FFR105
Stochastic Optimization Algorithms
HP1

Manish Suvarna + smanish

September 22, 2020

Problem 1.1: Penalty Method

1 1

The given function f(x) to be minimized is subjected to an inequality constraint g(x) as shown in Equation 1. Since penalty method converts a constrained optimization problem into one unconstrained optimization problem approach by penalizing the function whenever the predicted minima disobeys the given inequality constraint, which in this case would be to penalize the function whenever $g(x_1, x_2) \geq 0$ or simply $\max(0, g(x_1, x_2))$. Hence in order to convert this constrained problem to unconstrained problem, consider the sum of the function f(x) and a term $p(x; \mu)$ containing penalty factor μ multiplied to the maximum value the of square of non-negative values of constraints g(x). This forms the function $f_p(x; \mu)$ as shown below

$$\begin{aligned} f(x_1, x_2) &= (x_1 - 1)^2 + 2(x_2 - 2)^2 \\ g(x_1, x_2) &= x_1^2 + x_2^2 - 1 \leq 0 \end{aligned} \tag{1}$$

$$f_p(x; \mu) = \begin{cases} (x_1 - 1)^2 + 2(x_2 - 2)^2 + \mu(x_1^2 + x_2^2 - 1)^2 & \text{if } x_1^2 + x_2^2 - 1 \geq 0 \\ (x_1 - 1)^2 + 2(x_2 - 2)^2 & \text{otherwise} \end{cases}$$

2 2

The gradient of the function $f_p(x; \mu)$ will therefore be computed as a vector with partial derivatives of the function $f_p(x; \mu)$ wrt variables x_1 and x_2 as shown in the Equation 2.

$$\nabla(f_p(x; \mu)) = \begin{bmatrix} \frac{\partial f_p}{\partial x_1} & \frac{\partial f_p}{\partial x_2} \end{bmatrix}^T \tag{2}$$

On analytically computing the partial derivatives of the function $f_p(x, \mu)$ as shown in subsection 1 for the two conditions, one when the constraints are violated and the other otherwise, the gradient function was analytically found to be as shown below

$$\nabla(f_p(x;\mu)) = \begin{cases} \begin{bmatrix} 2x_1 + 4_1(x_1^2 + x_2^2 - 1) - 2 \\ 4x_2 + 4_2(x_1^2 + x_2^2 - 1) - 8 \end{bmatrix} & \text{if } x_1^2 + x_2^2 - 1 \geq 0 \\ \begin{bmatrix} 2x_1 - 2 \\ 4x_2 - 8 \end{bmatrix} & \text{otherwise} \end{cases}$$

3 3

The unconstrained minima can be found by equating the gradient of the function $f_p(x;\mu)$ (where the constraints are assumed to be satisfied) equal to 0 as shown in the Equation 3

$$\begin{aligned} 2(x_1 - 1) = 0 &\implies x_1^* = 1 \\ 4(x_2 - 2) = 0 &\implies x_2^* = 2 \end{aligned} \tag{3}$$

Therefore the unconstrained minima of the function where found to be at [1,2].

4 4

The MATLAB program for solving this constrained optimization method for a set of mu values i.e $\mu = [0 \;\; 1 \;\; 10 \;\; 100 \;\; 1000]$, step length $\eta = 0.0001$ and a threshold for the L2 norm of the gradient i.e$\left\|\nabla(f_p(x;\mu))\right\|$ can be found as the file `RunPenaltyMethod.m` along with function files namely `RunGradientDescent.m` and `ComputeGradient.m` respectively.

5 5

On running the MATLAB program `RunPenaltyMethod.m`, the output was found to be as follows

μ	x_1^*	x_2^*
0	1	2
1	0.434	1.21
10	0.331	0.996
100	0.314	0.955
1000	0.312	0.951

It is evident from these results that the stationary points with increasing penalty factor μ does seem to converge. Analysing the output data also shows that at $\mu = 100$ for example, the value of $g(x_1^*, x_2^*) = 0.106$ and as μ increases to 1000, the value $g(x_1^*, x_2^*) = 0.0017$. This proves that with increasing penalty factor the stationary point is getting predicted more closer to the feasible set by making the constraints more closer to getting satisfied.

Problem 1.2: Constrained Optimization

6 a

The function to be minimized is given by the Equation 4 on the closed set S as shown in the Figure 1 respectively. In order to determine the minima, all the possible stationary points that *could* be minima i.e stationary points in the interior of the closed space S, the stationary points on the boundaries ∂S and finally the corner points themselves.

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

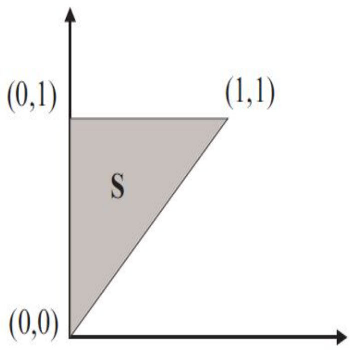

Figure 1: Closed set S of the function f(x_1, x_2)

Stationary points in the interior space S

Stationary points are found where the gradients of the function tend to vanish or become equal to 0. Therefore the stationary points in the interior space S is found by equating the gradient of the function $\nabla(f(x_1, x_2))$ to 0 as shown in Equation 5 which on computing gives the stationary point P1 as (0.0952,0.7619).

$$
\begin{aligned}
\nabla(f(x_1,x_2)) = \begin{bmatrix} \frac{\partial f}{\partial x_1} & \frac{\partial f}{\partial x_2} \end{bmatrix}^T &= [0 \;\; 0]^T \\
\implies \begin{bmatrix} 8x_1 - x_2 \\ -x_1 + 8x_2 - 6 \end{bmatrix} &= \begin{bmatrix} 0 \\ 0 \end{bmatrix} \\
\implies x_1^* = 0.0952 \;\; x_2^* &= 0.7619 \\
\text{Therefore, } P1 &= (0.0952, 0.7619)
\end{aligned}
\tag{5}
$$

Stationary Points on the three boundary lines ∂S

Starting with the line $x_2 = 1$, $0 < x_1 < 1$, the function of $f(x_1, x_2)$ in Equation 4 for this line boils down to the expression given in Equation 6 as function of only x_1 as the variable. On equating the gradinet of this function (wrt x_1) to 0, the stationary points on this line were found to be as P2 = $(\frac{1}{8}, 1)$ as shown in Equation 7

$$
f_{\partial S_1}(x_1, x_2) = 4x_1^2 - x_1 - 2 \tag{6}
$$

$$
\begin{aligned}
\nabla(f_{\partial S_1}(x_1, x_2)) &= 0 \\
\implies 8x_1 - 1 = 0 \implies x_1 &= \frac{1}{8} \\
\text{Therefore, } P2 &= (\frac{1}{8}, 1)
\end{aligned}
\tag{7}
$$

For the line where, $x_1 = 0$ and $0 < x_2 < 1$, the function $f_{\partial S_2}(x_1, x_2)$ is accordingly given by Equation 8. This is now a function of only x_2 as the variable, therefore the gradient of this function wrt x_2 is computed and equated to 0. This gives us the stationary point on the second line as $P3 = (0, \frac{3}{4})$ as shown in Equation 9.

$$
f_{\partial S_2}(x_1, x_2) = 4x_2^2 - 6x_2 \tag{8}
$$

$$
\begin{aligned}
\nabla(f_{\partial S_2}(x_1, x_2)) &= 0 \\
\implies 8x_2 - 6 = 0 \implies x_2 &= \frac{6}{8} \\
\text{Therefore, } P3 &= (0, \frac{3}{4})
\end{aligned}
\tag{9}
$$

Similarly on the line where $x_1 = x_2$, the function in the Equation 4 is evaluated by substituting x_2 as x_1 and the resulting function is given by Equation 10 as a function of only x_1 as the variable. Hence, equating its gradient wrt x_1 to 0 gives the stationary point as $P4 = (\frac{3}{7, \frac{3}{7}})$ as shown in the Equation 11

$$
f_{\partial S_3}(x_1, x_2) = 7x_1^2 - 6x_1 \tag{10}
$$

$$
\begin{aligned}
\nabla(f_{\partial S_3}(x_1, x_2)) &= 0 \\
\implies 14x_1 - 6 = 0 \implies x_1 &= \frac{3}{7} \\
\text{Therefore, } P3 &= (\frac{3}{7}, \frac{3}{7})
\end{aligned} \tag{11}
$$

Stationary points at the corners

We then consider the three corner points i.e $P5 = (0, 0)$, $P6 = (1, 1)$ and $P7 = (0, 1)$ which can also plausibly be the minima as well. Therefore we totally have 7 candidates for the global minimum of the function given in Equation 4. On evaluating this function value on all these 7 determined stationary points as shown in the Table 6, the lowest value of the function $f(x_1, x_2)$ in the Equation 4 was found at P1, i.e. the stationary point found in the interior of the closed space S (0.0952,0.7619). Thus, P1 is the global mimimum of the given function in Equation 4.

x_1	0.0952	0.125	0	0.429	0	1	0
x_2	0.7619	1	0.75	0.429	0	1	1
$f(x_1, x_2)$	-2.2857	-2.0625	-2.25	-1.285	0	1	-2

1.2 b : Lagrange Multiplier method

The function $f(x_1, x_2)$ to be minimized is given by the Equation 12 which is subjected to an equality constraint $h(x_1, x_2)$ as shown in Equation 12.Since the Lagrange multiplier method thrives on the idea that at the optima of the function $f(x_1, x_2)$, the gradient of the function and the constraint must be parallel, the Lagrange function $L(x_1, x_2, \lambda)$ can then be formulated as shown in 13.

$$
\begin{aligned}
f(x_1, x_2) &= 15 + 2x_1 + 3x_2 \\
h(x_1, x_2) &= x_1^2 + x_1x_2 + x_2^2 - 21 = 0
\end{aligned} \tag{12}
$$

$$
\begin{aligned}
L(x_1, x_2, \lambda) &= f(x_1, x_2) + \lambda h(x_1, x_2) \\
\implies L(x_1, x_2, \lambda) &= 15 + 2x_1 + 3x_2 + \lambda(x_1^2 + x_1x_2 + x_2^2 - 21)
\end{aligned} \tag{13}
$$

On equating the gradient of L to 0 we obtain, three equations corresponding to $\frac{\partial L}{\partial x_1}$, $\frac{\partial L}{\partial x_2}$ and $\frac{\partial L}{\partial \lambda}$ which is basically the constraint function. These are as shown in Equations 14, 15, 16.

$$
\frac{\partial L}{\partial x_1} = 2 + \lambda(2x_1 + x_2) = 0 \tag{14}
$$

$$
\frac{\partial L}{\partial x_2} = 3 + \lambda(x_1 + 2x_2) = 0 \tag{15}
$$

$$\frac{\partial L}{\partial \lambda} = x_1^2 + x_1 x_2 + x_2^2 - 21 = 0 \tag{16}$$

On proceeding further to find the stationary points of L by solving the Equations 14, 15, 16 as shown in Equations 17

$$\begin{aligned}
\lambda &= \frac{-2}{2x_1 + x_2} && \text{From Equation 14} \\
3 - \frac{2(x_1 + 2x_2)}{2x_1 + x_2} &= 0 && \text{From Equation 15} \\
\implies 4x_1 - x_2 &= 0 \\
\implies x_2 &= 4x_1 \\
x_1^2 + 4x_1^2 + 16x_1^2 - 21 &= 0 && \text{On substituting } x_2 = 4x_1 \text{ in the Equation 16} \\
21x_1^2 - 21 &= 0 \\
\implies x_1 &= \pm 1 \\
\implies x_2 = 4x_1 &= \pm 4
\end{aligned} \tag{17}$$

Therefore the Lagrange Multiplier method gives 4 stationary points for which the value of the function are tabulated below

x_1^*	x_2^*	$f(x_1^*, x_2^*)$
1	4	29
-1	4	25
1	-4	5
-1	-4	1

Therefore the global minimum of the function in Equation 12 is found to be at the point (-1,-4) since of all the 4 points determined, point (-1,-4) is where the function $f(x_1, x_2)$ takes the least value.

1.3 Basic GA program

a

The assosciated Matlab files for this program can be found in the folder named Problem 1.3. The main file that needs to be run to get the output is named as `FunctionOptimization.m`.

b

The given function $g(x_1, x_2)$ which needs to be minimized is as shown in the Equation 18.

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

The given range of search is [-10 10].The parameters set to run the GA had the chromosome length $m = 50$, population size $N = 100$, cross over probability 0.8, tournament selection parameter of 0.75. This GA was made to run for 100 generations for each mutation rate in the set [0 0.02 0.05] and the results obtained are as shown in the Table below.

Mutation Rate	Fitness Median	x_1^*	x_2^*
0	0.0175	-0.777	-0.222
0.02	0.3333	0	-0.999
0.05	0.265	0	-1

Its clearly evident that the maximum fitness is being achieved at the mutation rate of 0.02 at the point $x_1 = 0$ and $x_2 \approx -1$. Also, the fitness values seem to be worse if there is no mutation applied which can be verified by looking at the data corresponding to the case of 0 mutation rate.

c

The maximum fitness median was found with the variables being $x_1 = 0$ and $x_2 \approx -1$. In order to verify this analytically, we will have to compute the gradient of the function represented in the Equation 18 and then substitue the variables x_1 ad x_2 as 0 and -1. If this makes the gradient of the function $g(x_1, x_2)$ as 0 then we have found the stationary point i.e (0,-1).

Since this function seems to be relatively big to present all the terms during calculations in this report, we will proceed with this by first breaking down the function $g(x_1, x_2)$ as a product of two terms U and V respectively and then apply the gradient to them as shown in the Equation 19. This is done by deploying the chain rule as shown below. Again for simplicity, the longer terms in U and V in Equation 19 can be labelled as U_i and V_i respectively, whose corresponding values at the point(0,-1) is found to be

36 and -3 respectively.

$$\begin{aligned}
U &= (1 + (x_1 + x_2 + 1)^2(19 - 14x_1 + 3x_1^2 - 14x_2 + 6x_1x_2 + 3x_2^2)) \\
V &= (30 + (2x_1 - 3x_2)^2(18 - 32x_1 + 12x_1^2 + 48x_2 - 36x_1x_2 + 27x_2^2)) \\
U_i &= (19 - 14x_1 + 3x_1^2 - 14x_2 + 6x_1x_2 + 3x_2^2) \\
V_i &= (18 - 32x_1 + 12x_1^2 + 48x_2 - 36x_1x_2 + 27x_2^2) \\
\implies \frac{\partial g}{\partial x} &= \begin{bmatrix} \frac{\partial(UV)}{\partial x_1} \\ \frac{\partial(UV)}{\partial x_2} \end{bmatrix} = \begin{bmatrix} U\frac{\partial V}{\partial x_1} + V\frac{\partial U}{\partial x_1} \\ U\frac{\partial V}{\partial x_2} + V\frac{\partial U}{\partial x_2} \end{bmatrix}
\end{aligned} \tag{19}$$

Onwards from here, the partial derivatives can now be computed along with the substitution of the variable values $x_1 = 0$ and $x_2 = -1$ as follows

$$\begin{aligned}
\frac{\partial V}{\partial x_1} &= (2x_1 - 3x_2)^2(-32 + 24x_1 - 36x_2) + V_i(4(2x_1 - 3x_2)) \\
&= (9)(4) + (-3)(4 \cdot 3) = 36 - 36 = 0 \\
\frac{\partial U}{\partial x_1} &= (x_1 + x_2 + 1)^2(-14 + 6x_1 + 6x_2) + U_i \cdot 2(x_1 + x_2 + 1) \\
&= (0)(-20) + 36(0) = 0 \\
\frac{\partial V}{\partial x_2} &= 2(2x_1 - 3x_2)(-3)V_i + (2x_1 - 3x_2)^2(24x_1 + 48 + 54x_2) \\
&= 2(3)(-3)(-3) + 9(48 - 54) = 54 - 54 = 0 \\
\frac{\partial U}{\partial x_2} &= (x_1 + x_2 + 1)^2(-14 + 6x_2) + U_i \cdot 2(x_1 + x_2 + 1) \\
&= (0)(-14 - 6) + 36 \cdot 2(0) = 0
\end{aligned} \tag{20}$$

From the Equations in 20 it is clear that all the derivatives appearing in the gradient of the function $g(x_1, x_2)$ are 0. On substituting this in the equation of $\frac{\partial g}{\partial x}$, the values get nullified to 0 as shown in the Equation 21. This proves that the point (0,-1) is actually a stationary point and the reason being one of the stationary points' properties which is the gradient vanishing at the stationary point.

$$\frac{\partial g}{\partial x} = \begin{bmatrix} \frac{\partial(UV)}{\partial x_1} \\ \frac{\partial(UV)}{\partial x_2} \end{bmatrix} = \begin{bmatrix} U \cdot (0) + V \cdot (0) \\ U \cdot (0) + V \cdot (0) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{21}$$