltyMethod (returns a table equivalent to Table 1 when it is run) and the two auxiliary functions GradientDescent (carries out gradient descent until the convergence criteria, see below, is fulfilled), and Gradient (returns ∇f_p). Using the position of the minimum of $f(x_1,x_2)$ ((4) and (5)) as the starting point, the step length $\eta = 0.0001$, and the convergence threshold $T = 10^{-6}$ (i.e. the algorithm stopped when $|\nabla f_p(\boldsymbol{x},\mu)|$ went below T) for each of the μ-values 1, 10, 100, 1000, the points for the optimum displayed in Table 1 were obtained. Seemingly, the obtained points in the table converge for large values of μ. This is expected, since it can be shown that (under certain conditions) the minimum of $f_p(\boldsymbol{x};\mu)$ approaches the minimum of $f(\boldsymbol{x})$ subject to the constraints as μ increases[1].

$\boldsymbol{\mu}$	$\boldsymbol{x_1^*}$	$\boldsymbol{x_2^*}$
1	0.434	1.210
10	0.331	0.996
100	0.314	0.955
1000	0.312	0.951

Table 1: The positions in the x_1 and x_2 direction for the optimum of (3) for four different values of μ, obtained through gradient descent. It appears that the obtained points in the table converge for large values of μ as expected.

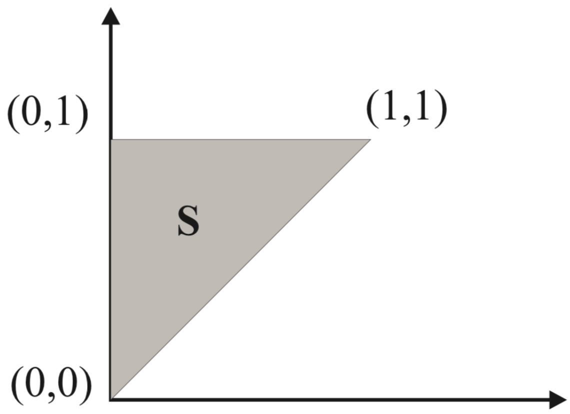

Figure 1: The set **S** used in Problem 1.2a. Adopted from [2].

Problem 1.2 - Constrained optimization

a)

To find the global minimum $(x_1^*, x_2^*)^{\mathrm{T}}$ and the corresponding function value of

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

on the closed set **S**, displayed in Figure 1, one may exploit the fact that it is a function under inequality constraints that defines a compact set **S**. Thus, it is enough to investigate all the stationary points of $f(\boldsymbol{x})$ in the interior of **S** and of the restriction of $f(\boldsymbol{x})$ to the boundary $\partial\mathbf{S}$ of **S** (including corner points)[1].

Starting with the interior of $\mathbf{S}$, evaluation of the gradient yields

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

Setting it equal to the zero vector renders two equations

$$\frac{\partial f(\boldsymbol{x})}{\partial x_1} = 0 \implies x_2 = 8x_1 \tag{7}$$

$$\frac{\partial f(\boldsymbol{x})}{\partial x_2} = 0 \implies \{(7)\} \implies 63x_1 = 6 \implies \begin{cases} x_1 = \frac{2}{21} \\ x_2 = 8x_1 = \frac{16}{21} \end{cases} \tag{8}$$

Since partial derivatives of $f(\boldsymbol{x})$ exist everywhere on $\mathbf{S}$, the only remaining part to consider is the boundary $\partial\mathbf{S}$. Considering one edge of the triangle at a time results in three cases:

$$\begin{cases} f(0,x_2) = 4x_2^2 - 6x_2 \implies \{\frac{\mathrm{d}f}{\mathrm{d}x_2} = 8x_2 - 6 = 0\} \implies x_2 = \frac{3}{4}, (x_1 = 0) \\ f(x_1,1) = 4x_1^2 - x_1 - 2 \implies \{\frac{\mathrm{d}f}{\mathrm{d}x_1} = 8x_1 - 1 = 0\} \implies x_1 = \frac{1}{8}, (x_2 = 1) \\ f(x_1,x_1) = 7x_1^2 - 6x_1 \implies \{\frac{\mathrm{d}f}{\mathrm{d}x_1} = 14x_1 - 6 = 0\} \implies x_1 = \frac{3}{7}, (x_2 = \frac{3}{7}) \end{cases} \tag{9}$$

Plugging each and every of the four sets of points obtained in (8) and (9) and also the three corners of the triangle ((0,0), (0,1) and (1,1)) into (6), Table 2 was obtained and it became apparent that $(x_1^*, x_2^*)^{\mathrm{T}} = (\frac{2}{21}, \frac{16}{21})$ was the global minimum and that $f(\frac{2}{21}, \frac{16}{21}) \approx -2.286$ was the corresponding minimal value of the function.

x_1	x_2	$f(x_1, x_2)$
0	0	0
0	1	-2.000
1	1	1.000
2/21	16/21	-2.286
0	3/4	-2.250
1/8	1	-2.063
3/7	3/7	-1.286

Table 2: The function values of (6) in the seven points that were candidates for the global minimum. As seen in the table, $(x_1^*, x_2^*)^{\mathrm{T}} = (\frac{2}{21}, \frac{16}{21})^{\mathrm{T}}$ corresponds to the global minimum.

b)

In order to determine the minimum $(x_1^*, x_2^*)^{\mathrm{T}}$ and the corresponding function value of

$$f(x_1, x_2) = 15 + 2x_1 + 3x_2 \tag{10}$$

subject to the constraint

$$h(x_1, x_2) = x_1^2 + x_1x_2 + x_2^2 - 21 = 0, \tag{11}$$

the Lagrange multiplier method was applied. Consequently, the function L is formed

$$\begin{aligned} L(x_1, x_2, \lambda) &= f(x_1, x_2) + \lambda h(x_1, x_2) \\ &= 15 + 2x_1 + 3x_2 + \lambda(x_1^2 + x_1x_2 + x_2^2 - 21). \end{aligned}$$

To find the stationary points, the gradient of L is set to the zero vector, i.e. one gets three equations

$$\frac{\partial L}{\partial x_1} = 2 + \lambda(2x_1 + x_2) = 0 \tag{12}$$
$$\frac{\partial L}{\partial x_2} = 3 + \lambda(x_1 + 2x_2) = 0 \tag{13}$$
$$\frac{\partial L}{\partial \lambda} = h(x_1, x_2) = x_1^2 + x_1x_2 + x_2^2 - 21 = 0. \tag{14}$$

From (12) the relation $\lambda = -2/(2x_1 + x_2)$ can be obtained, which in turn can be inserted into (13):

$$\begin{aligned} &3 + \lambda(x_1 + 2x_2) = 3 + \frac{-2}{(2x_1 + x_2)}(x_1 + 2x_2) = 0 \\ &\implies x_1 + 2x_2 = \frac{3}{2}(2x_1 + x_2) \implies \frac{x_2}{2} = 2x_1 \\ &\implies x_2 = 4x_1. \end{aligned}$$

Plugging the last expression into (14), one finally arrives at

$$\begin{aligned} &x_1^2 + x_1x_2 + x_2^2 - 21 = x_1^2 + 4x_1^2 + 16x_1^2 - 21 = 0 \\ &\implies 21x_1^2 = 21 \implies x_1 = \pm 1 \\ &\implies \{x_2 = 4x_1\} \implies x_2 = \pm 4. \end{aligned}$$

Now, since (10) solely consists of a sum of constant and linear terms, it is evident that f will take on its smallest value for the case where both x_1 and x_2 are negative, i.e. for $(x_1^*, x_2^*)^{\mathrm{T}} = (-1, -4)^{\mathrm{T}}$. Thus, the minimal function value is $f(x_1^*, x_2^*) = f(-1, -4) = 15 + 2 \cdot (-1) + 3 \cdot (-4) = 1$.

Problem 1.3 - Basic GA program

a)

A genetic algorithm was implemented in a Matlab program as specified in [2], with the objective to find the position and value of the global minimum of the function

$$g(x_1, x_2) = \Big(1 + (x_1 + x_2 + 1)^2(19 - 14x_1 + 3x_1^2 - 14x_2 + 6x_1x_2 + 3x_2^2)\Big) \times \Big(30 + (2x_1 - 3x_2)^2(18 - 32x_1 + 12x_1^2 + 48x_2 - 36x_1x_2 + 27x_2^2)\Big) \tag{15}$$

in the interval $x_1, x_2 \in [-10, 10]$. To this end, the fitness function $f = 1/g(x_1, x_2)$ was chosen. The chromosome length was set to 50, the population size to 100, the crossover probability to 0.8, the number of generations to 100, the mutation probability to one divided by the number of genes (i.e. 1/50), the tournament size to 2, and the tournament selection parameter to 0.75. In terms of elitism, a single copy of the best individual of each generation was automatically transferred to the next generation.

The program consists of the main script FunctionOptimization.m and multiple subfunctions in separate files in the same folder. When the former is run, it prints the location and function value at the acquired minimum of $g(x_1, x_2)$. Using the parameters stated above (hardcoded in FunctionOptimization), the global minimum $(x_1^*, x_2^*) \approx (-8 \cdot 10^{-5}, -1.00016) \approx (0, -1)$ with the corresponding function value $g(x_1^*, x_2^*) \sim 3.0000$ was obtained.

b)

In this subtask a parameter search was conducted for the mutation rate μ. 100 runs, each lasting 100 generations, were made for the mutation rates 0.00, 0.02 (i.e. one divided by the number of genes), 0.05, and 0.10. All other parameters were kept the same as in **1.3.a)**. The median fitness value obtained for each value of the mutation rate is displayed in Table 3.

Mutation rate μ	Median fitness value f
0	0.0899
0.02	0.333
0.05	0.332
0.1	0.317

Table 3: The median fitness value obtained for different values of the mutation rate, each subject to 100 runs lasting 100 generations each.

Assuming that the obtained function value in **1.3.a)**, corresponding to the fitness value $f = 1/3$, was the actual global minimum (at least it is a stationary point, see **1.3.c)**), it seems that without any mutations the algorithm did not get enough "genetic material" to work with, such that it never ended up in the region of the global minimum. Too much mutations, on the other hand, appears to make the search excessively random and the "driving evolutionary force" too weak, so that the search did not manage to converge completely before the run was finished. Out of these four mutation rate values, the rule of thumb value[1] one divided by the number of genes gave the smallest function value, albeit not by much in comparison to slightly higher values of μ.

c)

In order to show that

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

does in fact have a stationary point in the location obtained in **1.3.a)**, $(x_1^*, x_2^*) \approx (0, -1)$, it is convenient to first introduce a couple of variables for brevity reasons:

$$
\begin{aligned}
a_1 &= (x_1 + x_2 + 1)^2 \\
a_2 &= (19 - 14x_1 + 3x_1^2 - 14x_2 + 6x_1x_2 + 3x_2^2) \\
a &= \Big(1 + (x_1 + x_2 + 1)^2(19 - 14x_1 + 3x_1^2 - 14x_2 + 6x_1x_2 + 3x_2^2)\Big) \\
&= 1 + a_1a_2 \\
b_1 &= (2x_1 - 3x_2)^2 \\
b_2 &= (18 - 32x_1 + 12x_1^2 + 48x_2 - 36x_1x_2 + 27x_2^2) \\
b &= \Big(30 + (2x_1 - 3x_2)^2(18 - 32x_1 + 12x_1^2 + 48x_2 - 36x_1x_2 + 27x_2^2)\Big) \\
&= 30 + b_1b_2.
\end{aligned}
$$

Using the notation above, g can now be written as $g = ab$ and the partial derivatives become

$$\frac{\partial g}{\partial x_1} = a\frac{\partial b}{\partial x_1} + \frac{\partial a}{\partial x_1}b \tag{16}$$

$$\frac{\partial g}{\partial x_2} = a\frac{\partial b}{\partial x_2} + \frac{\partial a}{\partial x_2}b. \tag{17}$$

Evaluation of the partial derivatives on the RHS of (16) and (17) yields

$$\begin{aligned}
\frac{\partial a}{\partial x_1} &= \frac{\partial a_1}{\partial x_1}a_2 + a_1\frac{\partial a_2}{\partial x_1} = 2\left(x_1 + x_2 + 1\right)a_2 + a_1\left(-14 + 6x_1 + 6x_2\right)\\
\frac{\partial b}{\partial x_1} &= \frac{\partial b_1}{\partial x_1}b_2 + b_1\frac{\partial b_2}{\partial x_1} = 4\left(2x_1 - 3x_2\right)b_2 + b_1\left(-32 + 24x_1 - 36x_2\right)\\
\frac{\partial a}{\partial x_2} &= \frac{\partial a_1}{\partial x_2}a_2 + a_1\frac{\partial a_2}{\partial x_2} = 2\left(x_1 + x_2 + 1\right)a_2 + a_1\left(-14 + 6x_1 + 6x_2\right)\\
\frac{\partial b}{\partial x_2} &= \frac{\partial b_1}{\partial x_2}b_2 + b_1\frac{\partial b_2}{\partial x_2} = -6\left(2x_1 - 3x_2\right)b_2 + b_1\left(48 - 36x_1 + 54x_2\right)
\end{aligned}$$

Insertion of these partial derivatives into (16) and (17) then gives the following expressions for the two elements of the gradient of g:

$$\begin{aligned}
\frac{\partial g}{\partial x_1} &= a\left(4\left(2x_1 - 3x_2\right)b_2 + b_1\left(-32 + 24x_1 - 36x_2\right)\right)\\
&\quad + \left(2\left(x_1 + x_2 + 1\right)a_2 + a_1\left(-14 + 6x_1 + 6x_2\right)\right)b\\
\frac{\partial g}{\partial x_2} &= a\left(-6\left(2x_1 - 3x_2\right)b_2 + b_1\left(48 - 36x_1 + 54x_2\right)\right)\\
&\quad + \left(2\left(x_1 + x_2 + 1\right)a_2 + a_1\left(-14 + 6x_1 + 6x_2\right)\right)b.
\end{aligned}$$

Now it is time to evaluate these two expressions in $(x_1^*, x_2^*) = (0, -1)$ (by first calculating the values of the auxiliary variables in this point), which results in

$$
\begin{aligned}
a_1(0,-1) &= (0 + -1 + 1)^2 = 0 \\
a_2(0,-1) &= (19 - 0 + 0 + 14 + 0 + 3) = 36 \\
a(0,-1) &= (1 + 0 \cdot 36) = 1 \\
b_1(0,-1) &= (0 + 3)^2 = 9 \\
b_2(0,-1) &= (18 - 0 + 0 - 48 - 0 + 27) = -3 \\
b(0,-1) &= 30 + 9 \cdot -3 = 3. \\
\frac{\partial g(0,-1)}{\partial x_1} &= 1\,(4\,(0+3)\,(-3) + 9\,(-32 + 0 + 36)) \\
&\quad + (2\,(0 - 1 + 1)\,36 + 0\,(-14 + 0 - 6))\,3 \\
&= -36 + 36 + 0 + 0 = 0 \\
\frac{\partial g(0,-1)}{\partial x_2} &= 1\,(-6\,(0+3)\,(-3) + 9\,(48 - 0 - 54)) \\
&\quad + (2\,(0 - 1 + 1)\,36 + 0\,(-14 + 0 - 6))\,3 \\
&= 54 - 54 + 0 + 0 = 0.
\end{aligned}
$$

Hence, since both of the partial derivatives are equal to zero in this point, $\nabla g(0, -1)$ is equal to the zero vector and thus $(0, -1)$ corresponds to a stationary point.

References

[1] Wahde, M., *Biologically Inspired Optimization Methods: An Introduction.* WIT Press, 2008.

[2] Wahde, M., *Stochastic optimization algorithms 2018 - Home problems, set 1.*