Colton Cunov
Stochastic Optimization Algorithms
19940810-T732

Home problems 1

Problem 1.1, 3p, Penalty method

In this problem, we shall use the penalty method (see pp. 30-33 in the course book) to find the minimum of the function

$$f(x_1, x_2) = (x_1 - 1)^2 + 2(x_2 - 2)^2$$

subject to the constraint

$$g(x_1, x_2) = \; x_1^2 + x_2^2 - 1 \leq 0$$

1. Define (and specify clearly, in your report, as a function of x_1, x_2 and μ) the function $f_p(\boldsymbol{x}; \mu)$, consisting of the sum of $f(x_1, x_2)$ and the penalty term

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

2. Next, compute (analytically) the gradient $\nabla f_p(\boldsymbol{x}; \mu)$ and include it in your report. Make sure to include both he case where the constraints are fulfilled and the case where they are not.

$$\nabla f_p(\boldsymbol{x}; \mu) = \begin{cases} \begin{bmatrix} 2(x_1 - 1) + 4\mu({x_1}^3 + x_1{x_2}^2 - x_1) \\ 4(x_2 - 2) + 4\mu({x_2}^3 + {x_1}^2 x_2 - x_2) \end{bmatrix}, & {x_1}^2 + {x_2}^2 \geq 1 \\ \begin{bmatrix} 2(x_1 - 1) \\ 4(x_2 - 2) \end{bmatrix}, & otherwise \end{cases}$$

3. Find and report the unconstrained minimum (i.e. for $\mu = 0$) of the function. This point will be used as the starting point for gradient descent.

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

$$\boldsymbol{x_0} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

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.

See folder **1.1**.

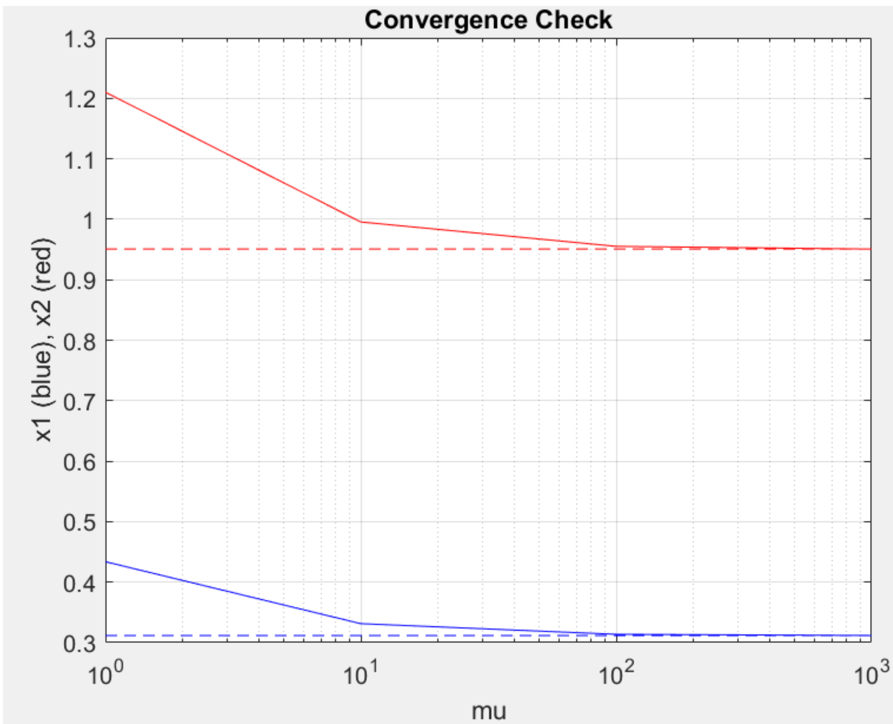
5. Run the program for a suitable sequence of μ values (which you may hard-code in **`RunPenaltyMethod.m`**). Select a suitable (small) value for the step length η and specify it clearly, along with the sequence of μ values, in your report.

$$\mu = [1; 10; 100; 1000] \qquad \eta = 0.0001 \qquad \mathrm{T} = 10^{-6} \qquad \boldsymbol{x}_0 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

Results of run with parameters set as above, rounded to 3 digits after the decimal:

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

To run the program with parameters as listed above, simply execute **`RunPenaltyMethod`** in the command window. Below is a plot with μ on the x-axis (scaled logarithmically), and x_1 and x_2 on the y-axis with their respective values for $\mu = 1000$ as a dashed line.



Colton Cunov
Stochastic Optimization Algorithms
19940810-T732

Problem 1.2, 3p, Constrained optimization

a) Use the analytical method described on pp. 29-30 in the course book to determine the global minimum $(x_1^*, x_2^*)^T$ (as well as the corresponding function value) of the function

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

on the (closed) set S, as shown in the figure. The corners of the triangle are located at $(0,0), (0,1)$ and $(1,1)$.

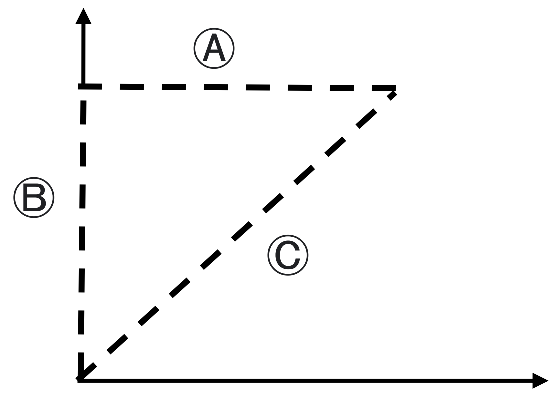
Stationary points:

$$\frac{\partial f}{\partial x_1} = 8x_1 - x_2 = 0 \;\; \Rightarrow \;\; 8x_1 = x_2$$

$$\frac{\partial f}{\partial x_2} = 8x_2 - x_1 - 6 = 0 \;\; \Rightarrow \;\; 8(8x_1) - x_1 - 6 = 63x_1 - 6 = 0$$

$$\boxed{x_1 = \frac{2}{21} \qquad x_2 = \frac{16}{21} \qquad f\left(\frac{2}{21}, \frac{16}{21}\right) = -\frac{16}{7}}$$

Boundaries:



Ⓐ $x_2 = 1\,, 0 < x_1 < 1$

$$f(x_1, 1) = 4{x_1}^2 - x_1 - 2$$

$$\frac{df(x_1,1)}{dx_1} = 8x_1 - 1 = 0$$

$$\boxed{x_1 = \frac{1}{8} \qquad x_2 = 1 \qquad f\left(\frac{1}{8}, 1\right) = \frac{-33}{16}}$$

Ⓑ $x_1 = 0\,, 0 < x_2 < 1$

$$f(0, x_2) = 4{x_2}^2 - 6x_2$$

$$\frac{df(0,x_2)}{dx_2} = 8x_2 - 6 = 0$$

$$\boxed{x_1 = 0 \qquad x_2 = \frac{3}{4} \qquad f\left(0, \frac{3}{4}\right) = -\frac{9}{4}}$$

Ⓒ $x_1 = x_2 \equiv x$

$$f(x) = 7x^2 - 6x$$

$$\frac{df(x)}{dx} = 14x - 6 = 0$$

$$\boxed{x_1 = \frac{3}{4} \qquad x_2 = \frac{3}{4} \qquad f\left(\frac{3}{4}, \frac{3}{4}\right) = -\frac{9}{7}}$$

Colton Cunov
Stochastic Optimization Algorithms
19940810-T732

Corners:

$\boxed{f(0,1) = -2}$

$\boxed{f(0,0) = 0}$

$\boxed{f(1,1) = 1}$

$\therefore$ the global minimum $(x_1^*, x_2^*)^T$ is: $\boxed{x_1^* = \frac{2}{21} \quad x_2^* = \frac{16}{21} \quad f\left(\frac{2}{21}, \frac{16}{21}\right) = -\frac{16}{7}}$

b) Use the Lagrange multiplier method described on pp. 25-28 in the course book to determine the minimum $(x_1^*, x_2^*)^T$ (as well as the corresponding function value) of the function $f(x_1, x_2) = 15 + 2x_1 + 3x_2$ subject to the constraint $h(x_1, x_2) = x_1^2 + x_1x_2 + x_2^2 - 21 = 0$.

$$\nabla f(x_1, x_2) = \begin{bmatrix} 2 \\ 3 \end{bmatrix} \qquad \nabla h(x_1, x_2) = \begin{bmatrix} 2x_1 + x_2 \\ 2x_2 + x_1 \end{bmatrix}$$

$$L(x_1, x_2, \lambda) = f(x_1, x_2) + \lambda h(x_1, x_2) \qquad \nabla f(x_1, x_2) = -\lambda \nabla h(x_1, x_2)$$

$$\begin{bmatrix} 2 \\ 3 \end{bmatrix} = -\lambda \begin{bmatrix} 2x_1 + x_2 \\ 2x_2 + x_1 \end{bmatrix}$$

$$Ⓐ \quad \frac{\partial L}{\partial x_1} = \frac{\partial f}{\partial x_1} + \lambda \frac{\partial h}{\partial x_1} = 2 + \lambda(2x_1 + x_2) = 0$$

$$Ⓑ \quad \frac{\partial L}{\partial x_2} = \frac{\partial f}{\partial x_2} + \lambda \frac{\partial h}{\partial x_2} = 3 + \lambda(2x_2 + x_1) = 0$$

$$Ⓒ \quad \frac{\partial L}{\partial \lambda} = h(x_1, x_2) = x_1^2 + x_1x_2 + x_2^2 - 21 = 0$$

Ⓐ $- 2$Ⓑ $= [2 + \lambda(2x_1 + x_2)] - 2[3 + \lambda(2x_2 + x_1)] = 0 \quad \Rightarrow \quad x_2 = -\frac{4}{3\lambda}$

Substituting $x_2 = -\frac{4}{3\lambda}$ into Ⓐ: $\quad 2 + 2\lambda x_1 + \lambda\frac{-4}{3\lambda} = 0 \quad \Rightarrow \quad x_1 = -\frac{1}{3\lambda}$

Substituting $x_1 = -\frac{1}{3\lambda}$, $x_2 = -\frac{4}{3\lambda}$ into Ⓒ:

$$\left(\frac{-1}{3\lambda}\right)^2 + \left(\frac{-1}{3\lambda}\right)\left(\frac{-4}{3\lambda}\right) + \left(\frac{-4}{3\lambda}\right)^2 - 21 = 0 \quad \Rightarrow \quad \boxed{\lambda = \pm\frac{1}{3} \quad x_1 = \pm 1 \quad x_2 = \pm 4}$$

Out of the four possibilities, $f(x_1, x_2)$ is minimized at $(x_1^*, x_2^*) = (-1, -4)$

Verify constraint: $h(-1, -4) = (-1)^2 + (-1)(-4) + (-1)(-4)^2 = 0 \qquad \boxed{f(-1, -4) = 1}$

Colton Cunov
Stochastic Optimization Algorithms
19940810-T732

Problem 1.3, 4p, Basic GA program

a) To execute the program with these parameters:

Number of variables (e.g. x_1, x_2)	2
Population size	200
Number of genes	50
Mutation probability	.05
Crossover probability	0.8
Tournament selection parameter	0.75
Tournament size	2
Variable range	10
Number of generations	200
Number of copies (# times to replicate fittest individual)	1

simply execute **`FunctionOptimization`** in the command window. This will output a statement like:

```
Global minimum found at:
     0
    -1

Global minimum:
     3
```

b) A table showing the median fitness value for each mutation probability (and parameters set per instructions) is shown below, with median fitness values rounded to 4 digits after the decimal point.

Mutation probability	Median fitness value
0	0.0329
0.02	0.3333
0.05	0.3333
0.1	0.3316

Some mutation clearly helps in our case indicating there is rarely, if ever, a linear combination of $(x_1^*, x_2^*)^T$ in our starting population. It appears the median fitness begins to fall with too much mutation. Running with a higher set of mutation probabilities yields:

Mutation probability	Median fitness value
0.15	0. 3136
0.20	0.2821
0.25	0.2309
0.30	0.1947

Here we conclude that the mutation probability should be high enough to expose the algorithm to a large set of genotypes, but not so high that it negates selection.

Colton Cunov
Stochastic Optimization Algorithms
19940810-T732

c) Prove *analytically* (i.e. by manual calculations, without the help of a computer!) that the point $(x_1^*, x_2^*)^T$ you found in part b) actually is a stationary point of function g. (You do not need to prove that it is a minimum). Make sure to include the relevant intermediate steps in your report, so that the calculation can be followed from beginning to end.

$$g(x_1, x_2) = \left(1 + \overbrace{(x_1 + x_2 + 1)^2}^{A}\overbrace{(19 - 14x_1 + 3{x_1}^2 - 14x_2 + 6x_1x_2 + 3{x_2}^2)}^{B}\right) *$$
$$\left(30 + \underbrace{(2x_1 - 3x_2)^2}_{C}\underbrace{(18 - 32x_1 + 12{x_1}^2 + 48x_2 - 36x_1x_2 + 27{x_2}^2)}_{D}\right)$$

$$= (1 + A * B)(30 + C * D) = u * v$$

Using the product rule, $\nabla g(x_1, x_2)^T = \begin{bmatrix} \frac{\partial g}{\partial x_1} \\ \frac{\partial g}{\partial x_2} \end{bmatrix} = \begin{bmatrix} \frac{\partial}{\partial x_1}(uv) \\ \frac{\partial}{\partial x_2}(uv) \end{bmatrix} = \begin{bmatrix} u\frac{\partial v}{\partial x_1} + v\frac{\partial u}{\partial x_1} \\ u\frac{\partial v}{\partial x_2} + v\frac{\partial u}{\partial x_2} \end{bmatrix}$

$$u = 1 + A * B$$
$$= 3{x_1}^4 + 12{x_1}^3x_2 - 8{x_1}^3 + 18{x_1}^2{x_2}^2 - 24{x_1}^2x_2 - 6{x_1}^2 + 12x_1{x_2}^3$$
$$- 24x_1{x_2}^2 - 12x_1x_2 + 24x_1 + 3{x_2}^4 - 8{x_2}^3 - 6{x_2}^2 + 24x_2 + 20$$

$$\frac{\partial u}{\partial x_1} = 12({x_1}^3 + {x_1}^2(3x_2 - 2) + x_1(3{x_2}^2 - 4x_2 - 1) + {x_2}^3 - 2{x_2}^2 - x_2 + 2)$$

$$\frac{\partial u}{\partial x_2} = 12({x_1}^3 + {x_1}^2(3x_2 - 2) + x_1(3{x_2}^2 - 4x_2 - 1) + {x_2}^3 - 2{x_2}^2 - x_2 + 2)$$

Both of the above partial derivatives, $\frac{\partial u}{\partial x_1}$ and $\frac{\partial u}{\partial x_2}$, when evaluated at $(x_1^*, x_2^*)^T = (0, -1)^T$, equal 0.

$$\begin{aligned} \mathrm{v} &= 30 + \text{Ⓒ} * \text{Ⓓ} \\ &= 48{x_1}^4 - 288{x_1}^3x_2 - 128{x_1}^3 + 648{x_1}^2{x_2}^2 + 576{x_1}^2x_2 + 72{x_1}^2 - 648x_1{x_2}^3 \\ &\quad - 864x_1{x_2}^2 - 216x_1x_2 + 243{x_2}^4 + 432{x_2}^3 + 162{x_2}^2 + 30 \end{aligned}$$

$$\frac{\partial v}{\partial x_1} = 24(8{x_1}^3 - 4{x_1}^2(9x_2 + 4) + 6x_1(9{x_2}^2 + 8x_2 + 1) - 9x_2(3{x_2}^2 + 4x_2 + 1))$$

$$\frac{\partial v}{\partial x_2} = -36(8{x_1}^3 - 4{x_1}^2(9x_2 + 4) + 6x_1(9{x_2}^2 + 8x_2 + 1) - 9x_2(3{x_2}^2 + 4x_2 + 1))$$

Both of the above partial derivatives, $\frac{\partial v}{\partial x_1}$ and $\frac{\partial v}{\partial x_2}$, when evaluated at $(x_1^*, x_2^*)^T = (0, -1)^T$, equal 0.

$$\nabla g(x_1^*, x_2^*)^T = \begin{bmatrix} u\frac{\partial v}{\partial x_1}({x_1}^*, {x_2}^*) + v\frac{\partial u}{\partial x_1}({x_1}^*, {x_2}^*) \\ u\frac{\partial v}{\partial x_2}({x_1}^*, {x_2}^*) + v\frac{\partial u}{\partial x_2}({x_1}^*, {x_2}^*) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$\therefore\ (x_1^*, x_2^*)^T = (0, -1)^T$ is a stationary point.