Exercise #3

January 31, 2023

Problem 1.

Assume that $B \in \mathbb{R}^{d\times d}$ is a non-singular matrix (not necessarily orthogonal), $\hat{c} \in \mathbb{R}^d$, and $f : \mathbb{R}^d \mapsto \mathbb{R}$ is a real-valued function. Define the function $g(x) = f(Bx + \hat{c})$.

a) Find expressions for $\nabla g(x)$ and $\nabla^2 g(x)$ in the terms of f, B and $\hat{c}$. (**Hint:** You can use the chain-rules for ∇ and ∇^2.)

b) Let $x \in \mathbb{R}^d$ and denote by x_1 the result of one Newton step starting at x for the minimization of g with the (Armijo) backtracking line search and the parameters $0 < c < 1$ (sufficient decrease parameter), $0 < \rho < 1$ (contraction factor), and $\hat{\alpha} = 1$ (initial step length). Moreover, let $y = Bx + \hat{c}$ and denote by y_1 the result of one Newton step starting at y for the minimization of f with the (Armijo) backtracking line search and the same parameters $0 < c < 1$, $0 < \rho < 1$, and $\hat{\alpha} = 1$. Show that

$$y_1 = Bx_1 + \hat{c}. \tag{1}$$

c) Show that the relation (1) in general does not hold for the gradient descent method unless B is orthogonal matrix.

Solution.

a) For clarity, we set $y(x) = Bx + \hat{c}$ so that $g(x) = f(y(x))$. Now, we have three ways to solve this.
The first one is the easiest way just by memorizing two basic chain-rules for ∇ and ∇^2. For any (smooth) maps $h : \mathbb{R}^d \mapsto \mathbb{R}$ and $H : \mathbb{R}^d \mapsto \mathbb{R}^d$,

$$\nabla(h \circ H)(x) = (J_H)^T \nabla h(H(x)) \text{ and } \nabla^2(h \circ H)(x) = (J_H)^T \nabla^2 h(H(x)) J_H,$$

where J_H is the Jacobian of H and $h \circ H$ represents the composite function. If we put $h = f$, $H = y(x) = Bx + \hat{c}$ and $J_H = B$, we obtain our desired solutions

$$\nabla(g(x)) = \nabla(f(y(x))) = B^T \nabla f(Bx + \hat{c}) \text{ and } \nabla^2(g(x)) = \nabla^2(f(y(x))) = B^T \nabla^2 f(Bx + \hat{c}) B.$$

The second way of computing $\nabla g(x)$ and $\nabla^2 g(x)$ is by doing calculations at index level (manually) which can be messy and exhausting. However, for a better picture of the concept of finding the expression for $\nabla g(x)$ and $\nabla^2 g(x)$, we can at least try to compute $\nabla g(x)$. We have $y : \mathbb{R}^d \mapsto \mathbb{R}^d$ such that $y(x) = Bx + \hat{c}$. If we choose the general form of the matrix B as

$$B = \begin{pmatrix} b_{11} & b_{12} & \dots & b_{1d} \\ b_{21} & b_{22} & \dots & b_{2d} \\ \dots & \dots & \dots & \dots \\ b_{d1} & b_{d2} & \dots & b_{dd} \end{pmatrix},$$

we can write

$$y = \begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_d \end{pmatrix} = \begin{pmatrix} b_{11} & b_{12} & \dots & b_{1d} \\ b_{21} & b_{22} & \dots & b_{2d} \\ \dots & \dots & \dots & \dots \\ b_{d1} & b_{d2} & \dots & b_{dd} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \dots \\ x_d \end{pmatrix} + \begin{pmatrix} c_1 \\ c_2 \\ \dots \\ c_d \end{pmatrix},$$

which gives

$$
\begin{aligned}
y_1 &= b_{11}x_1 + b_{12}x_2 + \ldots + b_{1d}x_d + c_1\\
y_2 &= b_{21}x_1 + b_{22}x_2 + \ldots + b_{2d}x_d + c_2\\
\ldots &\quad \ldots \quad \ldots \quad \ldots \quad \ldots\\
y_d &= b_{d1}x_1 + b_{d2}x_2 + \ldots + b_{dd}x_d + c_d.
\end{aligned} \tag{2}
$$

Now for $x \in \mathbb{R}^d$, we have

$$
\begin{aligned}
\nabla g(x) &= \left(\frac{\partial g}{\partial x_1}, \frac{\partial g}{\partial x_2}, \ldots, \frac{\partial g}{\partial x_d}\right)^T\\
&= \left(\frac{\partial f(y(x))}{\partial x_1}, \frac{\partial f(y(x))}{\partial x_2}, \ldots, \frac{\partial f(y(x))}{\partial x_d}\right)^T\\
&= \left(\frac{\partial f}{\partial y_1}.\frac{\partial y_1}{\partial x_1} + \frac{\partial f}{\partial y_2}.\frac{\partial y_2}{\partial x_1} + \ldots + \frac{\partial f}{\partial y_d}.\frac{\partial y_d}{\partial x_1}, \frac{\partial f}{\partial y_1}.\frac{\partial y_1}{\partial x_2} + \frac{\partial f}{\partial y_2}.\frac{\partial y_2}{\partial x_2} + \ldots + \frac{\partial f}{\partial y_d}.\frac{\partial y_d}{\partial x_2}, \ldots, \frac{\partial f}{\partial y_1}.\frac{\partial y_1}{\partial x_d} + \frac{\partial f}{\partial y_2}.\frac{\partial y_2}{\partial x_d} + \ldots + \frac{\partial f}{\partial y_d}.\frac{\partial y_d}{\partial x_d}\right)^T\\
&\qquad \text{by using (2)}\\
&= \left(\frac{\partial f}{\partial y_1}.b_{11} + \frac{\partial f}{\partial y_2}.b_{21} + \ldots + \frac{\partial f}{\partial y_d}.b_{d1}, \frac{\partial f}{\partial y_1}.b_{12} + \frac{\partial f}{\partial y_2}.b_{22} + \ldots + \frac{\partial f}{\partial y_d}.b_{d2}, \ldots, \frac{\partial f}{\partial y_1}.b_{1d} + \frac{\partial f}{\partial y_2}.b_{2d} + \ldots + \frac{\partial f}{\partial y_d}.b_{dd}\right)^T\\
&= \begin{pmatrix} b_{11} & b_{21} & \ldots & b_{d1}\\ b_{12} & b_{22} & \ldots & b_{d2}\\ \ldots & \ldots & \ldots & \ldots\\ b_{1d} & b_{2d} & \ldots & b_{dd}\end{pmatrix}\begin{pmatrix}\frac{\partial f}{\partial y_1}\\ \frac{\partial f}{\partial y_2}\\ \ldots\\ \ldots\\ \frac{\partial f}{\partial y_d}\end{pmatrix}\\
&= B^T\nabla f(y) \text{ (since } y = Bx + \hat{c})\\
&= B^T\nabla f(Bx + \hat{c}).
\end{aligned}
$$

Similarly, you can try to compute $\nabla^2 g(x)$.
Now, if you want to avoid the above-mentioned long method, you can compute only the possible partial derivatives instead of entire gradient of f, for $i = 1, 2, \ldots, d$ given as

$$\frac{\partial}{\partial x_i}[f(y(x))] = \left\langle \nabla f(y(x)), \frac{\partial}{\partial x_i}y(x)\right\rangle, \text{ where } \frac{\partial}{\partial x_i}y(x) = B_{\cdot,i} \ (i^{\text{th}}\text{column of } B),$$

which implies

$$(\nabla g(x))_i = \frac{\partial g(x)}{\partial x_i} = \frac{\partial}{\partial x_i}[f(y(x))] = \langle \nabla f(y(x)), B_{\cdot,i}\rangle = (B_{\cdot,i})^T\nabla f(y(x)).$$

From this we see that $\nabla g(x) = B^T\nabla f(y(x))$.
Next, we can write

$$(\nabla^2 g(x))_i = \nabla(\nabla g(x))_i = \left\langle \nabla^2 f(y(x))\frac{\partial y(x)}{\partial x_i}, \frac{\partial y(x)}{\partial x_i}\right\rangle + \langle \nabla f(y(x)), 0\rangle = \langle \nabla^2 f(y(x))B_{\cdot,i}, B_{\cdot,i}\rangle = (B_{\cdot,i})^T\nabla^2 f(y(x))B_{\cdot,i}.$$

From this we can see that

$$\nabla^2 g(x) = B^T\nabla^2 f(y(x))B.$$

b) By Newton's method we have $p_0 = -(\nabla^2 g(x_0))^{-1}\nabla g(x_0)$ and so $x_1 = x_0 + \alpha_0 p_0$ where the step length in the first step $\alpha_0 = \hat{\alpha}\rho^{k_0}$ and $k_0 \in \{0, 1, 2, \ldots\}$ is the smallest non-negative integer, at which Armijo condition $g(x_0 + \alpha_0 p_0) \le g(x_0) + c\alpha_0\nabla g(x_0)^T p_0$ is satisfied. Turning to the minimization of f with respect to y, we similarly obtain $\hat{p_0} = -(\nabla^2 f(y_0))^{-1}\nabla f(y_0)$ and $y_1 = y_0 + \hat{\alpha_0}\hat{p_0}$ where the step length in the first step $\hat{\alpha_0} = \hat{\alpha}\rho^{\hat{k_0}}$ and $\hat{k_0}$ is the smallest non-negative integer, at which the Armijo condition $f(y_0 + \hat{\alpha_0}\hat{p_0}) \le f(y_0) + c\hat{\alpha_0}\nabla f(y_0)^T\hat{p_0}$ is satisfied. Our task is to prove that $y_1 = Bx_1 + \hat{c}$ given that $y_0 = Bx_0 + \hat{c}$. we start by showing $Bp_0 = \hat{p}_0$. Indeed

$$\begin{aligned}
Bp_0 &= -B(\nabla^2 g(x_0))^{-1}\nabla g(x_0) \\
&= -B(B^T\nabla^2 f(Bx_0+\hat{c})B)^{-1}B^T\nabla f(Bx_0+\hat{c}) \text{ (by using (a))} \\
&= -BB^{-1}(B^T\nabla^2 f(Bx_0+\hat{c}))^{-1}B^T\nabla f(Bx_0+\hat{c}) \text{ (by using the property of matrix } (AB)^{-1} = B^{-1}A^{-1}) \\
&= -BB^{-1}(\nabla^2 f(Bx_0+\hat{c}))^{-1}(B^T)^{-1}B^T\nabla f(Bx_0+\hat{c}) \text{ (by using the property of matrix } AA^{-1} = A^{-1}A = I) \\
&= -(\nabla^2 f(y_0))^{-1}\nabla f(y_0) \\
&= \hat{p}_0.
\end{aligned}$$

This further implies that for all $\alpha \in \mathbb{R}$

$$\begin{aligned}
g(x_0+\alpha p_0) &= f(y(x_0+\alpha p_0)) \\
&= f(B(x_0+\alpha p_0)+\hat{c}) \\
&= f(Bx_0+\alpha Bp_0+\hat{c}) \\
&= f(Bx_0+\hat{c}+\alpha\hat{p}_0) \\
&= f(y_0+\alpha\hat{p}_0)
\end{aligned}$$

and

$$\begin{aligned}
g(x_0)+c_1\alpha\nabla g(x_0)^T p_0 &= f(y(x_0))+c_1\alpha(B^T\nabla f(Bx_0+\hat{c}))^T p_0 \text{ (by using (a))} \\
&= f(y_0)+c_1\alpha(B^T\nabla f(y(x_0)))^T p_0 \\
&= f(y_0)+c_1\alpha\nabla f(y(x_0))^T Bp_0 \text{ (by using the property of matrix } (AB)^T = B^TA^T \text{ and } (A^T)^T = A) \\
&= f(y_0)+c_1\alpha\nabla f(y(x_0))^T\hat{p}_0.
\end{aligned}$$

Therefore, the Armijo conditions for g and f are

$$g(x_0+\alpha p_0) \le g(x_0)+c_1\alpha\nabla g(x_0)^T p_0,$$

$$f(y_0+\alpha\hat{p}_0) \le f(y_0)+c_1\alpha\nabla f(y_0)^T\hat{p}_0, \text{ respectively.}$$

That means Armijo condition is satisfied in the (g, x)-realm when it is satisfied in the (f, y)-realm, that is $k_0 = \hat{k}_0$ and consequently $\alpha_0 = \hat{\alpha}_0$. Now, we arrive at the desired conclusion, i.e.,

$$\begin{aligned}
y_1 &= y_0+\hat{\alpha}_0\hat{p}_0 \\
&= (Bx_0+\hat{c})+\alpha_0 Bp_0 \\
&= B(x_0+\alpha_0 p_0)+\hat{c} \\
&= Bx_1+\hat{c}.
\end{aligned}$$

c) For the gradient descent method

$$\begin{aligned}
Bp_0 &= -B\nabla g(x_0) \\
&= -B(B^T\nabla f(Bx_0+\hat{c})) \text{ (by using (a))} \\
&= -BB^T\nabla f(y_0).
\end{aligned}$$

Now, if the matrix B is orthogonal then $BB^T = I$ (Identity matrix), which implies that $Bp_0 = \hat{p}_0$. Therefore, if B is not orthogonal, by following the same steps of (b)), we can prove the relation (1).

Problem 2.

Implement both the gradient descent method and Newton's method with backtracking line search. Apply your method to the minimization of *Rosenbrock* function $f(x, y) = 100(y-x^2)^2 + (1-x)^2$.
The Newton direction is not necessarily a descent direction for this function, as f is not convex, and thus it might be

necessary to modify the search directions in the Newton method. Do this by switching to the negative gradient direction, whenever the inequality $-\nabla f(x_k)^T p_k^{\text{Newton}} \leq \epsilon \|\nabla f(x_k)\| \|p_k^{\text{Newton}}\|$ holds (here $\epsilon > 0$ is some fixed small parameter).

Solution.

See possible solutions on the wiki page.

Problem 3.

Consider the function $f : \mathbb{R}^2 \mapsto \mathbb{R}$ (see Exercise 1, Problem 3a and Exercise 2, Problem 3)

$$f(x, y) = \frac{x^2}{2} + x \cos y.$$

We want to apply a line search method for the minimization of f, starting at the point $x_0 = (1, \frac{\pi}{4})$ and with the search direction $p_0 = (-1, 0)$. State the Wolfe conditions for this line search and explain their purpose! If the parameters are chosen as $c_1 = 0.1$ and $c_2 = 0.8$, what is the range of admissible values for the step length α?

Solution.

The Wolfe conditions (weak) are given as

$$f(x_k + \alpha_k p_k) \leq f(x_k) + c_1 \alpha_k \nabla f(x_k)^T p_k \text{ (Armijo condition)},$$

$$\nabla f(x_k + \alpha_k p_k)^T p_k \geq c_2 \nabla f(x_k)^T p_k \text{ (Curvature condition)}.$$

Choosing the step lengths such that the Wolfe conditions are satisfied ensures sufficiently decrease of the objective function from one iteration to the next one, and to ensure convergence of the line search method. For $x_0 = \left(1, \frac{\pi}{4}\right)$ and $p_0 = (-1, 0)$, the Armijo condition $f(x_0 + \alpha p_0) \leq f(x_0) + c_1 \alpha \nabla f(x_0)^T p_0$ gives

$$\frac{(1-\alpha)^2}{2} + \frac{(1-\alpha)}{\sqrt{2}} \leq \frac{1}{2} + \frac{1}{\sqrt{2}} - c_1 \alpha \left(1 + \frac{1}{\sqrt{2}}\right)$$

$$\begin{aligned}
&\Rightarrow \frac{(1-\alpha)(1-\alpha+\sqrt{2})}{2} \leq \frac{1+\sqrt{2}}{2} - c_1 \alpha \frac{(2+\sqrt{2})}{2} \\
&\Rightarrow \alpha(\alpha - 2 - \sqrt{2} + c_1(2+\sqrt{2})) \leq 0 \text{ (since } \alpha > 0) \\
&\Rightarrow \alpha - (2+\sqrt{2}) + c_1(2+\sqrt{2}) \leq 0 \\
&\Rightarrow \alpha \leq (1-c_1)(2+\sqrt{2}).
\end{aligned}$$

Similarly, curvature condition $\nabla f(x_0 + \alpha p_0)^T p_0 \geq c_2 \nabla f(x_0)^T p_0$ gives

$$\left(1 + \frac{1}{\sqrt{2}} - \alpha, \frac{\alpha - 1}{\sqrt{2}}\right) \begin{pmatrix} -1 \\ 0 \end{pmatrix} \geq c_2 \left(1 + \frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right) \begin{pmatrix} -1 \\ 0 \end{pmatrix}$$

$$\Rightarrow -1 - \frac{1}{\sqrt{2}} + \alpha \geq -c_2 \left(1 + \frac{1}{\sqrt{2}}\right)$$

$$\Rightarrow \alpha \geq (1 - c_2)\left(1 + \frac{1}{\sqrt{2}}\right).$$

By putting the values $c_1 = 0.1$ and $c_2 = 0.8$, we obtain the following admissible value range of α

$$0.3414 \leq \alpha \leq 3.0727.$$

Problem 4.

Consider the optimization problem

$$\min_{x\in\mathbb{R}^2} f(x), \tag{3}$$

where the objective function $f\colon \mathbb{R}^2 \mapsto \mathbb{R}$ is given by

$$f(x, y) = 2x^2 - 2xy^2 - 12x + y^4 + 2y^2 + 36.$$

a) Compute all stationary (critical) points of the optimization problem (3) and also find all its local and global minima.

b) Starting at the point $x_0 = (1, 2)$ compute one step of Newton's method with backtracking (Armijo) line search. Start with an initial step length $\alpha_0 = 1$ and use the parameters $c = 1/8$ (sufficient decrease parameter) and $\rho = 1/2$ (contraction factor).

Solution.

a)

$$(\text{Gradient of } f)\nabla f(x, y) = (4x - 2y^2 - 12, -4xy + 4y^3 + 4y)^T, \text{ and}$$

$$(\text{Hessian of } f)\nabla^2 f(x, y) = \begin{pmatrix} 4 & -4y \\ -4y & -4x + 12y^2 + 4 \end{pmatrix}.$$

The first-order necessary condition of optimality gives $\nabla f(x, y) = 0$. Therefore, we have

$$4x - 2y^2 - 12 = 0 \text{ and } -4xy + 4y^3 + 4y = 0.$$

$$-4xy + 4y^3 + 4y = 0 \implies 4y(-x + y^2 + 1) = 0 \implies \text{either } y = 0 \text{ or } -x + y^2 + 1 = 0.$$

$$\text{Now, } y = 0 \text{ and } 4x - 2y^2 - 12 = 0 \implies (x, y) = (3, 0).$$

$$\text{Further } -x + y^2 + 1 = 0 \text{ and } 4x - 2y^2 - 12 = 0 \implies (x, y) = (5, 2) \text{ and } (x, y) = (5, -2).$$

Eventually, we have three stationary points of the optimization problem (3), $(3, 0)$, $(5, 2)$, and $(5, -2)$.
Next, to check whether these are local or global minima, we have to find the Hessian matrix. Since the Hessian matrix at stationary point $(3, 0)$ is

$$\nabla^2 f(3, 0) = \begin{pmatrix} 4 & 0 \\ 0 & -8 \end{pmatrix},$$

which is indefinite because it is neither positive definite nor negative definite (since it has both positive and negative eigen values 4 and -8). Therefore, the stationary point $(3, 0)$ is saddle point, that means, at $(3, 0)$ the optimization problem (3) has neither local minimum nor local maximum.
Further, the Hessian matrix at $(5, 2)$ and $(5, -2)$ are

$$\nabla^2 f(5, 2) = \begin{pmatrix} 4 & -8 \\ -8 & 32 \end{pmatrix} \text{ and } \nabla^2 f(5, -2) = \begin{pmatrix} 4 & 8 \\ 8 & 32 \end{pmatrix}.$$

The above-mentioned Hessian matrices are symmetric and their approximate eigen values are 34.1245 and 1.8755, which are non-zero and positive. Consequently, both Hessian matrices are positive definite. We can also check the positive definiteness by checking the leading principal minors as; since, the diagonal entries and determinant of both Hessian matrices are positive, $\nabla^2 f(5, 2)$ and $\nabla^2 f(5, -2)$ are positive definite. Hence $(5, -2)$ and $(5, 2)$ are strict local minima.

Now, we have to check whether these are global minima. For that, we can check it by checking lower semi-continuity and coercivity of the objective function f. Since, f is the polynomial, it is continuous and then lower semi-continuous function. Next, f can be rewritten as $f(x, y) = (x-6)^2 + (x-y^2)^2 + 2y^2$. Every term on the right hand side of the function f has even power. Therefore, in both cases $x, y \to +\infty$ and $x, y \to -\infty$, $f \to +\infty$. Consequently, f is coercive function. Now, we can say that optimization problem (3) admits at least one global minimizer (by the lecture note 1). Thus, either both stationary points $(5, 2)$ and $(5, -2)$ could be global mimina or only one of these. In order to be sure about this, we find the value of function f at both stationary points, which yields that $f(5, 2) = f(5, -2) = 10$. Therefore, both strict local minima $(5, 2)$ and $(5, -2)$ are global minima.

b) First, we have to find the search direction for the starting point $x_0 = (1, 2)$, which is

$$p_0 = -(\nabla^2 f(x_0))^{-1} \nabla f(x_0). \tag{4}$$

$$\nabla f(x_0) = (-16, 32)^T,\ \nabla^2 f(x_0) = \begin{pmatrix} 4 & -8 \\ -8 & 48 \end{pmatrix} \text{ and } (\nabla^2 f(x_0))^{-1} = \frac{1}{128}\begin{pmatrix} 48 & 8 \\ 8 & 4 \end{pmatrix}.$$

By using above values in (4), we obtain $p_0 = (4, 0)^T$. Let's check whether the Armijo condition is satisfied for the step length $\alpha_0 = 1$. The Armijo condition is

$$f(x_0 + \alpha_0 p_0) \le f(x_0) + c\alpha_0 \nabla f(x_0)^T p_0,$$

which gives

$$10 \le 34.$$

Therefore, the Armijo condition is satisfied at $\alpha_0 = 1$. Consequently, we can choose the initial step length $\alpha_0 = 1$, which gives the one step of Newton's method as

$$x_1 = x_0 + \alpha_0 p_0 = (5, 2).$$

Note: Since $(5, 2)$ is a global minimizer too, the iteration will stop here.

Problem 5.

Consider the optimization problem

$$\min_{x \in \mathbb{R}^2} f(x), \tag{5}$$

where the objective function $f : \mathbb{R}^2 \mapsto \mathbb{R}$ is given by

$$f(x, y) = x^2 - 2x + x^2 y^2 - 2xy.$$

a) Compute all stationary (critical) points of the optimization problem (5) and also find all its local and global minima.

b) Starting at the point $x_0 = (0, 0)$ compute one step of the gradient (steepest) descent method. Ensure that the step length satisfies the Wolfe conditions with $c_1 = \frac{1}{4}$ and $c_2 = \frac{3}{4}$.

c) In part b), can you also use the Newton search direction?

Solution.

a)

$$(\text{Gradient of } f) \nabla f(x, y) = (2x - 2 + 2xy^2 - 2y, 2x^2 y - 2x)^T, \text{ and}$$

$$(\text{Hessian of } f) \nabla^2 f(x, y) = \begin{pmatrix} 2 + 2y^2 & 4xy - 2 \\ 4xy - 2 & 2x^2 \end{pmatrix}.$$

The first-order necessary condition of optimality requires that $\nabla f(x, y) = 0$. Therefore, we have to solve the equations

$$2x - 2 + 2xy^2 - 2y = 0 \text{ and } 2x^2y - 2x = 0.$$

$$2x^2y - 2x = 0 \implies 2x(xy - 1) = 0 \implies \text{either } x = 0 \text{ or } xy = 1.$$

$$\text{Now, } x = 0 \text{ and } 2x - 2 + 2xy^2 - 2y = 0 \implies (x, y) = (0, -1).$$

$$\text{Further } xy = 1 \text{ and } 2x - 2 + 2xy^2 - 2y = 0 \implies (x, y) = (1, 1).$$

Eventually, we have two stationary points of the optimization problem (5), $(0, -1)$ and $(1, 1)$.
Next, to check whether these are local or global minima, we have to find the Hessian matrix. Since the Hessian matrix at the stationary point $(0, -1)$ is

$$\nabla^2 f(0, -1) = \begin{pmatrix} 4 & -2 \\ -2 & 0 \end{pmatrix},$$

which is indefinite because it is neither positive definite nor negative definite (since it has both positive and negative approximate eigen values 4.8284 and -0.8285). Therefore, the stationary point $(0, -1)$ is saddle point, that means, at $(0, -1)$ the optimization problem (5) has neither a local minimum nor a local maximum.
Further, the Hessian matrices at $(1, 1)$ is

$$\nabla^2 f(1, 1) = \begin{pmatrix} 4 & 2 \\ 2 & 2 \end{pmatrix}.$$

This Hessian matrix is symmetric and its approximate eigen values are 5.3452 and 0.6548, which are non-zero and positive. Consequently, the Hessian matrix is positive definite. Hence $(1, 1)$ is a strict local minima.

Now, we have to check that whether $(1, 1)$ is global minimum. Start by checking lower semi-continuity and coercivity of the objective function f. Since, f is polynomial, it is continuous and hence lower semi-continuous. Further, we can write the objective function as $f(x, y) = (x - 1)^2 + (xy - 1)^2 - 2$, and $f(0, y) = 0$ which means f does not tends to $+\infty$ for all x. Therefore, f is not coercive function. Now for the existence of global minimizer of optimization problem (5), the objective function f should be coercive. However, if f is not coercive, we cannot be sure that optimization problem (5) does not attain any global minimizer. It could be and not be too. Let's check by the definition. Since, $-2 = f(1, 1) \leq f(x, y) = (x - 1)^2 + (xy - 1)^2 - 2$, for all $(x, y) \in \mathbb{R}^2$. Therefore, $(1, 1)$ is global minimizer.

b) First, we have to find the search direction for the starting point $x_0 = (0, 0)$, which is

$$p_0 = -\nabla f(x_0). \tag{6}$$

$$\nabla f(x_0) = (-2, 0)^T.$$

By using above values in (6), we obtain $p_0 = (2, 0)^T$. The Wolfe conditions (weak) are given as

$$f(x_0 + \alpha_0 p_0) \leq f(x_0) + c_1 \alpha_0 \nabla f(x_0)^T p_0 \text{ (Armijo condition)}, \tag{7}$$

$$\nabla f(x_0 + \alpha_0 p_0)^T p_0 \geq c_2 \nabla f(x_0)^T p_0 \text{ (Curvature condition)}. \tag{8}$$

Condition (7) with $c_1 = \frac{1}{4}$ gives $\alpha \leq \frac{3}{4}$, and (8) with $c_2 = \frac{3}{4}$ gives $\alpha \geq \frac{1}{8}$. Therefore, we have $\frac{1}{8} \leq \alpha \leq \frac{3}{4}$. Now, we can choose the initial step length $\alpha_0 = \frac{1}{2}$ for which the one step of gradient method is $x_1 = x_0 + \alpha_0 p_0 = (1, 0)$.

c) Let's check the search direction for the starting point $x_0 = (0, 0)$ of Newton's method, which is

$$p_0 = -(\nabla^2 f(x_0))^{-1} \nabla f(x_0). \tag{9}$$

$$\nabla f(x_0) = (-2, 0)^T, \ \nabla^2 f(x_0) = \begin{pmatrix} 2 & -2 \\ -2 & 0 \end{pmatrix} \text{ and } (\nabla^2 f(x_0))^{-1} = -\frac{1}{4} \begin{pmatrix} 0 & 2 \\ 2 & 2 \end{pmatrix}.$$

By using above values in (9), we obtain $p_0 = (0, -1)^T$. Further, $p_0^T \nabla f(x_0) = 0 \nless 0$. Therefore, the search direction $p_0 = (0, -1)^T$ is not descent which follows that we cannot use the Newton's search direction in part (b)).