Techniques in Operations Research

Assignment 2 – 2018

Christian Drozdowicz, 832391

Due: 10am April 18, 2018

Question 1: Let $\mathbf{f(x)} = \frac{\mathbf{x_1^4}}{\mathbf{4}} + \frac{\mathbf{x_2^2}}{\mathbf{2}} - \mathbf{x_1 x_2} + \mathbf{x_1} - \mathbf{x_2}$

Part A)

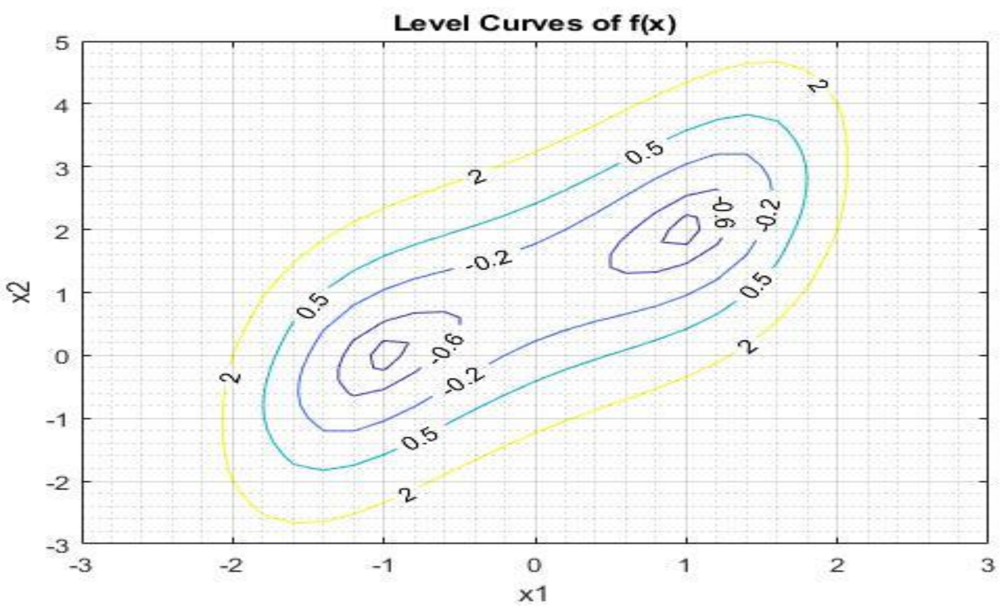
To locate the minimizers of our $f(x)$ we must first plot the level curves of this function in MATLAB,

MATLAB CODE:

```
x1 = -3:0.2:3;
x2 = -3:0.2:5;
[X1,X2] = meshgrid(x1,x2);
Z = ((X1.^4)./4)+((X2.^2)./2)-X1.*X2+X1-X2;

[c,h]=contour(X1,X2,Z,[-0.72 -0.6 -0.2 0.5 2])
clabel(c,h)
grid on
grid minor
```

PLOT OF LEVEL CURVES



As we can see by inspection of this plot, there seems to occur two minimizers which appear at the points $(x_1, x_2) = (-1,0)$ and $(x_1, x_2) = (1,2)$

Part B)

Now we shall go through the Newton Method for two variables to minimize $f(x)$ starting at the initial point $x^0 = (-1,1)$.

Step 1:

$$Select\ x^0 \in \mathbb{R}^n \ \Rightarrow x^0 = (-1,1)$$

$$Set\ k = 0$$

Step 2: $k = 0$

$$\nabla \mathrm{f(x)} = \begin{bmatrix} \mathrm{x}_1^3 - \mathrm{x}_2 + 1 \\ -\mathrm{x}_1 + \mathrm{x}_2 - 1 \end{bmatrix}$$

$$\nabla \mathrm{f(x^0)} = \begin{bmatrix} (-1)^3 - (1) + 1 \\ -(-1) + 1 - 1 \end{bmatrix} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

$$\nabla^2 \mathrm{f(x)} = \begin{bmatrix} 3\mathrm{x}_1^2 & -1 \\ -1 & 1 \end{bmatrix}$$

$$\nabla^2 \mathrm{f(x^0)} = \begin{bmatrix} 3 & -1 \\ -1 & 1 \end{bmatrix}$$

Now we must find the eigenvalues using $\det(A - \lambda I) = 0$:

$$\det\left(\begin{bmatrix} 3 & -1 \\ -1 & 1 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix}\right) = 0$$

$$\det\left(\begin{bmatrix} 3 - \lambda & -1 \\ -1 & 1 - \lambda \end{bmatrix}\right) = 0$$

We know if $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ then $\det(A) = ad - bc$, thus:

$$(3 - \lambda)(1 - \lambda) - 1 = 0$$

$$3 - 4\lambda + \lambda^2 - 1 = 0$$

$$\lambda^2 - 4\lambda + 2 = 0$$

We know the quadratic formula of $ax^2 + bx + c = 0$ is $x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$, thus:

$$\lambda = \frac{4 \pm \sqrt{16 - 8}}{2} = \frac{4 \pm \sqrt{8}}{2} = \frac{4 \pm 2\sqrt{2}}{2} = 2 \pm \sqrt{2}$$

Hence the eigenvalues are $\lambda_1 = 2 + \sqrt{2}$ and $\lambda_2 = 2 - \sqrt{2}$. Since both eigenvalues are strictly positive, the matrix $\nabla^2 f(x^0) > 0$, that is, the matrix is positive definite. So, in this iteration, the step direction is given by the Newton direction:

$$\mathrm{d^k} = -[\nabla^2 \mathrm{f}(\mathrm{x^k})]^{-1} \nabla \mathrm{f}(\mathrm{x^k})$$

$$\Rightarrow \mathrm{d}^0 = -[\nabla^2 \mathrm{f}(\mathrm{x}^0)]^{-1}\nabla \mathrm{f}(\mathrm{x}^0)$$

Now we know the inverse of matrix $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad-bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$, thus:

$$[\nabla^2 \mathrm{f}(\mathrm{x}^0)]^{-1} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{3}{2} \end{bmatrix}$$

$$\Longrightarrow \mathrm{d}^0 = -\begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{3}{2} \end{bmatrix}\begin{bmatrix} -1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ -1 \end{bmatrix}$$

Now we calculate $x^1 = x^0 + t \times d^0 = \begin{bmatrix} -1 \\ 1 \end{bmatrix} + t \times \begin{bmatrix} 0 \\ -1 \end{bmatrix} = \begin{bmatrix} -1 \\ t-1 \end{bmatrix}$

Step 3: To find t:

$$f(x^1) = \frac{(-1)^4}{4} + \frac{(t+1)^2}{2} - (-1)(\mathrm{t}+1) + (-1) - (\mathrm{t}+1)$$

$$f(x^1) = \frac{t^2}{2} - t - \frac{1}{4}$$

$$\nabla \mathrm{f}(\mathrm{x}^1) = t - 1$$

Now set $\nabla \mathrm{f}(\mathrm{x}^1) = 0$, solve for t:

$$t - 1 = 0 \Rightarrow t = 1$$

Step 4:

Set $k = k + 1 = 0 + 1 = 1$

Set $x^1 = (-1,0)$

Return to Step 2

Step 2: $k = 1$

Check $\left|\left|\nabla \mathrm{f}(\mathrm{x}^1)\right|\right| < \epsilon$

$$\nabla \mathrm{f}(\mathrm{x}^1) = \begin{cases} (-1)^3 - 0 + 1 = 0 \\ -(-1) + 0 - 1 = 0 \end{cases}$$

Therefore $\left|\left|\nabla \mathrm{f}(\mathrm{x}^1)\right|\right| = 0$ so **STOP**

We have found the magnitude of the gradient function of $x^1$is equal to zero so we have found a minimum, this is due to zero being less than any tolerance and thus we stop the algorithm.

From **Part A**, we know the point (-1,0) is a local minima for this function so we can be confident we have found an optimal point.

Part C)

We first notice that $\nabla^2 \mathrm{f}(\mathrm{x}) = \begin{bmatrix} 3\mathrm{x}_1^2 & -1 \\ -1 & 1 \end{bmatrix}$ so we realise the error provided in the code is located in the HESSF function:

Original:

```
function hess = hessf(x)

hess(1,1) = 3*x(1);
hess(1,2) = -1;
hess(2,1) =  -1;
hess(2,2) =  1;
```

Fixed:

```
function hess = hessf(x)

hess(1,1) = 3*(x(1)).^2;
hess(1,2) = -1;
hess(2,1) =  -1;
hess(2,2) =  1;
```

And now with the fix in place we can accurately answer question 1c.

In order to condense our findings, we shall use the provided program `sxript.m` to call the functions `steepestDescentMethod.m`, `NewtonMethod.m`, `BFGS.m`, together with our function: `f.m`, with its gradient and hessian matrix: `GRADF.m`, `HESSF.m`

And thus running the program `sxript.m`:

↓

```
>> sxript

Initial condition:

x0 =

    -1     1

   -1.0000    0.0001   -0.7500   17.0000

   -1.0000   -0.0000   -0.7500    1.0000

   -1.0000    0.0000   -0.7500    2.0000


Initial condition:

x0 =

    1.5000    1.0000

    1.0000    2.0000   -0.7500    4.0000

    1.0000    2.0000   -0.7500    3.0000

    1.0000    2.0000   -0.7500    5.0000
```

As we can see, if algorithms start at either (-1,1) or (1.5,1) for the same initial position they will converge to the same minimiser (-1,0) and (1,2) respectively within the given tolerances, although they each take a different number of iterations to get there.

However, the different starting positions do not always converge to the same minimiser point as we can see, starting at (-1,1) converges to the point (-1,0) within the given tolerances and starting at (1.5,1) converges to the point (1,2) within the given tolerances. As we know from part A, these two final points were located to be local minima of the function.

Thus we can see these algorithms will converge to the closest minimum to the initial guess, with each algorithm taking a different amount of iterations to get there.

This is expected as the Steepest Descent Method takes perpendicular directions to the previous direction and uses only first order methods to converge to a point. Newton Method and BFGS converge quicker as they converge quadratically as they are both second order methods.

Since the function we are minimising has more than one local minima, therefore different starting points will lead the algorithms to converge to the closest minimum: (-1,1) is closer to (-1,0) and (1.5,1) is closer to (1,2).

For this particular function, the Newton Method is the most efficient, Steepest Descent is the worst and BFGS lies in-between.

EXTRA: Further Discussion of Question 1)

In **Part A** If we were to go a bit further than inspection of the level curves and find when $\nabla \mathrm{f(x)} = 0$, this will indicate to use the stationary points of the function $f(x)$. Therefore:

$$\nabla \mathrm{f(x)} = \begin{cases} x_1^3 - x_2 + 1 = 0 \\ -x_1 + x_2 - 1 = 0 \end{cases}$$

$$\Rightarrow x^* = (-1,0) \; or \; x^* = (0,1) \; or \; x^* = (1,2)$$

Therefore, it seems the assumptions from the graph in part A is correct to have found the stationary points, but to determine the nature of each point, we must find whether the second-order sufficient condition is satisfied for each point. Therefore:

We will see in part B how to calculate $\nabla^2 \mathrm{f(x)}$ which for our function is

$$\nabla^2 \mathrm{f(x)} = \begin{bmatrix} 3\mathrm{x}_1^2 & -1 \\ -1 & 1 \end{bmatrix}$$

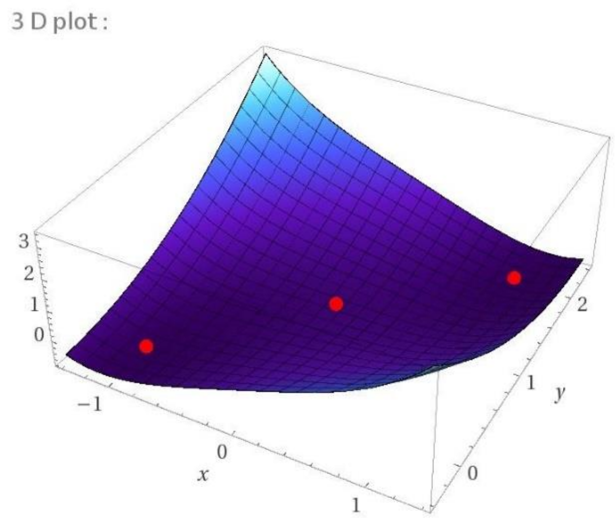
Therefore, it is easy to see that $\nabla^2 \mathrm{f}(-1,0) = \nabla^2 \mathrm{f}(1,2) = \begin{bmatrix} 3 & -1 \\ -1 & 1 \end{bmatrix}$

Now we must find the eigenvalues using $\det(A - \lambda I) = 0$ which can be seen in part B also, the final eigenvalues for these points are $\lambda_1 = 2 + \sqrt{2}$ and $\lambda_2 = 2 - \sqrt{2}$. Since both eigenvalues are strictly positive, the matrix is positive definite and thus both $x^* = (-1,0) \; or \; x^* = (1,2)$ are local minimums, thus confirming our assumption from the plots of the level curves.

However, there is a third point, $x^* = (0,1)$ and its Hessian matrix is $\nabla^2 \mathrm{f}(0,1) = \begin{bmatrix} 0 & -1 \\ -1 & 1 \end{bmatrix}$

The final eigen values for this stationary point is $\lambda_1 = \frac{\sqrt{5}+1}{2}$ and $\lambda_2 = \frac{\sqrt{5}-1}{2}$ thus not all the eigenvalues are strictly positive, the matrix is not positive definite and is thus shown to be a saddle point.

3D PLOT WITH EACH STATIONARY POINT



(where $x = x_1$ and $y = x_2$)

Now in **Part C**, if we were to take random guesses of initial conditions, let's say a point in-between the two local minima and a point very far away from the local minima and we have already noticed that these algorithms will converge to the closest local minima from the starting point then:

Starting at [1,0]

```
>> sxript
Initial condition:
x0 =
    1     0
   0.0000    1.0000   -0.5000    1.0000
   1.0000    2.0000   -0.7500    1.0000
   0.0000    1.0000   -0.5000    1.0000
```

Starting at [-100,100]

```
>> sxript
Initial condition:
x0 =
  -100   100
    1.0000    2.0001   -0.7500   78.0000
   -1.0000   -0.0000   -0.7500    7.0000
    1.0000    2.0000   -0.7500    5.0000
```

We can see that these algorithms do not always converge to the same position, especially if the initial guess is 'bad' (very far away from the local minima), this is due to the way each algorithm converges to find a minimum.

From this we can also see starting at [1,0], Steepest Descent Method as well as BFGS both converged to a saddle point which is not optimal, therefore revealing that in some instances these algorithms can indeed fail to converge to an optimal point.

These are things we must take into consideration when attempting to implement these algorithms to achieve desired results and to have a complete understanding of the limitations of each algorithm.

END OF DISCUSSION

Question 2: min $f(x) = x_1^2 + 3x_2^2 + x_3$

$$\text{s.t } x_1^2 + x_2^2 + x_3^2 - 4 = 0$$

Part A)

First Note: To find all the stationary points and check the constraint qualifications at each of them, we can use the constraint condition to turn the function into a function of just 3 variables:

$$\Rightarrow x_1^2 + x_2^2 + x_3^2 - 4 = 0$$

$$\Rightarrow x_3 = \sqrt{4 - x_1^2 - x_2^2}$$

Substituting this into $f(x)$ and using basic algebra techniques, the gradient $\nabla f(x)$, and setting $\nabla f(x) = 0$ we find:

$$x^* = \left(-\sqrt{\frac{15}{4}}, 0, \frac{1}{2}\right), \left(0, -\sqrt{\frac{143}{36}}, \frac{1}{6}\right), (0,0,-2), (0,0,2), \left(\sqrt{\frac{15}{4}}, 0, \frac{1}{2}\right), \left(0, \sqrt{\frac{143}{36}}, \frac{1}{6}\right)$$

6 stationary points in terms of the three variables x_1, x_2 and x_3.

<u>We now seek a solution by examining the Lagrange function:</u>

Let $g(x) = x_1^2 + x_2^2 + x_3^2 - 4$

So the Lagrangian and gradient:

$$L(x, \eta) = f(x) + \langle \eta, g(x) \rangle$$

$$L(x, \eta) = x_1^2 + 3x_2^2 + x_3 + \eta(x_1^2 + x_2^2 + x_3^2 - 4)$$

$$\nabla_x L(x, \eta) = \begin{bmatrix} 2x_1 + 2\eta x_1 \\ 6x_2 + 2\eta x_2 \\ 1 + 2\eta x_3 \end{bmatrix}$$

We assume the constraint qualifications hold for now and we will check this fact later.

By Theorem 4 on slide 327 of the lecture notes: The Condition on the Lagrangian for minimality: if x is a local optimum, it must be that $g(x) = 0$ and there must exist η such that:

$$\nabla_x L(x, \eta) = \begin{bmatrix} 2x_1 + 2\eta x_1 \\ 6x_2 + 2\eta x_2 \\ 1 + 2\eta x_3 \end{bmatrix} = 0$$

The system of equations we must solve:

1. $2x_1 + 2\eta x_1 = 0$
2. $6x_2 + 2\eta x_2 = 0$
3. $1 + 2\eta x_3 = 0$
4. $x_1^2 + x_2^2 + x_3^2 - 4 = 0$

From 3.$\Rightarrow \eta = \frac{-1}{2x_3}$

Because here we are dividing by x_3, we have assumed that $x_3 \neq 0$, therefore now we have:

5. $2x_1 + 2(\frac{-1}{2x_3})x_1 = 0$
6. $6x_2 + 2(\frac{-1}{2x_3})x_2 = 0$
7. $x_1^2 + x_2^2 + x_3^2 - 4 = 0$

Rearrange 6. for x_3 we find:

$$x_3 = \frac{1}{6}, x_1 = 0, x_2 = \pm\frac{\sqrt{143}}{36}, \eta = -3$$

Rearrange 5. for x_3 we find:

$$x_3 = \frac{1}{2}, x_1 = \pm\frac{\sqrt{15}}{2}, x_2 = 0, \eta = -1$$

Thus, we also find when $x_1 = 0$ and $x_2 = 0$:

$$x_3 = 2, \eta = -\frac{1}{4}$$

OR

$$x_3 = -2, \eta = \frac{1}{4}$$

Therefore, the Lagrange multipliers for:

$x^* = \left(\pm\sqrt{\frac{15}{4}}, 0, \frac{1}{2}\right)$ is $\eta = -1$

$x^* = \left(0, \pm\sqrt{\frac{143}{36}}, \frac{1}{6}\right)$ is $\eta = -3$

$x^* = (0,0,2)$ is $\eta = -\frac{1}{4}$

$x^* = (0,0,-2)$ is $\eta = \frac{1}{4}$

We now check the constrain qualifications at each of these points in order to confirm we could apply Theorem 4. We know that $g(x)$ is not affine so we must look at the constraint gradients at each point:

$$\nabla g(x) = \begin{bmatrix} 2x_1 \\ 2x_2 \\ 2x_3 \end{bmatrix}$$

Now by Gauss-Jordan Elimination to row reduce each vector:

$$\nabla g\left(\begin{bmatrix} \sqrt{\frac{15}{4}} \\ 0 \\ \frac{1}{2} \end{bmatrix}\right) = \begin{bmatrix} \sqrt{15} \\ 0 \\ 1 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

$$\nabla g\left(\begin{bmatrix} -\sqrt{\frac{15}{4}} \\ 0 \\ \frac{1}{2} \end{bmatrix}\right) = \begin{bmatrix} -\sqrt{15} \\ 0 \\ 1 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

$$\nabla g\left(\begin{bmatrix} 0 \\ \sqrt{\frac{143}{36}} \\ \frac{1}{6} \end{bmatrix}\right) = \begin{bmatrix} 0 \\ \frac{\sqrt{143}}{18} \\ \frac{1}{3} \end{bmatrix} \longrightarrow \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

$$\nabla g\left(\begin{bmatrix} 0 \\ -\sqrt{\frac{143}{36}} \\ \frac{1}{6} \end{bmatrix}\right) = \begin{bmatrix} 0 \\ -\frac{\sqrt{143}}{18} \\ \frac{1}{3} \end{bmatrix} \longrightarrow \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

$$\nabla g\left(\begin{bmatrix} 0 \\ 0 \\ 2 \end{bmatrix}\right) = \begin{bmatrix} 0 \\ 0 \\ 4 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

$$\nabla g\left(\begin{bmatrix} 0 \\ 0 \\ -2 \end{bmatrix}\right) = \begin{bmatrix} 0 \\ 0 \\ -4 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

Each of which have full column rank, so the constraint qualifications hold at all stationary points. Thus, by Theorem 4, if this NLP has a locally optimal point, it must be one of these 6 stationary points.

$$x^* = \left(-\sqrt{\frac{15}{4}}, 0, \frac{1}{2}\right), \left(0, -\sqrt{\frac{143}{36}}, \frac{1}{6}\right), (0,0,-2), (0,0,2), \left(\sqrt{\frac{15}{4}}, 0, \frac{1}{2}\right), \left(0, \sqrt{\frac{143}{36}}, \frac{1}{6}\right)$$

(Question 2 Part B omitted from this assignment as per email instructions)