Stochastic optimization algorithms 2020
Home problems, set 1

David Tonderski (davton)
199812101773

Problem 1.1, Penalty method

The task is to use the penalty method 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

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

1 The function $f_p(\mathbf{x}; \mu)$ is defined as follows:

$$f_p(\mathbf{x}; \mu) = \begin{cases} (x_1 - 1)^2 + 2(x_2 - 2)^2 + \mu(x_1^2 + x_2^2 - 1)^2 & \text{if } x_1^2 + x_2^2 - 1 > 0, \\ (x_1 - 1)^2 + 2(x_2 - 2)^2 & \text{otherwise.} \end{cases} \tag{3}$$

2 Next, we compute the gradient. We compute $\frac{\partial f_p}{\partial x_1}$, $\frac{\partial f_p}{\partial x_2}$ and $\frac{\partial f_p}{\mu}$ for both of the above cases to get[1]

$$\nabla f_p(\mathbf{x}; \mu) = \begin{cases} \begin{bmatrix} 2(x_1 - 1) + 4\mu x_1(x_1^2 + x_2^2 - 1) \\ 4(x_2 - 2) + 4\mu x_2(x_1^2 + x_2^2 - 1) \\ (x_1^2 + x_2^2 - 1)^2 \end{bmatrix} & \text{if } (x_1^2 + x_2^2 - 1) > 0, \\ \\ \begin{bmatrix} 2(x_1 - 1) \\ 4(x_2 - 2) \\ 0 \end{bmatrix} & \text{otherwise.} \end{cases} \tag{4}$$

3 Because the function $f(x_1, x_2)$ consists of a sum of two non-negative terms, the minimum must occur when both terms are zero. Therefore, the minimum is found at $x_1 = 1, x_2 = 2$.

[1] Note that $\frac{\partial f_p}{\mu}$ is not used in the program.

4,5 The parameter values suggested in the instructions are used, i.e. $\eta = 0.0001$, $T = 10^{-6}$ and $\mu \in \{1, 10, 100, 1000\}$. The program's output is shown in table 1.

Table 1: The obtained values of $\mathbf{x}^*$ for the given values of μ.

μ	x_1^*	x_2^*
1	0.434	1.210
10	0.331	0.996
100	0.314	0.955
1000	0.312	0.951

The results seem to be reasonable, as they seem to converge to a point.

Problem 1.2, Constrained optimization

a) The task is to find the minimum of the function

$$f(x_1, x_2) = 4x_1^2 - x_1x_2 + 4x_2^2 - 6x_2 \tag{5}$$

on the closed set S, which is a triangle with corners in (0,0), (0,1) and (1,1). First, the gradient of the function should be computed to find stationary points. Then, the boundaries of the S should be checked, and lastly, the corners of S should be checked.

Let's start with the gradient. We have

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

Setting $\nabla f(x_1, x_2) = 0$ gives a simple system of equation with the solution $x_1 = \frac{2}{21},\ x_2 = \frac{16}{21}$. Because this point belongs to S, it is our first point of interest.

Now, we move on to check the boundaries. Let's begin with the section connecting the points (0,0) and (1,1). Here, we have $x_1 = x_2 \in [0, 1]$. We define $g_1(s) = f(s, s) = 7s^2 - 6s$ on $s \in [0, 1]$. Taking the derivative gives

$$\frac{\mathrm{d}g_1}{\mathrm{d}s} = 14s - 6, \tag{7}$$

with a stationary point in $s = \frac{3}{7}$, which is in the interval $[0, 1]$. This gives us our next point of interest: $x_1 = x_2 = \frac{3}{7}$.

Next, let's check the section between (0,1) and (1,1). Here, we define $g_2(s) = f(s, 1) = 4s^2 - s - 2$ on $s \in [0, 1]$. Taking the derivative, we get

$$\frac{\mathrm{d}g_2}{\mathrm{d}s} = 8s - 1, \tag{8}$$

and a stationary point in $s = \frac{1}{8}$, which belongs to the interval. Therefore, our next point of interest is $x_1 = \frac{1}{8},\ x_2 = 1$.

The last part of the boundary is the section between (0,0) and (0,1). We define $g_3(s) = f(0,s) = 4s^2 - 6s$ on $s \in [0,1]$, with the derivative

$$\frac{\mathrm{d}g_3}{\mathrm{d}s} = 8s - 6, \tag{9}$$

and the stationary point $s = 3/4$, which belongs to the interval. The point of interest is $x_1 = 0, x_2 = \frac{3}{4}$.

The next step is to check the values of the found points (and the corners of S). This is done in table 2.

Table 2: The points of interest and their function values.

x_1	x_2	$f(x_1, x_2)$
$\frac{2}{21}$	$\frac{16}{21}$	$-\frac{16}{7}$
$\frac{3}{7}$	$\frac{3}{7}$	$-\frac{9}{7}$
$\frac{1}{8}$	1	$-\frac{33}{16}$
0	$\frac{3}{4}$	$-\frac{9}{4}$
0	0	0
0	1	-2
1	1	1

We see that the minimum function value is $f(x_1^*, x_2^*) = \frac{-16}{7}$ found at $x_1^* = \frac{2}{21}, x_2^* = \frac{16}{21}$.

b) The task is to use the Lagrange multiplier method to find the minimum of the function

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

subject to the constraint

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

We begin by defining

$$L(x_1, x_2, \lambda) = 15 + 2x_1 + 3x_2 + \lambda(x_1^2 + x_1x_2 + x_2^2 - 21). \tag{12}$$

Next, we calculate the gradient and set it to 0

$$\nabla L(x_1, x_2, \lambda) = \begin{bmatrix} 2 + 2\lambda x_1 + \lambda x_2 \\ 3 + \lambda x_1 + 2\lambda x_2 \\ x_1^2 + x_1x_2 + x_2^2 - 21 \end{bmatrix} = 0, \tag{13}$$

resulting in a system of three equations. We multiply the first equation by (-2) and add it to the second to get $x_1 = -\frac{1}{3\lambda}$. Then, we multiply the second equation by (-2) and add it to the first to get $x_2 = -\frac{4}{3\lambda}$. This is then inserted into the third equation to get

$$\frac{1}{9\lambda^2} + \frac{4}{9\lambda^2} + \frac{16}{9\lambda^2} - 21 = 0. \tag{14}$$

This can then be rearranged to $\lambda^2 = \frac{1}{9}$, with the solutions $\lambda = \pm\frac{1}{3}$, which results in either $x_1 = 1, x_2 = 4$ or $x_1 = -1, x_2 = -4$. Calculating the function value for both points shows that the minimum value is $f(x_1^*, x_2^*) = 1$ at $x_1^* = -1, x_2^* = -4$.

Problem 1.3, Basic GA program

a The task is to create a genetic algorithm that finds the minimum of the function

$$\begin{aligned} g\left(x_1, x_2\right) = &\left(1 + \left(x_1 + x_2 + 1\right)^2 \left(19 - 14x_1 + 3x_1^2 - 14x_2 + 6x_1x_2 + 3x_2^2\right)\right) \times \\ &\left(30 + \left(2x_1 - 3x_2\right)^2 \left(18 - 32x_1 + 12x_1^2 + 48x_2 - 36x_1x_2 + 27x_2^2\right)\right). \end{aligned} \tag{15}$$

Running the program with the parameters from 1.3b) and a mutation rate of 0.02 usually yields the minimum g(0, -1) = 3 (with a precision of 2 decimals). Around 1 in 10-20 runs gets stuck in a minimum around $(1.793, 0.195)$, with a function value of approximately 84, but this is easily avoided by running the program multiple times and taking the median as in 1.3b).

b The median fitness values of 100 runs for the given mutation rates are shown in table 3.

Table 3: The obtained median fitness values for the given mutation rates.

Mutation rate	Median fitness value
0.00	0.076
0.02	0.333
0.05	0.332
0.10	0.317

The results show that mutation is absolutely necessary for the GA to work. However, they also show that too high of a mutation rate can lead to the algorithm not converging to the correct value (or to any value at all).

A further analysis of the array with the fitness values found for each of the 100 runs for a mutation rate of 0.1 shows that it almost always gets close to finding the minimum (the array has a mean of 0.2959 and a standard deviation of 0.04), but is quite inaccurate. I am not sure if questions are allowed here, but could a large mutation rate be used for more difficult problems to approximately locate the minimum, followed by a smaller mutation rate in the area to locate it more accurately?

c The task is to make sure that the point (0,-1) found previously is a stationary point of g. To make the calculations easier, we begin by defining

$$g_{11}(x_1, x_2) = (x_1 + x_2 + 1)^2 \tag{16}$$

$$g_{12}(x_1, x_2) = (19 - 14x_1 + 3x_1^2 - 14x_2 + 6x_1x_2 + 3x_2^2) \tag{17}$$

$$g_{21}(x_1, x_2) = (2x_1 - 3x_2)^2 \tag{18}$$

$$g_{22}(x_1, x_2) = (18 - 32x_1 + 12x_1^2 + 48x_2 - 36x_1x_2 + 27x_2^2), \tag{19}$$

which allows us to write (dropping the (x_1, x_2) to save space)

$$g = (1 + g_{11}g_{12})\,(30 + g_{21}g_{22})\,. \tag{20}$$

We then have

$$\nabla g = \begin{bmatrix} (1 + g_{11}g_{12})(\frac{\partial g_{21}}{\partial x_1}g_{22} + g_{21}\frac{\partial g_{22}}{\partial x_1}) + (\frac{\partial g_{11}}{\partial x_1}g_{12} + g_{11}\frac{\partial g_{12}}{\partial x_1})(30 + g_{21}g_{22}) \\ \\ (1 + g_{11}g_{12})(\frac{\partial g_{21}}{\partial x_2}g_{22} + g_{21}\frac{\partial g_{22}}{\partial x_2}) + (\frac{\partial g_{11}}{\partial x_2}g_{12} + g_{11}\frac{\partial g_{12}}{\partial x_2})(30 + g_{21}g_{22}) \end{bmatrix}. \tag{21}$$

Firstly, we can see that the partial derivatives of g_{11} are both $2(x_1 + x_2 + 1)$, which is equal to 0 in (0,-1). Furthermore, $g_{11}(0, -1)$ is also equal to 0 in (0,-1). This allows us to simplify equation 21 to the following

$$\nabla g(0, -1) = \begin{bmatrix} (\frac{\partial g_{21}}{\partial x_1}g_{22} + g_{21}\frac{\partial g_{22}}{\partial x_1}) \\ \\ (\frac{\partial g_{21}}{\partial x_2}g_{22} + g_{21}\frac{\partial g_{22}}{\partial x_2}) \end{bmatrix}. \tag{22}$$

Now, we only need to calculate the values and partial derivatives of g_{21}, g_{22} in (0, -1). The results of these calculations are shown in table 4.

Table 4: The values and partial derivatives of g_{21} and g_{22} in (0, -1).

Function	Value	$\frac{\partial}{\partial x_1}$	$\frac{\partial}{\partial x_2}$
g_{21}	9	12	-18
g_{22}	-3	4	-6

Inserting this into equation 22, we get

$$\nabla g(0, -1) = \begin{bmatrix} (12 \cdot (-3) + 9 \cdot 4) \\ ((-18) \cdot (-3) + 9 \cdot (-6)) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \tag{23}$$

which shows that the point (0,-1) is a stationary point of g. $\square$