Student name: Adi Dereviani Prager
Student ID: 305674731

Numerical Optimization with Python (2024B)

Dry HW 02 – Solutions

Solution 1:

a. Given the function $f\colon \mathbb{R}^2 \to \mathbb{R}$ defined by $f(x) = 2x_1^2 + 4x_2^2 + 6x_1x_2$, we are to express this as: $f(x) = \frac{1}{2}x^TQx$ where Q is a symmetric matrix.

First, we equate the given function with the general form of a quadratic function:

$$2x_1^2 + 4x_2^2 + 6x_1x_2 = \mathbf{x}^T \begin{bmatrix} q_{11} & q_{12} \\ q_{21} & q_{22} \end{bmatrix} \mathbf{x}$$

Where $q_{12} = q_{21}$ due to symmetry. Expanding the right-hand side:

$$x_1^2q_{11} + 2x_1x_2q_{12} + x_2^2q_{22}$$

Matching coefficients, we find:

- $q_{11} = 4(coefficient\ of\ x_1^2)$
- $q_{22} = 8(coefficient\ of\ x_2^2)$
- $q_{12} = q_{21} = 3$(half the coefficient of x_1x_2, since it appears twice)

Thus, $\mathbf{Q} = \begin{bmatrix} 4 & 3 \\ 3 & 8 \end{bmatrix}$.

b. The gradient ∇f of a quadratic function $f(x) = \frac{1}{2}x^TQx$ is given by $\nabla f(x) = Qx$. Plugging in $x = [1,1]^T$:

$$\nabla f([1,1]^T) = \begin{bmatrix} 4 & 3 \\ 3 & 8 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 4+3 \\ 3+8 \end{bmatrix} = \begin{bmatrix} 7 \\ 11 \end{bmatrix}$$

c. For a quadratic function $f(x) = \frac{1}{2}x^TQx$, the Hessian matrix $\nabla^2 f$ is constant and equal to Q :

$$\nabla^2 f = \begin{bmatrix} 4 & 3 \\ 3 & 8 \end{bmatrix}$$

d. The directional derivative of f at x in the direction of v is given by $\nabla f(x) \cdot v$:

$$\nabla f([-2,0]^T) = \begin{bmatrix} 4 & 3 \\ 3 & 8 \end{bmatrix} \begin{bmatrix} -2 \\ 0 \end{bmatrix} = \begin{bmatrix} -8 \\ -6 \end{bmatrix}$$

Student name: Adi Dereviani Prager
Student ID: 305674731

$$\frac{\partial f}{\partial \mathbf{v}} = \begin{bmatrix} -8 \\ -6 \end{bmatrix} \cdot \begin{bmatrix} -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{bmatrix} = \frac{\sqrt{2}}{2}(8-6) = \sqrt{2}$$

e. The second directional derivative of f in direction v is $v^T \nabla^2 f v$:

$$\frac{\partial^2 f}{\partial \mathbf{v}^2} = \begin{bmatrix} -\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{bmatrix} \begin{bmatrix} 4 & 3 \\ 3 & 8 \end{bmatrix} \begin{bmatrix} -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{bmatrix}$$
$$= \begin{bmatrix} -\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{bmatrix} \begin{bmatrix} -2\sqrt{2} + \frac{3\sqrt{2}}{2} \\ -\frac{3\sqrt{2}}{2} + 4\sqrt{2} \end{bmatrix}$$
$$= \frac{2-3+3-8}{2} = -3$$

<u>**Solution 2:**</u>

Proof that the Intersection of Any Number of Convex Sets is Convex:

Let $\{C_i\}_{i \in I}$ be a family of convex sets in $\mathbb{R}^n$, where I is an index set that could be infinite and not necessarily countable. To prove that the intersection $\bigcap_{i \in I} C_i$ is convex, consider any two points $x,\ y\ \in \bigcap_{i \in I} C_i$. By the definition of intersection, $x,\ y\ \in C_i for\ all\ i\ \in I$.

Let $\alpha \in [0,\ 1]$ and consider the point $z\ =\ \alpha x\ +\ (1\ -\ \alpha)y$. Since each C_i is convex, and $x,\ y\ \in C_i$, we have $z\ \in C_i$ for each i because convexity of C_i implies that any linear combination z of x and y with coefficients summing to 1 also belongs to C_i.

Thus, $z\ \in C_i$ for all $i\ \in I$, which means $z\ \in \bigcap_{i \in I} C_i$. Therefore, $\bigcap_{i \in I} C_i$ is convex.

Analysis of the Union of Convex Sets:

To address whether the union of any number of convex sets is convex, we consider whether for any two points $x,\ y$ in the union of convex sets, the line segment connecting them also lies entirely within the union.

Consider two convex sets A and B that are not subsets of one another but both contain the origin in $\mathbb{R}^n$. Let $x\ \in A \setminus B$ and $y\ \in B \setminus A$. The union $A\ \cup B$ contains both x and y, but the point $\frac{1}{2}x + \frac{1}{2}y$ might not necessarily belong to either A or B if x and y are chosen appropriately (for example, if A and B are disjoint except at the origin).

As a concrete example, let A be the set $\{(x, 0) \in \mathbb{R}^2 : x \geq 0\}$ and B be the set $\{(0, y) \in \mathbb{R}^2 : y \geq 0\}$. Then, $x\ =\ (1,0) \in A$ and $y\ =\ (0,1) \in B$, but the point $\frac{1}{2}(1,0) + \frac{1}{2}(0,1) = \frac{1}{2}(1,1)$ does not belong to either A or B, hence not in $A\ \cup B$. Thus, $A\ \cup B$ is not convex.

Student name: Adi Dereviani Prager
Student ID: 305674731

This example demonstrates that the union of convex sets is not necessarily convex, contrasting sharply with the intersection property.

Conclusion:

The intersection of any number of convex sets is convex, utilizing the fundamental property of convex sets and linear combinations within those sets. However, the union of convex sets is generally not convex, as demonstrated by specific counterexamples showing that the line segment between two points in different sets might not lie entirely within the union.

<u>**Solution 3:**</u>

a. To prove that S_+^n, the set of all symmetric, positive semidefinite matrices in $\mathbb{R}^{n\times n}$, is convex, consider any two matrices $A,\ B\ \in S_+^n$ and any scalar $\alpha \in [0,\ 1]$. We need to show that the matrix $C\ =\ \alpha A\ +\ (1\ -\ \alpha)B$ also belongs to S_+^n.

Since A and B are positive semidefinite, for any vector $x\ \in \mathbb{R}^n$, we have:

1. $x^T A\, x\ \geq 0$

2. $x^T B\, x\ \geq 0$

Considering the matrix C :

$$x^T Cx = x^T(\alpha A + (1-\alpha)B)x = \alpha x^T Ax + (1-\alpha)x^T Bx$$

Since both $x^T A\, x$ and $x^T B\, x$ are non-negative and $\alpha,\ (1-\alpha)$ are non-negative (as they are weights in a convex combination and lie between 0 and 1), the linear combination $\alpha x^T A\, x\ +\ (1\ -\ \alpha)x^T B\, x$ is also non-negative:

$$\alpha x^T A\, x\ +\ (1\ -\ \alpha)x^T B\, x\ \geq 0$$

Hence, $x^T C\, x\ \geq 0$ for all $x\ \in \mathbb{R}^2$, which shows that C is positive semidefinite. Therefore, $C\ \in S_+^n$, proving that S_+^n is convex.

b. To demonstrate that S_+^n is not only a convex set but also a cone, we need to show that for any matrix $A\ \in S_+^n$ and any non-negative scalar $\alpha \geq 0$, the matrix αA also belongs to S_+^n. Furthermore, we need to consider sums of such scaled matrices.

1. **Positive Semidefinite under Scaling:**

For $A\ \in S_+^n$ and $\alpha \geq 0$, consider $x^T(\alpha A)x\ =\ \alpha(x^T A\, x)$. Since $x^T A\, x\ \geq 0$ (by the positive semidefiniteness of A) and $\alpha \geq 0$, it follows that $\alpha(x^T A\, x) \geq 0$. Thus, αA is positive semidefinite, implying $\alpha A\ \in S_+^n$.

Student name: Adi Dereviani Prager
Student ID: 305674731

2. **Addition of Scaled Matrices:**

Let $A,\ B\ \in S_+^n$ and $\alpha,\ \beta \geq 0$. Consider $C\ =\ \alpha A\ +\ \beta B$. For any $x\ \in \mathbb{R}^n$, we have:

$$x^T Cx = x^T(\alpha A + \beta B)x = \alpha x^T Ax + \beta x^T Bx$$

Both $x^T A\,x$ and $x^T B\,x$ are non-negative, and with $\alpha,\ \beta \geq 0$, $\alpha x^T A\,x\ +\ \beta x^T B\,x\ \geq 0$. Hence, C is positive semidefinite, so $C\ \in S_+^n$.

Since S_+^n is closed under non-negative scalar multiplication and addition of such scaled matrices, S_+^n forms a cone.

Solution 4:

a. Let $S = \{x \in \mathbb{R}^n: Ax = b\}$be the set of all solutions x to the matrix equation $Ax\ =\ b$, where A is a $m\ \times n$ matrix and b is a vector in $\mathbb{R}^m$. To show that S is convex, consider two points $x_1,\ x_2 \in S$, which implies that $Ax_1 =\ b$ and $Ax_2 =\ b$.

Let α be a scalar in the interval $[0,\ 1]$. We need to prove that the linear combination $x\ = \alpha x_1 +\ (1\ -\ \alpha)x_2$ also belongs to S . Calculate Ax as follows:

$Ax\ =\ A(\alpha x_1 +\ (1\ -\ \alpha)x_2) = \alpha Ax_1 + (1 - \alpha)Ax_2$

Since $Ax_1 =\ b$ and $Ax_2 =\ b$, substituting these into the equation gives:

$Ax = \alpha b\ +\ (1\ -\ \alpha)b = \big(\alpha +\ (1\ -\ \alpha)\big)b = b$

Thus, $Ax\ =\ b$, which means x is also a solution to the equation $Ax\ =\ b$, and therefore$x\ \in S$. Since $x_1,\ x_2$ were arbitrary points in S and α was any number in $[0,\ 1]$, this shows that S is convex.

b. An affine set is defined as a set which, if it contains any two points, contains the entire line that passes through these points, not just the line segment between them. To show this property for S , consider again two points $x_1,\ x_2 \in S$ such that $Ax_1 =\ b$ and $Ax_2 =\ b$.

Consider any point x on the line supported by x_1 and x_2, given by the equation $x\ =\ x_1 + t(x_2 -\ x_1)$ for any scalar $t\ \in \mathbb{R}$. Compute Ax to check if it equals b :

$Ax = A\big(x_1 +\ t(x_2 -\ x_1)\big) = Ax_1 +\ tA(x_2 -\ x_1) = b + t(Ax_2 - Ax_1)$

Since $Ax_1 =\ b$ and $Ax_2 =\ b$, the term $t(Ax_2 -\ Ax_1)$ simplifies to $t(b\ -\ b) =\ 0$. Thus:

$Ax = b\ +\ 0\ = b$

Student name: Adi Dereviani Prager
Student ID: 305674731

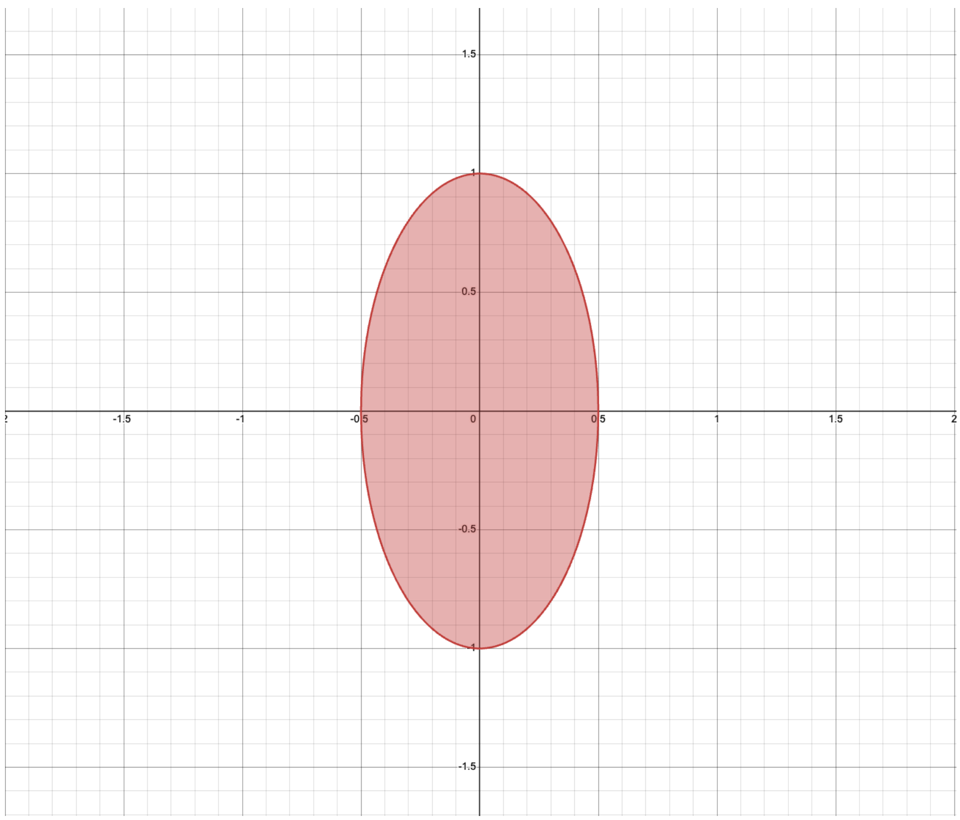
Hence, $x \in S$ for any $t \in \mathbb{R}$, demonstrating that the entire line through x_1 and x_2 is contained within S .

Solution 5:

a. For the function $f(x, y) = 4x^2 + y^2$, the sub-level sets for $c = 1$ and $c = 2$ can be described as follows:

1. **For** $c = 1$ **:**

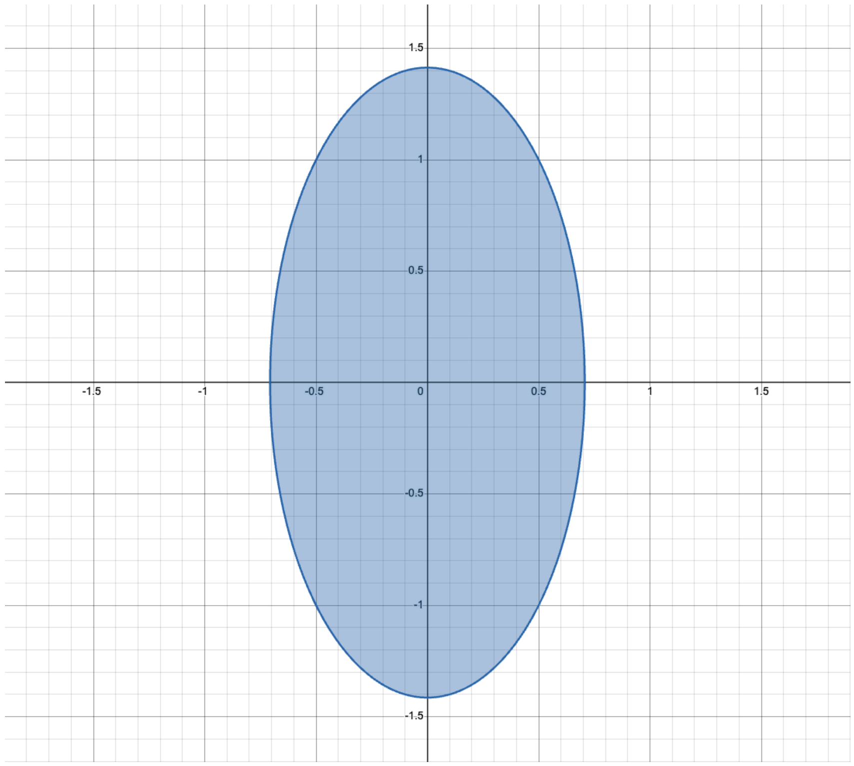
The sub-level set is defined by $\{(x, y) \in \mathbb{R}^2 : 4x^2 + y^2 \leq 1\}$. This set represents an ellipse centered at the origin, with the major axis along the y -axis and the minor axis along the x -axis. The lengths of the semi-axes are determined by setting $4x^2$ and y^2 equal to 1, giving $x = \pm\frac{1}{2}$ and $y = \pm 1$.



2. **For** $c = 2$ **:**

The sub-level set is defined by $\{(x, y) \in \mathbb{R}^2 : 4x^2 + y^2 \leq 2\}$. Similarly, this set is also an ellipse centered at the origin. The lengths of the semi-axes in this case are found by setting $4x^2$ and y^2 equal to 2, giving $x = \pm\frac{1}{\sqrt{2}}$ and $y = \pm\sqrt{2}$.

Student name: Adi Dereviani Prager
Student ID: 305674731



b. Proof:

Assume $u,\ v$ are in the sub-level set $S = \{x \in \mathbb{R}^n: f(x) \leq c\}$ for some convex function $f\colon \mathbb{R}^n \to \mathbb{R}$. Let $w = \alpha u + (1 - \alpha)v$ for some $\alpha \in [0,\ 1]$. We need to show that w also belongs to S, which requires demonstrating that $f(w) \leq c$.

Given the convexity of f, we have:

$f(w) = f(\alpha u + (1 - \alpha)v) \leq \alpha f(u) + (1 - \alpha)f(v)$

Since $u,\ v \in S$, it follows that $f(u) \leq c$ and $f(v) \leq c$. Applying these inequalities, we obtain:

$$\alpha f(u) + (1 - \alpha)f(v) \leq \alpha c + (1 - \alpha)c = c$$

In the first inequality, the convex nature of f is applied, while in the second, we leverage the non-negativity of α and $1 - \alpha$ along with $u,\ v$ being elements of S. Consequently, $f(w) \leq c$ demonstrates that $w \in S$, confirming the convexity of S.

<u>**Solution 6:**</u>

a. Given that $f\colon D \subset \mathbb{R}^n \to \mathbb{R}$ is a convex function and $h\colon \mathbb{R} \to \mathbb{R}$ is convex and monotone increasing, we need to prove that the composition $h \circ f$ is convex.

Student name: Adi Dereviani Prager
Student ID: 305674731

Consider two points $u,\ v\ \in D$ and a scalar $\alpha \in [0,\ 1]$. From the convexity of f , we know:

$$f(\alpha u + (1-\alpha)v) \leq \alpha f(u) + (1-\alpha)f(v)$$

Using the monotonicity of h , where $t_1 >\ t_2$ implies $h(t_1) \geq h(t_2)$, we can apply h to both sides of the inequality:

$$h\big(f(\alpha u + (1-\alpha)v)\big) \leq h\big(\alpha f(u) + (1-\alpha)f(v)\big)$$

Since h is also convex, the right-hand side can be further bounded by:

$$h\big(\alpha f(u) + (1-\alpha)f(v)\big) \leq \alpha h\big(f(u)\big) + (1-\alpha)h\big(f(v)\big)$$

Thus, combining these inequalities gives:

$h\ \circ f(\alpha u\ +\ (1\ -\ \alpha)v) \leq \alpha h \circ f(u) + (1-\alpha)h \circ f(v)$

This confirms that $h\ \circ f$ is convex, as it satisfies the definition of convexity.

b. Consider the function $f(x) =\ x^2$, which is convex. Let $h(y) =\ \ -y$, which is neither monotone increasing nor convex, but for this part, the key is the lack of monotonicity.

The composition $h\ \circ f(x) =\ \ -x^2$ results in a function that is concave, not convex. To see this:

$-(x^2)$ has a downward-opening parabola, clearly not satisfying the convexity condition.

This shows that dropping the monotonicity of h can lead to $h\ \circ f$ being non-convex.

c. Consider $f(x) =\ x$, which is linear and therefore convex. Let $h(y) = \log y$, which is monotone increasing but not convex over its entire domain (since the second derivative $\frac{d^2}{dy^2}\log y = -\frac{1}{y^2}$ is negative).

The composition $h\ \circ f(x)$ = $\log x$ is defined for $x\ >\ 0$ and is not convex, as it does not satisfy the definition of convexity for all $x\ >\ 0$. A specific example can illustrate the failure:

$$\log(0.5x_1 + 0.5x_2) > 0.5\log(x_1) + 0.5\log(x_2)$$

for certain values of x_1 and x_2 (e.g., $x_1 =\ 1,\ x_2 =\ 4$).

<u>Solution 7:</u>

a. Let A be a matrix in $\mathbb{R}^{m\times n}$. To show that $A^T A$ is positive semi-definite, consider any vector $x\ \in \mathbb{R}^n$. We need to show that $x^T A^T A\, x\ \geq 0$.

Compute $x^T A^T A\, x$ as follows:

Student name: Adi Dereviani Prager
Student ID: 305674731

$x^T A^T A\, x = (Ax)^T (Ax)$

This expression represents the dot product of the vector Ax with itself, which is always non-negative:

$$(Ax)^T (Ax) = \| Ax \|^2 \geq 0$$

The squared norm $\| Ax \|^2$ is non-negative for all x, confirming that $x^T A^T Ax \geq 0$ for any x. Therefore, $A^T A$ is positive semi-definite.

b. Assume $\dim \ker(A) = 0$, which implies that A has trivial kernel, or in other words, A is of full rank. We need to show under this condition that $A^T A$is not just positive semi-definite, but positive definite.

To show that $A^T A$ is positive definite, we must prove that for any non-zero vector $x \in \mathbb{R}^n$, $x^T A^T A\, x > 0$.

From the assumption $\dim \ker(A) = 0$, it follows that $Ax \neq 0$ whenever $x \neq 0$. Therefore, the squared norm $\| Ax \|^2$ is strictly positive for all non-zero x:

$$x^T A^T Ax = \| Ax \|^2 > 0 \; for \; all \; x \neq 0$$

This strictly positive result for all non-zero x satisfies the definition of positive definiteness.

<u>Solution 8:</u>

a. Given the function $f(x_1, x_2) = 2 + 6x_1 + 2x_2 + 3x_1^2 - x_2^2$, we are to express this function in the form $f(x) = \frac{1}{2} x^T Q\, x + q^T x + c$.

1. Quadratic Form:

The quadratic terms in f are $3x_1^2$ and $-x_2^2$. To represent these in matrix form, we recognize that each coefficient in the quadratic form $x^T Q\, x$ must be halved in the matrix when writing in the form $\frac{1}{2} x^T Q\, x$. Thus:

$$Q = \begin{bmatrix} 6 & 0 \\ 0 & -2 \end{bmatrix}$$

The matrix Q is constructed such that

$$\frac{1}{2} [x_1 \; x_2] \begin{bmatrix} 6 & 0 \\ 0 & -2 \end{bmatrix} [x_1 \; x_2]^T$$

yields $3x_1^2 - x_2^2$.

2. Linear and Constant Terms:

The linear terms $6x_1$ and $2x_2$, and the constant term 2 can be directly taken as q and c, respectively:

$$q = \begin{bmatrix} 6 \\ 2 \end{bmatrix}$$

c = 2

Hence, $f(x)$ can be written as:

$$f(x) = \frac{1}{2}[x_1\ x_2] \begin{bmatrix} 6 & 0 \\ 0 & -2 \end{bmatrix} [x_1\ x_2]^T + [6\ 2][x_1\ x_2]^T + 2$$

b. To find the stationary points, compute the gradient $\nabla f(x)$ and set it to zero:

$\nabla f(x_1,\ x_2)= Qx + q$

$$\nabla f(x_1, x_2) = \begin{bmatrix} 6 & 0 \\ 0 & -2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 6 \\ 2 \end{bmatrix} = \begin{bmatrix} 6x_1 + 6 \\ -2x_2 + 2 \end{bmatrix}$$

Setting $\nabla f(x_1,\ x_2)$ to zero, we solve:

$6x_1 +\ 6 = 0 \Rightarrow\ x_1 = -1$

$-2x_2 +\ 2 = 0 \Rightarrow\ x_2 = 1$

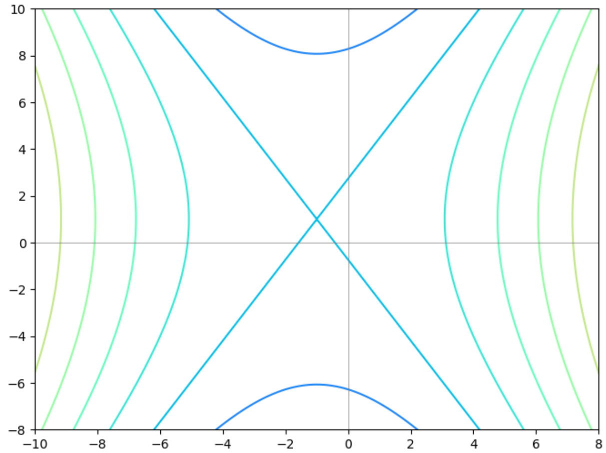
The stationary point is $(-1,\ 1)$.

Checking if the Point is a Saddle Point:

The nature of the stationary point is determined by the eigenvalues of Q. If Q has both positive and negative eigenvalues, the stationary point is a saddle point. The matrix Q has eigenvalues 6 and -2, which are of opposite signs. This indicates the presence of both upward and downward curvature in different directions.

Therefore, the stationary point $(-1,\ 1)$ is not a maximum or a minimum, but a saddle point, as the eigenvalues of Q $(6,\ -2)$ are sign-indefinite.

Student name: Adi Dereviani Prager
Student ID: 305674731

c. 

<u>Solution 9:</u>

a. For the Rosenbrock function defined as $f(x) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2$, we calculate the gradient $\nabla f(x)$ and the Hessian $\nabla^2 f(x)$ by straightforward differentiation:

$$\nabla f(x_1, x_2) = \begin{bmatrix} 200(x_2 - x_1^2)(-2x_1) - 2(1 - x_1) \\ 200(x_2 - x_1^2) \end{bmatrix} = \begin{bmatrix} -400(x_1 x_2 - x_1^3) - 2(1 - x_1) \\ 200(x_2 - x_1^2) \end{bmatrix}$$

Therefore, the Hessian matrix is:

$$\nabla^2 f(x_1, x_2) = \begin{bmatrix} -400x_2 + 1200x_1^2 + 2 & -400x_1 \\ -400x_1 & 200 \end{bmatrix}$$

b. To determine a stationary point, we first calculate the partial derivative with respect to x_2, given by $\frac{\partial f}{\partial x_2} = 200(x_2 - x_1^2)$. Setting this to zero, we have: $x_2 - x_1^2 = 0 \Rightarrow x_2 = x_1^2$.

Next, we substitute $x_2 = x_1^2$ into $\frac{\partial f}{\partial x_2} = -400(x_1 x_2 - x_1^3) - 2(1 - x_1)$, yielding:

$-400(x_1^3 - x_1^3) - 2(1 - x_1) = -2(1 - x_1)$.

Setting this to zero, we find $x_1 = 1$. Consequently, $x_2 = 1^2 = 1$, leading us to the stationary point at $[1,1]^T$.

For the Hessian matrix at this point, we compute:

$$\nabla^2 f(1,1) = \begin{bmatrix} -400 + 1200 + 2 & -400 \\ -400 & 200 \end{bmatrix} = \begin{bmatrix} 802 & -400 \\ -400 & 200 \end{bmatrix}$$

Student name: Adi Dereviani Prager
Student ID: 305674731

Checking for positive definiteness, the leading principal minors, 802 and $802 \times 200 - (-400) \times (-400) = 400$, are positive.

Therefore, the Hessian is positive definite, indicating that $[1,1]^T$ is a local minimizer and the only one.