Home problem 1

FFR105 Stochastic optimization algorithms

Axel Johansson
axejoh@student.chalmers.se
990305-0715

20 Sept 2022

Solution proposal

Problem 1.1 - Penalty method

Solutions for problem 1.1 and it's subproblems is presented here.

1.1.1

Given the function

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

subject to the constraint

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

With the penalty term

$$p(\mathbf{x};\mu) = \mu\left(\sum_{i=1}^{m}(\max\{g_i(\mathbf{x}),o\})^2 + \sum_{i=1}^{k}(h_i(\mathbf{x}))^2\right), \tag{3}$$

where $p(\mathbf{x};\mu) = 0$ only when the constraint $g(x_1,x_2) \leq 0$ is satisfied. One can define the function

$$f_p(\boldsymbol{x};\mu) = f(\boldsymbol{x}) + p(\boldsymbol{x};\mu) \tag{4}$$

which consists of the function and the penalty term.

1.1.2

The gradient of the function $f_p(\boldsymbol{x};\mu)$ when the constraint is satisfied is given as

$$\nabla f_p(\boldsymbol{x};\mu) = \begin{bmatrix} 2(x_1-1) \\ 4(x_2-2) \end{bmatrix}. \tag{5}$$

When the constraint is not satisfied the penalty term contributes and the gradient is given

$$\nabla f_p(\boldsymbol{x};\mu) = \begin{bmatrix} 2(x_1-1) + 4x_1\mu(x_1^2+x_2^2-1) \\ 4(x_2-2) + 4x_2\mu(x_1^2+x_2^2-1) \end{bmatrix}. \tag{6}$$

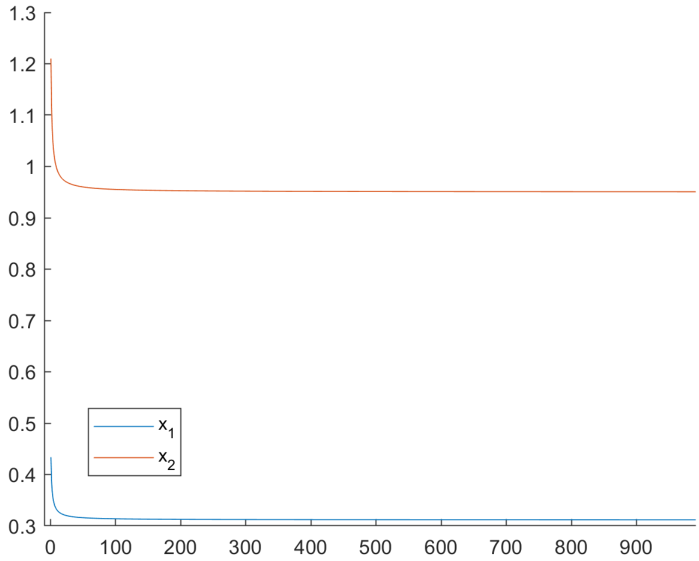

Figure 1: *The figure shows a clear convergence of x_1 and x_2 from gradient descent with $\mu \in [0,1000]$ on the x-axis and parameters: step length $\eta = 10^{-4}$ and tolerance $T = 10^{-6}$.*

1.1.3

One can analytically calculate the unconstrained minimum by letting the gradient of the function equal the zero vector and $\mu = 0$

$$\nabla f_p(\boldsymbol{x};\mu) = \begin{bmatrix} 2(x_1 - 1) \\ 4(x_2 - 2) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \implies \begin{bmatrix} x_1 = 1 \\ x_2 = 2 \end{bmatrix}. \tag{7}$$

Here $x^* = [1,2]^T$ is the point for the unconstrained minimum which will be used to carry out gradient descent with different μ.

1.1.5

The gradient descent computed for $\mu = [0,1,10,100,1000]$ with step length eta $\eta = 10^{-4}$ and threshold $T = 10^{-6}$ can be seen in table 1. In the table 1 one can see that the distance between points

Table 1: Table over the minimums of the function $f_p(\boldsymbol{x};\mu)$ for different μ and eta $\eta = 10^{-4}$ and $T = 10^{-6}$.

μ	x_1^*	x_2^*
0	1.0000	2.0000
1	0.4338	1.2102
10	0.3313	0.9955
100	0.3138	0.9553
1000	0.3118	0.9507

decreases less and less as μ increases. And its even more clear in figure 1.

Problem 1.2 - Constrained optimization

Given the function

$$f(x_1,x_2) = 4x_1^2 - x_1x_2 + 4x_2^2 - 6x_2. \tag{8}$$

Where the the domain is defined as the closed set S of the triangle with corners in (0, 0), (0, 1) and (1, 1). To find the global global minimum $(x_1^*,x_2^*)^T$ and corresponding function value of the function on S one can use the Lagrange multiplier method. This done by first computing points for local minimums on the surface, edges, corners and then comparing the function values to decide the global minimum. But first compute the gradient

$$\nabla f(x_1,x_2) = \begin{bmatrix} 8x_1 - x_2 \\ 8x_2 - x_1 - 6 \end{bmatrix}. \tag{9}$$

Surface

Compute if there is a local minimum on the surface by letting

$$\nabla f(x_1,x_2) = \vec{0} \Longrightarrow \begin{bmatrix} x_1 = \frac{2}{21} \\ x_2 = \frac{16}{21} \end{bmatrix} \Longrightarrow \vec{P_1} = \left(\frac{2}{21}, \frac{16}{21}\right). \tag{10}$$

Note that $\vec{P_1} \in S$.

Edges

The edges is done one by one. First look at the edge where $x_1 = 0$ and $x_2 \in [0,1]$,

$$\nabla f(0,x_2) = 0 \text{ for } \Longrightarrow \vec{P_2} = (0, \frac{6}{8}). \tag{11}$$

Secondly let's look at the edge where $x_2 = 1$ and $x_1 \in [0,1]$,

$$\nabla f(x_1,1) = 0 \Longrightarrow \vec{P_3} = (\frac{1}{8}, 1). \tag{12}$$

And thirdly let $x_1 = x_2 = x \in [0,1]$,

$$\nabla f(x,x) = 0 \Longrightarrow \vec{P_4} = (\frac{3}{7}, \frac{3}{7}). \tag{13}$$

Corners

The three corners of the set S is $\vec{P_5} = (0,0)$, $\vec{P_6} = (0,1)$ and $\vec{P_7} = (1,1)$.

Comparison

Given these points one can find the global minimum by inserting in the function (8).

$$\text{Surface: } f(\vec{P_1}) = -\frac{16}{7}. \tag{14}$$

$$\text{Edges: } \begin{matrix} f(\vec{P_2}) = -\frac{9}{4} \\ f(\vec{P_3}) = -\frac{33}{16} \\ f(\vec{P_4}) = -\frac{9}{7} \end{matrix}. \tag{15}$$

$$\text{Corners: } \begin{matrix} f(\vec{P_5}) = -\frac{9}{4} \\ f(\vec{P_6}) = -\frac{33}{16} \\ f(\vec{P_7}) = -\frac{9}{7} \end{matrix}. \tag{16}$$

One can see that $f(\vec{P_1}) = -\frac{16}{7}$ is the smallest and thus has a global minimum in $(x_1^*,x_2^*)^T = \left(\frac{2}{21}, \frac{16}{21}\right)^T$.

1.3 - Genetic Algorithm

a)

In tabular 2 the different parameters that were used to run "RunSingle.m" is presented. The program was run ten times and the result is presented in tabular 3.

Table 2: The tabular presents the changeable parameters that were used to run "RunSingle.m".

Parameter name	Value
tournamentSize	2
tournamentProbability	0.75
crossoverProbability	0.8
mutationProbability	0.02
numberOfGenerations	2000

Table 3: Table with x_1, x_2 from ten runs and corresponding function value $g(x_1,x_2)$. As can be seen the function is very small which gives a high fitness from equation $1/(1+g(x_1,x_2))$.

x_1	x_2	$g(x_1,x_2)$
3.0000459552	0.5000116080	3.3905702670e-10
3.0000307560	0.5000077337	1.5166713642e-10
2.9999976754	0.4999993891	8.9234323501e-13
2.9999908209	0.4999993891	7.7437459716e-11
2.9999988675	0.4999996871	2.2901676647e-13
2.9999964833	0.4999990910	2.0110342298e-12
2.9999988675	0.4999996871	2.2901676647e-13
3.0000021458	0.5000005811	7.9239225091e-13
2.9999923110	0.4999981970	9.7106612929e-12
3.0000689030	0.5000169724	7.6014465843e-10

b)

From figure 2 one can see that the fitness as a function of the mutation rate reaches a maximum when P_{mut}=1/(number of genes). A higher mutation rate seems to lead to worse performance. This is because after a while the individuals get highly adapted to the environment and therefore a random change is not likely to be positive. Also when the mutation probability is around 50% the new genes are basically created randomly, removing the impact of tournament select and crossover. And a completely random algorithm is not likely to be better which is seen again in the figure. But giving new material to the algorithm/nature will eventually pay dividends and thus a small but non zero mutation rate, around 1/(number of genes), is desirable. With that in mind the fitness seems to reach a local maximum when $P_{mut} \in [93{,}98]\%$ as can be seen in 2.

To check the legitimacy of this algorithmic findings one can compute the optimal mutation rate. Let m be the number of genes and l be the number of zeors.

$$P(l,P_{mut}) = (1-P_{mut})^{m-l}\cdot(1-(1-P_{mut}))^{l} = \{x = 1-P_{mut}\} = (x)^{m-l}\cdot(1-x)^{l} \equiv \prod(l,x). \quad (17)$$

To find the maximum let the derivative equal to zero which provides with two solutions. First solution

$$\prod(l,x) = 0 \implies x = 0 \implies P_{mut} = 1. \quad (18)$$

The second solution

$$\prod(l,x) = 0 \implies (m-1)x^{-l} = m \implies x = \sqrt[l]{1-\frac{l}{m}} \approx \{l \ll m\} \approx 1 - \frac{l}{m}\frac{1}{l} = 1 - \frac{1}{m}. \quad (19)$$

Table 4: Table of the median fitness over 100 runs for each mutation probability. The same parameters as in tabular 2 are used. Note that a mutation probability around 1.5-2.5% gives a fitness value very close to 1.

P_{mut}	fitness
0	0.991868085509962
0.005	0.997820730461261
0.01	0.999838164935663
0.015	0.999999996987517
0.02	0.999999964786222
0.025	0.999999959373140
0.03	0.999999833634615
0.3	0.999021748923845
0.4	0.998660745893594
0.5	0.998344931219959
0.6	0.998223452111308
0.85	0.999243192267959
0.9	0.999492975890554
0.925	0.999762840915092
0.95	0.999831364613678
0.97	0.999808093154126
0.98	0.999699565276780
0.985	0.999339666810507
0.99	0.999171433969267
0.995	0.996893790116966
1	0.989379025669266

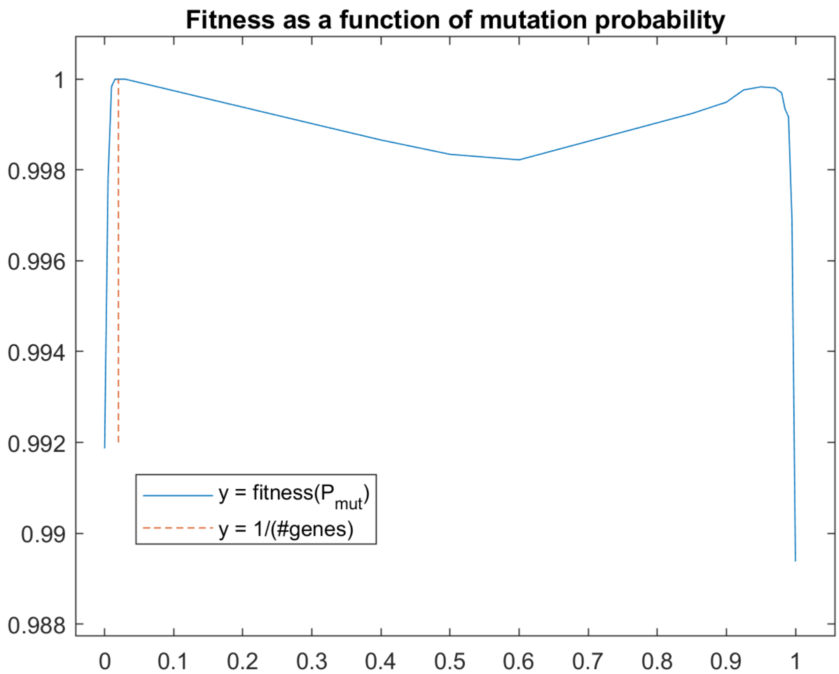

Figure 2: *The figure is generated from the data in tabular 4 and shows how the median fitness over 100 runs varies with* P_{mut}. *The red dotted line is drawn at one over the number of genes i.e.* $P_{mut} = 0.02$.

Which gives $P_{mut} = 1/m$ where m is the number of genes. This further strengthens the findings presented in figure 2. From equation (18) one can see that there is a theoretical optimal mutation probability close to one. This is also found in figure 2. However this solution is discarded since a high mutation probability could be dangerous and thus not common in nature.

c)

Guided from subsection 1.3 a) and tabular 3 an educated guess is that the true minimum occurs at $x_1^* = 3$ and $x_2^* = 0.5$. To verify this one can compute $g'(x_1,x_2) = 0$ analytically. But given this educated guess one can instead check if the guess minimizes the function. Given

$$g(x_1,x_2) = (1.5 - x_1 + x_1x_2)^2 + (2.25 - x_1 + x_1x_2^2)^2 + (2.625 - x_1 + x_1x_2^3)^2 \tag{20}$$

and the guess that $x_1^* = 3$ and $x_2^* = 0.5$ one should compute

$$g(3,0.5) = (1.5 - 3 + 3\cdot 0.5)^2 + (2.25 - 3 + 3\cdot 0.5^2)^2 + (2.625 - 3 + 3\cdot 0.5^3)^2 = 0^2 + 0^2 + 0^2 = 0. \tag{21}$$

So the educated guess proves to minimize the function since $g(x_1,x_2) \geq 0$ and therefore $(x_1^*, x_2^*)^T = (3,0.5)^T$ is a stationary point for the global minimum of the function. But the entire derivation is asked for so here it is..

$$\frac{\partial}{\partial x_1} g(x_1,x_2) = 2(1.5 - x_1 + x_1x_2)(x_2 - 1) + 2(2.25 - x_1 + x_1x_2^2)(x_2^2 - 1) + 2(2.625 - x_1 + x_1x_2^3)(x_2^3 - 1). \tag{22}$$

Insert point

$$\begin{aligned}\frac{\partial}{\partial x_1} g(3,0.5) &= 2(1.5 - 3 + 3\cdot 0.5)(3 - 1) + 2(2.25 - 3 + 3\cdot 0.5^2)(0.5^2 - 1) + 2(2.625 - 3 + 3\cdot 0.5^3)(0.5^3 - 1) \\ &= 2(0)(2) + 2(0)(-0.75) + 2(0)(-0.875) = 0 + 0 + 0 = 0.\end{aligned} \tag{23}$$

Now with respect to x_2

$$\frac{\partial}{\partial x_2} g(x_1,x_2) = 2(1.5 - x_1 + x_1x_2)(x_1) + 2(2.25 - x_1 + x_1x_2^2)(2x_1x_2) + 2(2.625 - x_1 + x_1x_2^3)(3x_1x_2^2). \tag{24}$$

Insert point

$$\begin{aligned}\frac{\partial}{\partial x_2} g(3,0.5) &= 2(1.5 - 3 + 3\cdot 0.5)(3) + 2(2.25 - 3 + 3\cdot 0.5^2)(2\cdot 3\cdot 0.5) + 2(2.625 - 3 + 3\cdot 0.5^3)(3\cdot 3\cdot 0.5^2) \\ &= 2(0)(3) + 2(0)(3) + 2(0)(9/4) = 0 + 0 + 0 = 0.\end{aligned} \tag{25}$$

Here one can see that the educated guess satisfies $\nabla g = 0$ and is therefore a stationary point.