YZV202E - Homework 1

Student: İbrahim Bancar, bancar22@itu.edu.tr
Student Number: 150220313,

Question 1

Consider the function:

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

This is a non-linear function of two variables. The goal is to analyze its critical points, classify them, and visualize its behavior.

(a) **Find all stationary points and their corresponding values.**

In order to find the stationary (or critical) points of the function, first compute the gradient vector and set it equal to zero. This gives us the necessary conditions for optimality (also known as first-order necessary conditions, FONC). The stationary points are candidates for local minima, local maxima, or saddle points.

$$\nabla f = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \end{bmatrix} = \begin{bmatrix} 4x_1^3 - 4x_1 x_2 + 2x_1 \\ -2x_1^2 + 4x_2^3 - 2x_2 \end{bmatrix} = \vec{0}$$

Solving the following nonlinear system:

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

This system has two primary cases based on equation (1):

Case 1: $x_1 = 0$

Substituting this into equation (2):

$$-2(0)^2 + 4x_2^3 - 2x_2 = 0 \Rightarrow 4x_2^3 - 2x_2 = 0 \Rightarrow 2x_2(2x_2^2 - 1) = 0$$

This gives:

$$x_2 = 0, \quad x_2 = \pm\frac{1}{\sqrt{2}}$$

Hence, three stationary points:

$$(0, 0), \quad \left(0, \frac{1}{\sqrt{2}}\right), \quad \left(0, -\frac{1}{\sqrt{2}}\right)$$

Case 2: From equation (1), suppose $x_1 \neq 0$. Then by solve equation (1):

$$4x_1^2 - 4x_2 + 2 = 0$$
$$x_1^2 = x_2 - \frac{1}{2}$$

Now, substituting this expression only into the x_1^2 term in equation (2):

$$-2x_1^2 + 4x_2^3 - 2x_2 = 0$$

$$\Rightarrow -2(x_2 - \tfrac{1}{2}) + 4x_2^3 - 2x_2 = 0$$

$$\Rightarrow -2x_2 + 1 + 4x_2^3 - 2x_2 = 0 \Rightarrow 4x_2^3 - 4x_2 + 1 = 0$$

Solving the the cubic equation:

$$g(x_2) = 4x_2^3 - 4x_2 + 1 = 0$$

This is a degree-3 polynomial, which can have up to three real roots. To analyze the number and location of the roots, I visualized the function in Python using the plot below:

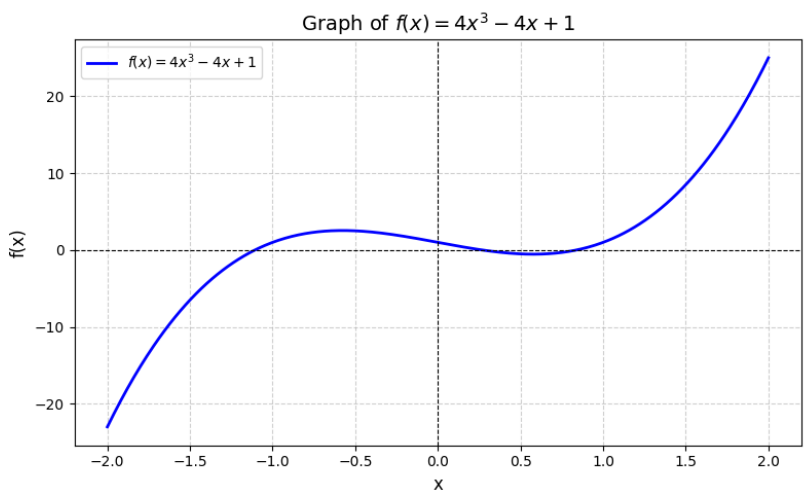

Figure 1: Plot of $f(x) = 4x^3 - 4x + 1$

As seen in the graph, the function intersects the x-axis at three distinct locations, confirming the existence of three real roots. These are approximately located at:

$$x_2^{(1)} \approx -1.107$$
$$x_2^{(2)} \approx 0.215$$
$$x_2^{(3)} \approx 0.896$$

Now test which of these roots correspond to valid stationary points by checking whether the implied value of $x_1^2 = x_2 - \frac{1}{2}$ is non-negative:

- For $x_2^{(1)} \approx -1.107$, it has $x_1^2 \approx -1.607 \rightarrow$ not real
- For $x_2^{(2)} \approx 0.215$, it has $x_1^2 \approx -0.285 \rightarrow$ not real
- For $x_2^{(3)} \approx 0.896$, it has $x_1^2 \approx 0.396 \rightarrow$ valid

Thus, only the third root corresponds to real values of both x_1 and x_2. Specifically:

$$x_1 \approx \pm\sqrt{0.396} \approx \pm 0.63, \quad x_2 \approx 0.896$$

Therefore, Case 2 yields two additional stationary points in the real plane:

$$\boxed{(0.63,\ 0.896), \quad (-0.63,\ 0.896)}$$

Updated list of stationary points:

- $(0,0)$
- $\left(0, \frac{1}{\sqrt{2}}\right)$:
- $\left(0, -\frac{1}{\sqrt{2}}\right)$:
- $(0.63,\ 0.896)$
- $(-0.63,\ 0.896)$

Computing the function values at these points:

$$\begin{aligned}
f(0,0) &= 0^4 - 2(0)^2(0) + 0^2 + 0^4 - 0^2 = 0 \\
f\left(0, \frac{1}{\sqrt{2}}\right) &= 0 + 0 + 0 + \left(\frac{1}{\sqrt{2}}\right)^4 - \left(\frac{1}{\sqrt{2}}\right)^2 = \frac{1}{4} - \frac{1}{2} = -\frac{1}{4} \\
f\left(0, -\frac{1}{\sqrt{2}}\right) &= 0 + 0 + 0 + \left(-\frac{1}{\sqrt{2}}\right)^4 - \left(-\frac{1}{\sqrt{2}}\right)^2 = \frac{1}{4} - \frac{1}{2} = -\frac{1}{4}
\end{aligned}$$

For stationary points obtained from Case 2, numerical approximations are used to estimate the function values. Using the relations:

$$x_1^2 \approx 0.396, \quad x_1^4 = (x_1^2)^2 \approx 0.157, \quad x_2^2 \approx 0.803, \quad x_2^4 = (x_2^2)^2 \approx 0.645$$

Substituting these into the function:

$$\begin{aligned}
f(x_1, x_2) &= x_1^4 - 2x_1^2 x_2 + x_1^2 + x_2^4 - x_2^2 \\
&\approx 0.157 - 2(0.396)(0.896) + 0.396 + 0.645 - 0.803 \\
&= 0.157 - 0.709 + 0.396 + 0.645 - 0.803 \\
&\approx -0.314
\end{aligned}$$

Since the function is symmetric in x_1, both points $(0.63, 0.896)$ and $(-0.63, 0.896)$ yield the same value.

Final list of all stationary points and corresponding function values:

Stationary Point	**Function Value**
$(0,0)$	0
$\left(0, \frac{1}{\sqrt{2}}\right)$	$-\frac{1}{4}$
$\left(0, -\frac{1}{\sqrt{2}}\right)$	$-\frac{1}{4}$
$(0.63,\ 0.896)$	≈ -0.314
$(-0.63,\ 0.896)$	≈ -0.314

(b) **Classification of stationary points**

To determine the nature of each stationary point, the Hessian matrix is evaluated at that point and its definiteness is analyzed. The Hessian matrix, which is the matrix of second-order partial derivatives, provides information about the curvature of the function. In the context of a function $f(x_1, x_2)$, the Hessian is given by:

$$H(x_1, x_2) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} \end{bmatrix}$$

There are multiple methods to determine the nature of a stationary point, such as the eigenvalue test or leading principal minors, but in this report the eigenvalue method is used. A stationary point is classified according to the following criteria:

(a) **Positive definiteness:** If the Hessian matrix at the stationary point is positive definite (i.e., all eigenvalues are positive), then the function is locally convex at that point, and the stationary point is a local minimum.

(b) **Positive semidefiniteness:** If all eigenvalues are non-negative (i.e., some may be zero), then the Hessian is positive semidefinite. In this case, the point is a candidate for a local minimum, but further analysis is required to determine its nature.

(c) **Negative definiteness:** If all eigenvalues of the Hessian are negative, the function is locally concave, and the stationary point is a local maximum.

(d) **Negative semidefiniteness:** If all eigenvalues are non-positive (i.e., some may be zero), then the Hessian is negative semidefinite. The point is a candidate for a local maximum, but again, further investigation is needed to confirm this.

(e) **Indefiniteness:** If the Hessian has both positive and negative eigenvalues, the function curves in different directions, and the stationary point is classified as a saddle point.

Thus, by computing the Hessian at each stationary point and applying these conditions, the nature of the point (local minimum, local maximum, or saddle point) is determined.

The Hessian matrix of the function

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

is computed as:

$$H(x_1, x_2) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} \end{bmatrix} = \begin{bmatrix} 12x_1^2 - 4x_2 + 2 & -4x_1 \\ -4x_1 & 12x_2^2 - 2 \end{bmatrix}$$

Point 1: $(0, 0)$

$$H(0,0) = \begin{bmatrix} 2 & 0 \\ 0 & -2 \end{bmatrix} \quad \Rightarrow \quad \text{Eigenvalues: } \lambda_1 = 2 > 0, \ \lambda_2 = -2 < 0$$

The Hessian has eigenvalues of opposite signs **saddle point**.

Point 2: $\left(0, \frac{1}{\sqrt{2}}\right)$

$$x_1 = 0, \quad x_2 = \frac{1}{\sqrt{2}}, \quad x_2^2 = \frac{1}{2}$$

$$H\left(0, \frac{1}{\sqrt{2}}\right) = \begin{bmatrix} 2 - 4 \cdot \frac{1}{\sqrt{2}} & 0 \\ 0 & 12 \cdot \frac{1}{2} - 2 \end{bmatrix} = \begin{bmatrix} 2 - 2\sqrt{2} & 0 \\ 0 & 4 \end{bmatrix}$$

$$\lambda_1 \approx -0.828 < 0, \quad \lambda_2 = 4 > 0 \Rightarrow \textbf{saddle point}$$

Point 3: $\left(0, -\frac{1}{\sqrt{2}}\right)$

$$x_1 = 0, \quad x_2 = -\frac{1}{\sqrt{2}}, \quad x_2^2 = \frac{1}{2}$$

$$H\left(0, -\frac{1}{\sqrt{2}}\right) = \begin{bmatrix} 2 + 2\sqrt{2} & 0 \\ 0 & 4 \end{bmatrix} \Rightarrow \lambda_1 \approx 4.828 > 0, \ \lambda_2 = 4 > 0$$

Both eigenvalues positive so it's positive definite and the function has a **local minimum**

Point 4 and 5: $(\pm 0.63, \ 0.896)$

Using approximations:

$$x_1^2 \approx 0.396, \quad x_2 \approx 0.896, \quad x_2^2 \approx 0.803$$

Then the Hessian becomes:

$$H \approx \begin{bmatrix} 12(0.396) - 4(0.896) + 2 & -4(0.63) \\ -4(0.63) & 12(0.803) - 2 \end{bmatrix} = \begin{bmatrix} 3.168 & -2.52 \\ -2.52 & 7.636 \end{bmatrix}$$

The eigenvalues of this symmetric matrix are computed as:

$$\lambda_{1,2} = \frac{\mathrm{tr}(H) \pm \sqrt{\mathrm{tr}(H)^2 - 4 \cdot \det(H)}}{2}$$

where:

$$\mathrm{tr}(H) = 3.168 + 7.636 = 10.804, \quad \det(H) = 3.168 \cdot 7.636 - (-2.52)^2 \approx 24.188 - 6.350 = 17.838$$

Compute discriminant:

$$\Delta = \mathrm{tr}(H)^2 - 4 \cdot \det(H) = (10.804)^2 - 4(17.838) \approx 116.7 - 71.35 = 45.35$$

Then:

$$\lambda_1 \approx \frac{10.804 + \sqrt{45.35}}{2} \approx \frac{10.804 + 6.73}{2} \approx 8.77, \quad \lambda_2 \approx \frac{10.804 - 6.73}{2} \approx 2.04$$

Both eigenvalues are positive, hence the Hessian is **positive definite**, and the point is classified as a **local minimum**.

Summary of classification:

Stationary Point	Type
$(0, 0)$	Saddle point
$\left(0, \frac{1}{\sqrt{2}}\right)$	Saddle point
$\left(0, -\frac{1}{\sqrt{2}}\right)$	Local minimum
$(0.63,\ 0.896)$	Local minimum
$(-0.63,\ 0.896)$	Local minimum

(c) **3D surface plot of the function**

To visualize the behavior of the function $f(x_1, x_2)$, a 3D surface plot was created using the interval $x_1, x_2 \in [-2, 2]$, which includes all stationary points identified in part (a). The purpose of this visualization is to gain an intuitive understanding of the surface geometry and the nature of each stationary point (minima,maxima and saddle points).

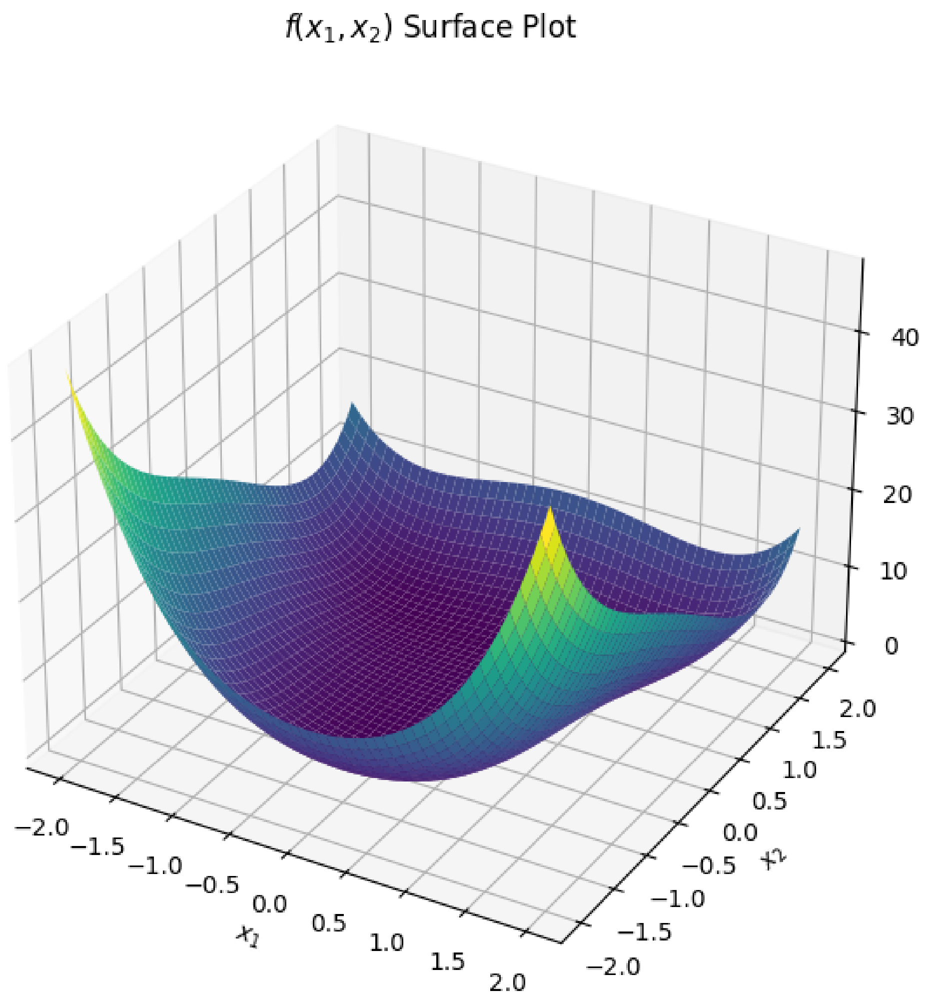

Figure 2: 3D surface plot of $f(x_1, x_2) = x_1^4 - 2x_1^2x_2 + x_1^2 + x_2^4 - x_2^2$

The plot reveals a saddle-like behavior around the origin, and distinct valleys corresponding to local minima near the numerically obtained stationary points. The function surface contains regions of both convexity and concavity. All five stationary points are located within the chosen plotting interval.

(d) **Contour plot with overlaid gradient directions**

To visualize the local behavior of the function and validate the stationary points identified, a contour plot of $f(x_1, x_2)$ was generated with the gradient vectors overlaid on top. This allows us to observe how the gradient field behaves in different regions of the domain.

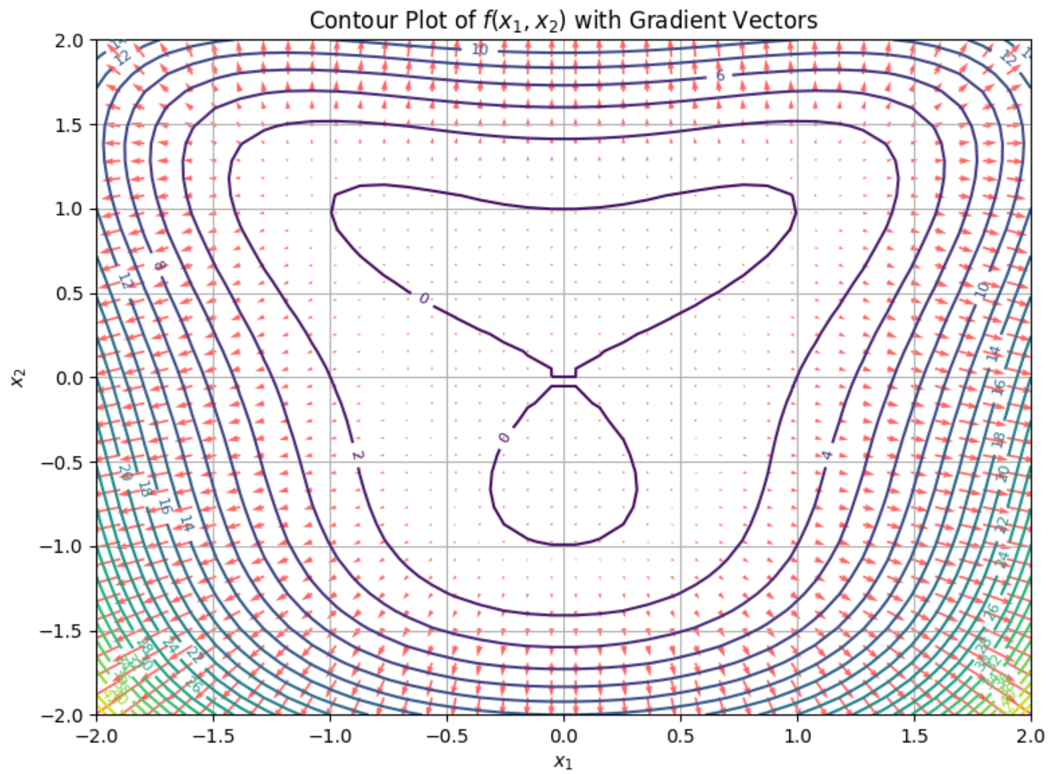

Figure 3: Contour plot of $f(x_1, x_2)$ with overlaid gradient vectors

The red arrows represent the direction and magnitude of the gradient $\nabla f(x_1, x_2)$, indicating the direction of steepest ascent. As expected, the gradient vectors are orthogonal to the contour lines. In regions near saddle points (e.g., around the origin), the directions change rapidly, while near local minima the vectors tend to point inward. This graphical representation provides further insight into the curvature and optimization landscape of the function.

Question 2

Consider a linear map $T : \mathbb{R}^3 \to \mathbb{R}^3$ defined by the matrix

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 2 & 2 \\ 3 & 2 & 1 \end{bmatrix}$$

(a) **The range of T is a subspace of $\mathbb{R}^3$.**

To show that the range (or image) of a transformation is a subspace, one may verify the following three conditions:

- The zero vector is contained in the range;
- The range is closed under vector addition;
- The range is closed under scalar multiplication.

These properties define any subspace of $\mathbb{R}^n$. However, since T is a linear transformation, all three conditions are automatically satisfied by the definition of linearity. That is, a linear map $T : \mathbb{R}^3 \to \mathbb{R}^3$ satisfies

$$T(\alpha\vec{x} + \beta\vec{y}) = \alpha T(\vec{x}) + \beta T(\vec{y}),$$

for all $\vec{x}, \vec{y} \in \mathbb{R}^3$ and all scalars $\alpha, \beta \in \mathbb{R}$. This implies that the image of T must be closed under addition and scalar multiplication, and since $T(\vec{0}) = \vec{0}$, it contains the zero vector as well. Therefore, the range of any linear transformation is always a subspace of its codomain.

In this particular case, the transformation T is defined by the matrix A, and the range of T corresponds to the column space of A. Since the column space of a matrix is always a subspace of $\mathbb{R}^n$, it follows again that the range of T is a subspace of $\mathbb{R}^3$.

To illustrate this concretely, consider the matrix:

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 2 & 2 \\ 3 & 2 & 1 \end{bmatrix}$$

Let the columns of A be denoted by $\vec{a}_1, \vec{a}_2, \vec{a}_3$, where:

$$\vec{a}_1 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \quad \vec{a}_2 = \begin{bmatrix} 2 \\ 2 \\ 2 \end{bmatrix}, \quad \vec{a}_3 = \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix}$$

To determine the dimension of the range, examining whether these vectors are linearly independent by computing the rank of A. Applying row reduction:

$$\begin{bmatrix} 1 & 2 & 3 \\ 2 & 2 & 2 \\ 3 & 2 & 1 \end{bmatrix} \xrightarrow{R_2 \leftarrow R_2 - 2R_1} \begin{bmatrix} 1 & 2 & 3 \\ 0 & -2 & -4 \\ 3 & 2 & 1 \end{bmatrix} \xrightarrow{R_3 \leftarrow R_3 - 3R_1} \begin{bmatrix} 1 & 2 & 3 \\ 0 & -2 & -4 \\ 0 & -4 & -8 \end{bmatrix} \xrightarrow{R_3 \leftarrow R_3 - 2R_2} \begin{bmatrix} 1 & 2 & 3 \\ 0 & -2 & -4 \\ 0 & 0 & 0 \end{bmatrix}$$

The matrix has two pivot positions, so $\text{rank}(A) = 2$. This means the columns span a two-dimensional subspace of $\mathbb{R}^3$, i.e., a plane through the origin.

$$\boxed{\text{The range of } T \text{ is a 2-dimensional subspace of } \mathbb{R}^3}$$

(b) **Diagonalizability of A and construction of an eigenbasis for $\mathbb{R}^3$**

To determine whether the matrix A is diagonalizable, consider computing its eigenvalues. This involves solving the characteristic equation $\det(A - \lambda I) = 0$:

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 2 & 2 \\ 3 & 2 & 1 \end{bmatrix}$$

Computing the characteristic polynomial by evaluating $\det(A - \lambda I)$, where:

$$A - \lambda I = \begin{bmatrix} 1-\lambda & 2 & 3 \\ 2 & 2-\lambda & 2 \\ 3 & 2 & 1-\lambda \end{bmatrix}$$

Now compute the determinant using cofactor method:

$$\begin{aligned} \det(A - \lambda I) = (1-\lambda) &\begin{vmatrix} 2-\lambda & 2 \\ 2 & 1-\lambda \end{vmatrix} - 2 \begin{vmatrix} 2 & 2 \\ 3 & 1-\lambda \end{vmatrix} + 3 \begin{vmatrix} 2 & 2-\lambda \\ 3 & 2 \end{vmatrix} \\ &= (1-\lambda)\left[(2-\lambda)(1-\lambda) - (2)(2)\right] - 2\left[2(1-\lambda) - 2(3)\right] + 3\left[2(2) - 3(2-\lambda)\right] \\ &= (1-\lambda)\left[(2-\lambda)(1-\lambda) - 4\right] - 2\left[2 - 2\lambda - 6\right] + 3\left[4 - 6 + 3\lambda\right] \\ &= (1-\lambda)\left[(2-\lambda)(1-\lambda) - 4\right] - 2(-4 + 2\lambda) + 3(-2 + 3\lambda) \end{aligned}$$

Expanding each term:

$$(2-\lambda)(1-\lambda) = 2(1-\lambda) - \lambda(1-\lambda) = 2 - 2\lambda - \lambda + \lambda^2 = \lambda^2 - 3\lambda + 2$$

Thus:

$$\begin{aligned} \det(A - \lambda I) &= (1-\lambda)(\lambda^2 - 3\lambda + 2 - 4) + 8 - 4\lambda + (-6 + 9\lambda) \\ &= (1-\lambda)(\lambda^2 - 3\lambda - 2) + 8 - 4\lambda - 6 + 9\lambda \\ &= (1-\lambda)(\lambda^2 - 3\lambda - 2) + (2 + 5\lambda) \end{aligned}$$

After full expansion and simplification:

$$\det(A - \lambda I) = \lambda(\lambda - 6)(\lambda + 2)$$

Thus, the eigenvalues are:

$$\boxed{\lambda_1 = 0, \quad \lambda_2 = 6, \quad \lambda_3 = -2}$$

Since all three eigenvalues are real and distinct, the matrix A is diagonalizable. Moreover, a 3×3 matrix with three distinct eigenvalues always has three linearly independent eigenvectors.

To diagonalize A, the eigenvectors corresponding to each eigenvalue should be computed.

- **For $\lambda = 0$:**
  Solving the homogeneous system:

$$(A - 0I)\vec{v} = A\vec{v} = 0 \Rightarrow \begin{bmatrix} 1 & 2 & 3 \\ 2 & 2 & 2 \\ 3 & 2 & 1 \end{bmatrix} \vec{v} = 0$$

  Row reduce the augmented matrix:

$$\begin{bmatrix} 1 & 2 & 3 \\ 2 & 2 & 2 \\ 3 & 2 & 1 \end{bmatrix} \xrightarrow{R_2 \leftarrow R_2 - 2R_1} \begin{bmatrix} 1 & 2 & 3 \\ 0 & -2 & -4 \\ 3 & 2 & 1 \end{bmatrix}$$

$$\xrightarrow{R_3 \leftarrow R_3 - 3R_1} \begin{bmatrix} 1 & 2 & 3 \\ 0 & -2 & -4 \\ 0 & -4 & -8 \end{bmatrix} \xrightarrow{R_3 \leftarrow R_3 - 2R_2} \begin{bmatrix} 1 & 2 & 3 \\ 0 & -2 & -4 \\ 0 & 0 & 0 \end{bmatrix}$$

  From the reduced matrix:

$$\begin{bmatrix} 1 & 2 & 3 \\ 0 & -2 & -4 \\ 0 & 0 & 0 \end{bmatrix} \Rightarrow \begin{cases} x_1 + 2x_2 + 3x_3 = 0 \\ -2x_2 - 4x_3 = 0 \end{cases}$$

  Solving the second equation:

$$x_2 = -2x_3 \Rightarrow x_1 + 2(-2x_3) + 3x_3 = x_1 - 4x_3 + 3x_3 = x_1 - x_3 = 0 \Rightarrow x_1 = x_3$$

  Let $x_3 = t$. Then:

$$\vec{v}_0 = t \cdot \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix} \Rightarrow \boxed{\vec{v}_0 = \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}}$$

- **For $\lambda = 6$:**
  Solving the homogeneous system:

$$(A - 6I)\vec{v} = 0 \Rightarrow \begin{bmatrix} -5 & 2 & 3 \\ 2 & -4 & 2 \\ 3 & 2 & -5 \end{bmatrix} \vec{v} = 0$$

  Apply row reduction:

$$\begin{bmatrix} -5 & 2 & 3 \\ 2 & -4 & 2 \\ 3 & 2 & -5 \end{bmatrix} \xrightarrow{R_1 \leftrightarrow R_2} \begin{bmatrix} 2 & -4 & 2 \\ -5 & 2 & 3 \\ 3 & 2 & -5 \end{bmatrix} \xrightarrow{R_2 \leftarrow R_2 + \frac{5}{2}R_1} \begin{bmatrix} 2 & -4 & 2 \\ 0 & -8 & 8 \\ 3 & 2 & -5 \end{bmatrix}$$

$$\xrightarrow{R_3 \leftarrow R_3 - \frac{3}{2}R_1} \begin{bmatrix} 2 & -4 & 2 \\ 0 & -8 & 8 \\ 0 & 8 & -8 \end{bmatrix} \xrightarrow{R_3 \leftarrow R_3 + R_2} \begin{bmatrix} 2 & -4 & 2 \\ 0 & -8 & 8 \\ 0 & 0 & 0 \end{bmatrix}$$

  From this reduced matrix, getting the system:

$$\begin{cases} 2x_1 - 4x_2 + 2x_3 = 0 \\ -8x_2 + 8x_3 = 0 \end{cases} \Rightarrow x_2 = x_3, \quad x_1 = x_3$$

Let $x_3 = t$, then:

$$\vec{v}_6 = t \cdot \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \Rightarrow \boxed{\vec{v}_6 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}}$$

- **For $\lambda = -2$:**
  Solving:

$$(A + 2I)\vec{v} = 0 \Rightarrow \begin{bmatrix} 3 & 2 & 3 \\ 2 & 4 & 2 \\ 3 & 2 & 3 \end{bmatrix} \vec{v} = 0$$

  Apply row reduction:

$$\begin{bmatrix} 3 & 2 & 3 \\ 2 & 4 & 2 \\ 3 & 2 & 3 \end{bmatrix} \xrightarrow{R_2 \leftarrow R_2 - \frac{2}{3}R_1} \begin{bmatrix} 3 & 2 & 3 \\ 0 & \frac{8}{3} & 0 \\ 3 & 2 & 3 \end{bmatrix}$$

$$\xrightarrow{R_1 \leftarrow \frac{1}{3}R_1, \quad R_2 \leftarrow \frac{3}{8}R_2} \begin{bmatrix} 1 & \frac{2}{3} & 1 \\ 0 & 1 & 0 \\ 3 & 2 & 3 \end{bmatrix} \xrightarrow{R_3 \leftarrow R_3 - 3R_1} \begin{bmatrix} 1 & \frac{2}{3} & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

  From this, the reduced system is:

$$\begin{cases} x_1 + \frac{2}{3}x_2 + x_3 = 0 \\ x_2 = 0 \end{cases} \Rightarrow x_2 = 0, \quad x_1 + x_3 = 0 \Rightarrow x_1 = -x_3$$

  Let $x_3 = t$, then:

$$x_1 = -t, \quad x_2 = 0, \quad x_3 = t \Rightarrow \vec{v}_{-2} = t \cdot \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} \Rightarrow \boxed{\vec{v}_{-2} = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}}$$

Conclusion:

Since the matrix A has three distinct real eigenvalues, it admits three linearly independent eigenvectors. These eigenvectors span $\mathbb{R}^3$ and form a basis consisting entirely of eigenvectors of A. Therefore, A is diagonalizable.

The diagonalization of A can be expressed as:

$$A = PDP^{-1}$$

where D is a diagonal matrix whose diagonal entries are the eigenvalues of A, and P is an invertible matrix whose columns are the corresponding eigenvectors.

(c) **Definiteness analysis of the quadratic form $\vec{x}^T A\vec{x}$**

To determine whether the matrix A is positive definite, negative definite, or indefinite, the associated quadratic form is analyzed:

$$q(\vec{x}) = \vec{x}^T A\vec{x}$$

The classification of definiteness based on the sign of $q(\vec{x})$ is as follows:

- **Positive definite** if $\vec{x}^T A\vec{x} > 0$ for all nonzero $\vec{x} \in \mathbb{R}^n$,
- **Negative definite** if $\vec{x}^T A\vec{x} < 0$ for all nonzero $\vec{x}$,
- **Indefinite** if $\vec{x}^T A\vec{x}$ takes both positive and negative values.

The relationship between eigenvalues and definiteness was already discussed in detail in Question 1(b). To summarize:

- If all eigenvalues are positive, the matrix is positive definite.
- If all eigenvalues are negative, the matrix is negative definite.
- If the eigenvalues include both positive and negative values, the matrix is indefinite.

As previously computed in part (b), the eigenvalues of A are:

$$\lambda_1 = 6, \quad \lambda_2 = 0, \quad \lambda_3 = -2$$

Since the matrix has both positive and negative eigenvalues, and one of them is zero, the quadratic form takes both positive and negative values depending on the direction of $\vec{x}$. Therefore:

$$\boxed{\text{The matrix } A \text{ is indefinite.}}$$

Question 3: Convexity

(a) **Analysis of the convexity properties of the function** $f(x) = |x|$

The function $f(x) = |x|$ is piecewise defined as:

$$f(x) = \begin{cases} x, & x \geq 0 \\ -x, & x < 0 \end{cases}$$

This function is convex, which can be shown using the definition of convexity: A function $f : \mathbb{R} \to \mathbb{R}$ is convex if for all $x_1, x_2 \in \mathbb{R}$ and $\lambda \in [0, 1]$, the following inequality holds:

$$f(\lambda x_1 + (1 - \lambda)x_2) \leq \lambda f(x_1) + (1 - \lambda)f(x_2)$$

The absolute value function satisfies this condition for all real values. Geometrically, this corresponds to the graph of f lying below any chord connecting two points on the graph — a hallmark of convexity.

To formally verify convexity, the definition can be directly tested for $f(x) = |x|$ as follows. For any $x_1, x_2 \in \mathbb{R}$ and $\lambda \in [0, 1]$, now:

$$\begin{aligned} f(\lambda x_1 + (1 - \lambda)x_2) &= |\lambda x_1 + (1 - \lambda)x_2| \\ &\leq \lambda|x_1| + (1 - \lambda)|x_2| = \lambda f(x_1) + (1 - \lambda)f(x_2) \end{aligned}$$

The inequality step follows from the triangle inequality:

$$|a+b| \leq |a| + |b| \quad \text{for all } a, b \in \mathbb{R}$$

Here, $a = \lambda x_1,\ b = (1-\lambda)x_2$, so convexity is satisfied for all real numbers. Therefore, $f(x) = |x|$ is convex.

However, $f(x) = |x|$ is **not strictly convex**. A function is strictly convex if the inequality above is strict whenever $x_1 \neq x_2$. This fails for $f(x) = |x|$ in linear regions such as $x_1, x_2 \geq 0$ or $x_1, x_2 \leq 0$, where the function behaves linearly:

$$f(\lambda x_1 + (1-\lambda)x_2) = \lambda f(x_1) + (1-\lambda) f(x_2)$$

Equality holds exactly, so strict convexity does not apply on those intervals. Therefore:

$$\boxed{f(x) = |x| \text{ is convex but not strictly convex.}}$$

(b) **Convexity analysis of the binary cross-entropy loss function**

Consider the binary cross-entropy loss:

$$\mathcal{L}(y, \hat{y}) = -\left[y \log \hat{y} + (1-y)\log(1-\hat{y})\right]$$

where $y \in \{0, 1\}$ is the true label and $\hat{y} \in (0, 1)$ is the predicted probability.

To analyze convexity, one can treat y as fixed (a constant label) and consider $\mathcal{L}$ as a function of $\hat{y} \in (0, 1)$. This is because $y \in \{0, 1\}$ represents the ground truth label, which is not a variable in the optimization process. Convexity is assessed with respect to the prediction $\hat{y}$, so y acts as a constant during this analysis.

$$\begin{aligned}
&\text{If } y = 1, \quad \text{the loss becomes } \mathcal{L}(\hat{y}) = -\log(\hat{y}) \\
&\text{If } y = 0, \quad \text{the loss becomes } \mathcal{L}(\hat{y}) = -\log(1-\hat{y})
\end{aligned}$$

Since both expressions are univariate functions defined on the interval (0,1), their convexity can be determined by inspecting the sign of the second derivative. A positive second derivative on the entire interval confirms that the function is convex (and strictly convex if the derivative is strictly positive).

- For $y = 1$:

$$\mathcal{L}''(\hat{y}) = \frac{1}{\hat{y}^2} > 0 \quad \text{for } \hat{y} \in (0, 1)$$

- For $y = 0$:

$$\mathcal{L}''(\hat{y}) = \frac{1}{(1-\hat{y})^2} > 0 \quad \text{for } \hat{y} \in (0, 1)$$

Hence, $\mathcal{L}(y, \hat{y})$ is **strictly convex** in $\hat{y} \in (0, 1)$ for any fixed $y \in \{0, 1\}$.

$$\boxed{\mathcal{L}(y, \hat{y}) \text{ is strictly convex in } \hat{y} \in (0, 1)}$$

Question 4: Quadratic Fit Line Search Implementation and Analysis

Consider the quadratic fit line search method discussed in class. The goal is to apply it to the following function:

$$f(x) = \left(x - \frac{b_0}{2}\right)\left(x - \frac{b_1}{2}\right)\left(x - \frac{b_2}{2}\right)\left(x - \frac{b_3}{2}\right)$$

Here, b_0, b_1, b_2, b_3 are the last four digits of the student number. If any digit is zero, it should be replaced by 1. For student number `150220313`, the values become:

$$b_0 = 1, \quad b_1 = 3, \quad b_2 = 1, \quad b_3 = 3$$

Thus, the function becomes:

$$f(x) = \left(x - \frac{3}{2}\right)^2 \left(x - \frac{1}{2}\right)^2$$

(a) **2D Visualization and Unimodal Interval Identification**

The function is defined as:

$$f(x) = \left(x - \frac{1}{2}\right)^2 \left(x - \frac{3}{2}\right)^2$$

A 2D plot of the function is generated below:

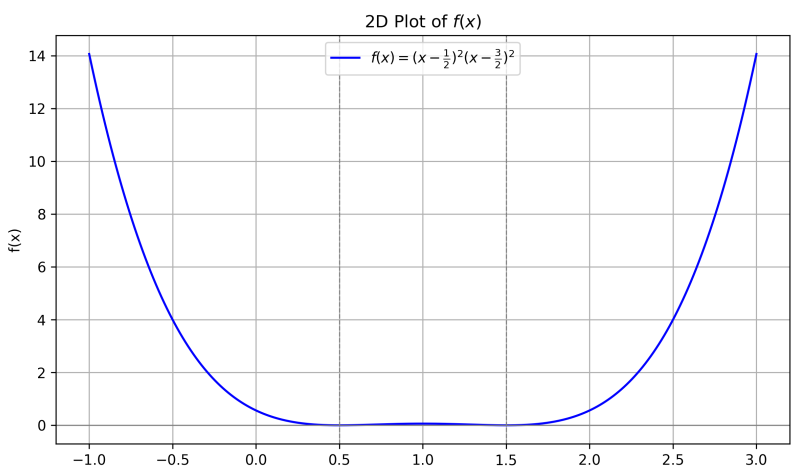

Figure 4: Plot of $f(x) = (x - 0.5)^2(x - 1.5)^2$

From the graph, one may observe that the function is non-negative and symmetric, with a single global minimum between $x = 0.5$ and $x = 1.5$. In this interval, the function decreases, reaches a minimum, and increases again, without any other local extrema.

Therefore, the function is **unimodal** on the interval:

$$\boxed{[0.5,\ 1.5]}$$

(b) **Quadratic Fit Search Iterations and Visualization**

The quadratic fit line search method was applied to the function

$$f(x) = \left(x - \frac{1}{2}\right)^2 \left(x - \frac{3}{2}\right)^2$$

on the interval $[0.5,\ 1.5]$, which is known to be unimodal.

Three initial points were selected:

$$x_0 = 0.55, \quad x_1 = 1.2, \quad x_2 = 1.45$$

and the method was run for three iterations, resulting in three new approximation points, totaling six evaluated points in the process.

The following figures display the function curve along with the evaluated points connected in order of appearance. Labels indicate the sequence of iterations:

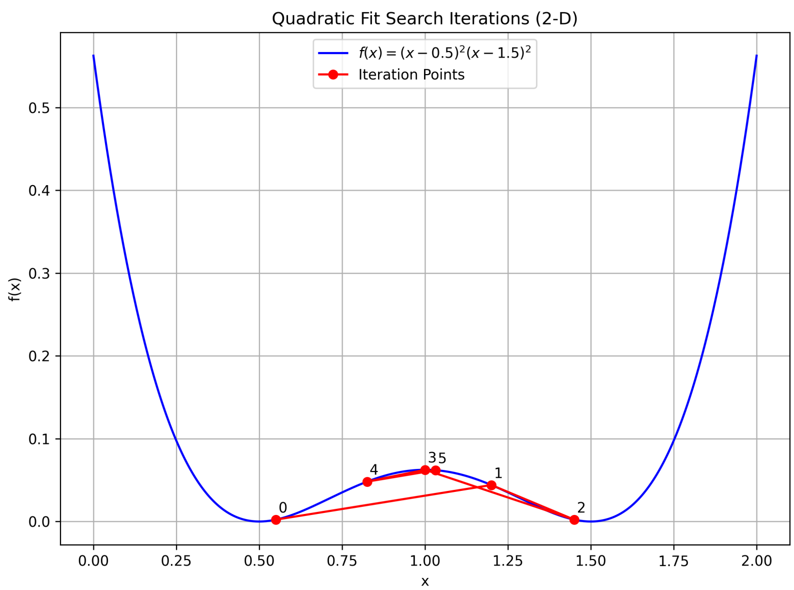

Figure 5: Quadratic Fit Line Search Iteration Points on the Function $f(x)$

The graph shows that the method successfully refines its estimate of the minimum by progressively narrowing the search area. The choice of initial points ensures numerical stability in the polynomial fit and avoids symmetric degeneracies.

In addition to the 2D function plot, a 3D function plot visualizes the iteration process in terms of search step progression.

In this graph:

- The x-axis shows the evaluated input values x,
- The y-axis shows the corresponding function values $f(x)$,
- The z-axis corresponds to the iteration step (i.e., the order in which the points were evaluated).

This representation helps to visualize how the quadratic fit method converges step by step along the surface of the function. The movement in the graph illustrates the method's refinement path across iterations.

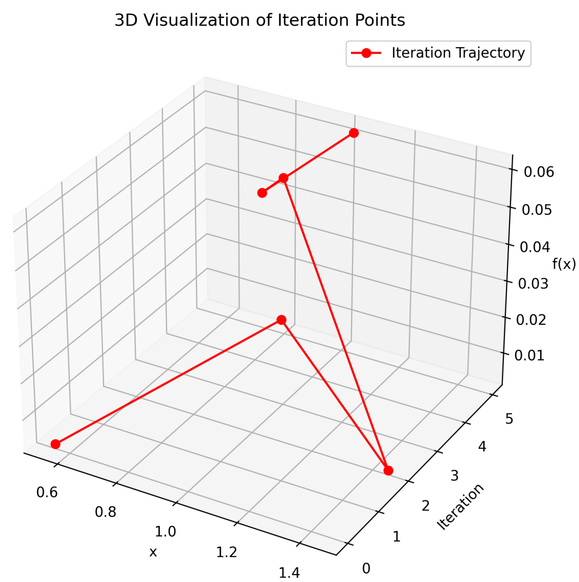

Figure 6: 3D Visualization of Quadratic Fit Search Iterations

(c) **Failure Behavior of the Quadratic Fit Search Method**

In general, the quadratic fit line search method can fail under certain conditions. These failures typically arise when the three sampled points used to fit the quadratic are poorly chosen. Some common reasons include:

- The three points are nearly collinear, resulting in a poorly conditioned system for fitting a quadratic curve.
- The function is flat or nearly flat in the selected region, making it difficult to determine a unique minimum.
- The selected points are symmetrically placed around a flat region or the global minimum, causing the fitted parabola to have an inaccurate curvature or degenerate shape.
- Numerical instability due to round-off errors when the function values are too close or the x-values are too clustered.

To test for failure in our specific function:

$$f(x) = \left(x - \frac{1}{2}\right)^2 \left(x - \frac{3}{2}\right)^2$$

The quadratic fit method was applied on the interval $[0.5,\ 1.5]$, which is unimodal and symmetric around its minimum at $x = 1$.

Three initial points were selected:

$$x_0 = 0.55, \quad x_1 = 1.2, \quad x_2 = 1.45$$

and the method was executed for three iterations. The result showed stable convergence toward the minimum without any sign of instability or failure.

Conclusion: No failure was observed in this test case. The primary reason for success is that the function is smooth, strictly convex in the chosen interval, and the selected points are sufficiently spaced and asymmetrically placed. This avoids degenerate fits and ensures that the quadratic approximation is both accurate and well-conditioned.

Hence, for this specific problem setup, the quadratic fit search performs reliably and does not exhibit failure.

Question 5: Steepest Descent

Given function:

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

Steepest Descent Method with Fixed Step Size: $\alpha = 0.05$ **Initial Point:** $(-1, 1)$

(a) Manual Computation of First Two Iterations

The goal is to minimize a differentiable function using the steepest descent method. This iterative technique proceeds by moving from the current point in the direction of the negative gradient of the function, which represents the direction of steepest decrease.

The gradient vector provides the direction of maximum increase of the function. Consequently, taking a step in the opposite direction — the negative gradient — yields the most rapid local decrease. This strategy ensures that each update in the iterative scheme brings the point closer to a local minimum, provided that the step size is appropriately chosen.

The steepest descent algorithm updates the current point by moving in the direction of the negative gradient, scaled by a fixed step size α. The general update rule is given by:

$$\vec{x}_{k+1} = \vec{x}_k - \alpha \nabla f(\vec{x}_k)$$

Here:

- $\vec{x}_k$ denotes the current point at iteration k,
- $\alpha > 0$ is the fixed step size,
- $\nabla f(\vec{x}_k)$ is the gradient of the function evaluated at $\vec{x}_k$.

This formula ensures that the algorithm takes a step in the direction of steepest descent, with the step size determining how far to move in that direction. In this question, the step size is fixed at $\alpha = 0.05$ for all iterations.

For the given function:

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

the gradient is computed as:

$$\nabla f(x_1, x_2) = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \end{bmatrix} = \begin{bmatrix} 4x_1^3 - 4x_1 x_2 + 2x_1 \\ -2x_1^2 + 4x_2^3 - 2x_2 \end{bmatrix}$$

Iteration 0:

At $\vec{x}^{(0)} = (-1, 1)$:

$$\nabla f(-1,1) = \begin{bmatrix} 4(-1)^3 - 4(-1)(1) + 2(-1) \\ -2(-1)^2 + 4(1)^3 - 2(1) \end{bmatrix} = \begin{bmatrix} -4 + 4 - 2 \\ -2 + 4 - 2 \end{bmatrix} = \begin{bmatrix} -2 \\ 0 \end{bmatrix}$$

Descent direction:

$$\vec{d}^{(0)} = -\nabla f^{(0)} = \begin{bmatrix} 2 \\ 0 \end{bmatrix}$$

Update:

$$\vec{x}^{(1)} = \vec{x}^{(0)} + \alpha \cdot \vec{d}^{(0)} = \begin{bmatrix} -1 \\ 1 \end{bmatrix} + 0.05 \cdot \begin{bmatrix} 2 \\ 0 \end{bmatrix} = \begin{bmatrix} -0.9 \\ 1 \end{bmatrix}$$

Iteration 1:

At $\vec{x}^{(1)} = (-0.9, 1)$, compute:

$$x_1^2 = 0.81, \quad x_1^3 = -0.729$$

$$\nabla f(-0.9,1) = \begin{bmatrix} 4(-0.729) - 4(-0.9)(1) + 2(-0.9) \\ -2(0.81) + 4(1)^3 - 2(1) \end{bmatrix} = \begin{bmatrix} -2.916 + 3.6 - 1.8 \\ -1.62 + 4 - 2 \end{bmatrix} \approx \begin{bmatrix} -1.116 \\ 0.38 \end{bmatrix}$$

Descent direction:

$$\vec{d}^{(1)} = -\nabla f^{(1)} = \begin{bmatrix} 1.116 \\ -0.38 \end{bmatrix}$$

Update:

$$\vec{x}^{(2)} = \vec{x}^{(1)} + \alpha \cdot \vec{d}^{(1)} \approx \begin{bmatrix} -0.9 \\ 1 \end{bmatrix} + 0.05 \cdot \begin{bmatrix} 1.116 \\ -0.38 \end{bmatrix} \approx \begin{bmatrix} -0.8442 \\ 0.981 \end{bmatrix}$$

Summary of Iterations

$$\begin{aligned} \vec{x}^{(0)} &= (-1,\ 1) \\ \vec{x}^{(1)} &= (-0.9,\ 1) \\ \vec{x}^{(2)} &\approx (-0.8442,\ 0.981) \end{aligned}$$

The steepest descent method successfully reduces the function value by moving in the direction of the negative gradient. The steps are consistent with the function's local curvature.

(b) Manual Descent from Two Different Initial Points

To evaluate the impact of the initial point on convergence behavior, two additional initial points are selected:

$$\vec{x}_A^{(0)} = (1,\ -1), \quad \vec{x}_B^{(0)} = (1,\ 1)$$

The same descent rule is used as in part (a), with fixed step size $\alpha = 0.05$. The gradient and update steps are computed manually for two iteration for each point.

Initial Point A: $\vec{x}^{(0)} = (1, -1)$

Iteration 0:

$$\nabla f(1, -1) = \begin{bmatrix} 4(1)^3 - 4(1)(-1) + 2(1) \\ -2(1)^2 + 4(-1)^3 - 2(-1) \end{bmatrix} = \begin{bmatrix} 4 + 4 + 2 \\ -2 - 4 + 2 \end{bmatrix} = \begin{bmatrix} 10 \\ -4 \end{bmatrix}$$

Descent direction:

$$\vec{d}^{(0)} = -\nabla f(1, -1) = \begin{bmatrix} -10 \\ 4 \end{bmatrix}$$

Update:

$$\vec{x}^{(1)} = \vec{x}^{(0)} + \alpha \cdot \vec{d}^{(0)} = \begin{bmatrix} 1 \\ -1 \end{bmatrix} + 0.05 \cdot \begin{bmatrix} -10 \\ 4 \end{bmatrix} = \begin{bmatrix} 0.5 \\ -0.8 \end{bmatrix}$$

Iteration 1:

Compute:

$$x_1^2 = 0.25, \quad x_1^3 = 0.125$$

$$\nabla f(0.5, -0.8) = \begin{bmatrix} 4(0.5)^3 - 4(0.5)(-0.8) + 2(0.5) \\ -2(0.5)^2 + 4(-0.8)^3 - 2(-0.8) \end{bmatrix} = \begin{bmatrix} 0.5 + 1.6 + 1 \\ -0.5 - 2.048 + 1.6 \end{bmatrix} \approx \begin{bmatrix} 3.1 \\ -0.948 \end{bmatrix}$$

Descent direction:

$$\vec{d}^{(1)} = -\nabla f(0.5, -0.8) = \begin{bmatrix} -3.1 \\ 0.948 \end{bmatrix}$$

Update:

$$\vec{x}^{(2)} = \vec{x}^{(1)} + \alpha \cdot \vec{d}^{(1)} \approx \begin{bmatrix} 0.5 \\ -0.8 \end{bmatrix} + 0.05 \cdot \begin{bmatrix} -3.1 \\ 0.948 \end{bmatrix} \approx \begin{bmatrix} 0.345 \\ -0.7526 \end{bmatrix}$$

Initial Point B: $\vec{x}^{(0)} = (1, 1)$

Iteration 0:

$$\nabla f(1, 1) = \begin{bmatrix} 4(1)^3 - 4(1)(1) + 2(1) \\ -2(1)^2 + 4(1)^3 - 2(1) \end{bmatrix} = \begin{bmatrix} 4 - 4 + 2 \\ -2 + 4 - 2 \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \end{bmatrix}$$

Descent direction:

$$\vec{d}^{(0)} = -\nabla f(1, 1) = \begin{bmatrix} -2 \\ 0 \end{bmatrix}$$

Update:

$$\vec{x}^{(1)} = \vec{x}^{(0)} + \alpha \cdot \vec{d}^{(0)} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} + 0.05 \cdot \begin{bmatrix} -2 \\ 0 \end{bmatrix} = \begin{bmatrix} 0.9 \\ 1 \end{bmatrix}$$

Iteration 1:

$$x_1^2 = 0.81, \quad x_1^3 = 0.729$$

$$\nabla f(0.9, 1) = \begin{bmatrix} 4(0.729) - 4(0.9)(1) + 2(0.9) \\ -2(0.81) + 4(1)^3 - 2(1) \end{bmatrix} \approx \begin{bmatrix} 2.916 - 3.6 + 1.8 \\ -1.62 + 4 - 2 \end{bmatrix} \approx \begin{bmatrix} 1.116 \\ 0.38 \end{bmatrix}$$

Descent direction:

$$\vec{d}^{(1)} = -\nabla f(0.9, 1) = \begin{bmatrix} -1.116 \\ -0.38 \end{bmatrix}$$

Update:

$$\vec{x}^{(2)} = \vec{x}^{(1)} + \alpha \cdot \vec{d}^{(1)} \approx \begin{bmatrix} 0.9 \\ 1 \end{bmatrix} + 0.05 \cdot \begin{bmatrix} -1.116 \\ -0.38 \end{bmatrix} \approx \begin{bmatrix} 0.8442 \\ 0.981 \end{bmatrix}$$

Initial Point	**First Iteration $\vec{x}^{(1)}$**	**Second Iteration $\vec{x}^{(2)}$**
$(-1,\ 1)$	$(-0.9,\ 1)$	$(-0.8442,\ 0.981)$
$(1,\ -1)$	$(0.5,\ -0.8)$	$(0.3475,\ -0.7102)$
$(1,\ 1)$	$(0.9,\ 1)$	$(0.8442,\ 0.981)$

Table 1: Comparison of Two Iterations from Different Initial Points

Discussion: The initial point $(1, -1)$ exhibits a significantly steeper gradient norm compared to the other two points. This results in a larger update in the first descent direction, as observed in the table. The rapid initial movement implies that the method may reach regions of lower function value more quickly, potentially accelerating convergence.

In contrast, the initial points $(-1, 1)$ and $(1, 1)$ yield smaller gradient norms. Consequently, the descent directions are less aggressive, leading to smaller updates and slower initial progress. Despite having similar function structure, these symmetric points behave identically in the descent process due to the symmetry in gradient expressions.

This comparison illustrates an important property of the steepest descent method: its strong dependence on the local gradient magnitude and curvature of the objective function. A large gradient norm can provide faster descent if the direction aligns well with the curvature of the function. However, it can also lead to overshooting or oscillation in poorly conditioned regions.

Therefore, the effectiveness of steepest descent is not solely determined by gradient magnitude but also by the local geometry. Well-conditioned regions with smooth curvature support stable updates, whereas regions with steep or anisotropic contours may require adaptive step sizes or preconditioning for reliable convergence.

Question 6: Steepest Descent – Programming

(a) Backtracking Line Search Implementation

In this part, the steepest descent method is implemented using the backtracking line search algorithm that satisfies the sufficient decrease (Armijo) condition. The function selected for optimization is the Rosenbrock function, defined as:

$$f(x_1, x_2) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2$$

The update rule at each iteration is given by:

$$\vec{x}^{(k+1)} = \vec{x}^{(k)} - \alpha_k \nabla f(\vec{x}^{(k)})$$

where α_k is selected via backtracking line search based on the Armijo condition.

The implementation proceeds until the norm of the gradient drops below a small threshold (e.g., $\|\nabla f\| < 10^{-4}$) or a maximum number of iterations is reached.

(b) 2D and 3D Visualizations

The figures below show the optimization trajectory on both 2D contour and 3D surface plots. Each step of the descent is connected by arrows and shown with a color indicating the iteration progress.

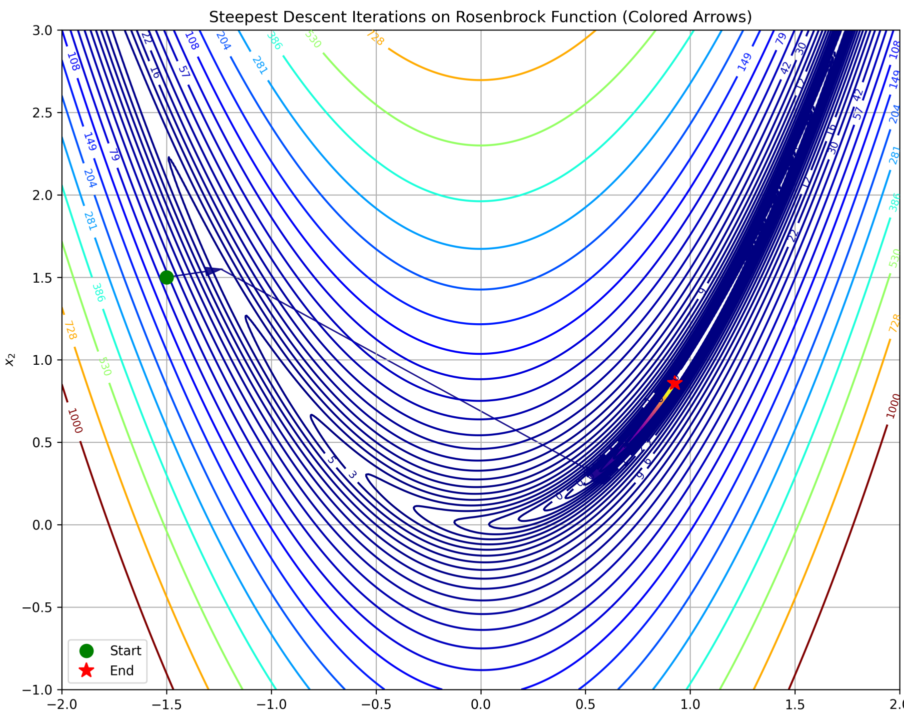

Figure 7: 2D Contour Plot of Rosenbrock Function with Steepest Descent Trajectory

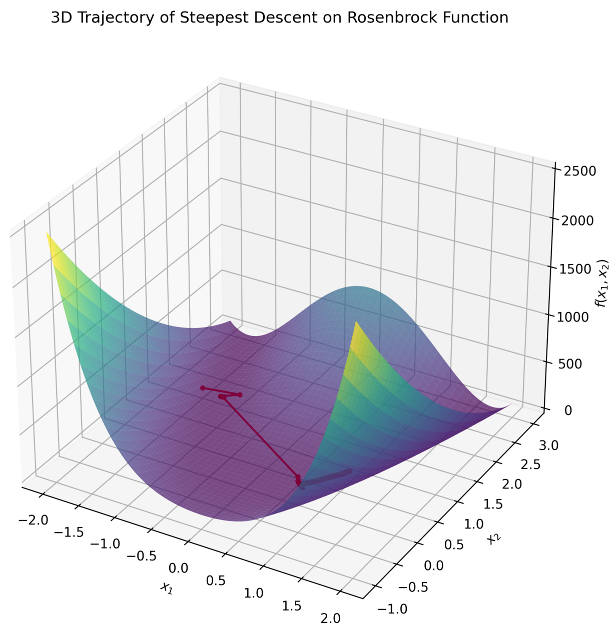

Figure 8: 3D Surface Plot of Rosenbrock Function with Steepest Descent Trajectory

Comment: The visualizations reveal the challenges of optimizing the Rosenbrock function, particularly due to its narrow, curved valley that forms a classic banana-shaped landscape. The initial steps of the algorithm are relatively large and direct, reflecting the steep gradient encountered at the edges of the domain. As the algorithm progresses closer to the valley, the curvature increases dramatically, causing the gradient directions to oscillate and the step sizes to shrink due to backtracking.

This behavior demonstrates a well-known limitation of steepest descent: poor convergence in regions with ill-conditioning or high anisotropy. The 2D contour plot clearly shows the zigzag pattern of the trajectory, while the 3D surface highlights the dramatic height drop toward the minimum. These patterns underscore the importance of line search strategies and potentially second-order information when optimizing functions like Rosenbrock. Such visual tools are invaluable for diagnosing optimization behavior and guiding improvements to the algorithm.

(c) Manually Calculated Points Overlay

The manually computed steepest descent points from Question 5 for each of the three initial points are:

$$\begin{aligned}
&\textbf{From } (-1,1): \quad \vec{x}^{(0)} = (-1,1), \quad \vec{x}^{(1)} = (-0.9,1), \quad \vec{x}^{(2)} \approx (-0.8442, 0.981) \\
&\textbf{From } (1,-1): \quad \vec{x}^{(0)} = (1,-1), \quad \vec{x}^{(1)} = (0.5,-0.8), \quad \vec{x}^{(2)} \approx (0.3475, -0.7102) \\
&\textbf{From } (1,1): \quad \vec{x}^{(0)} = (1,1), \quad \vec{x}^{(1)} = (0.6,0.84), \quad \vec{x}^{(2)} \approx (0.4896, 0.7235)
\end{aligned}$$

These manually calculated paths have been overlaid on both the 2D contour and 3D surface plots using distinct colors for each trajectory to differentiate the descent behavior from different starting points.

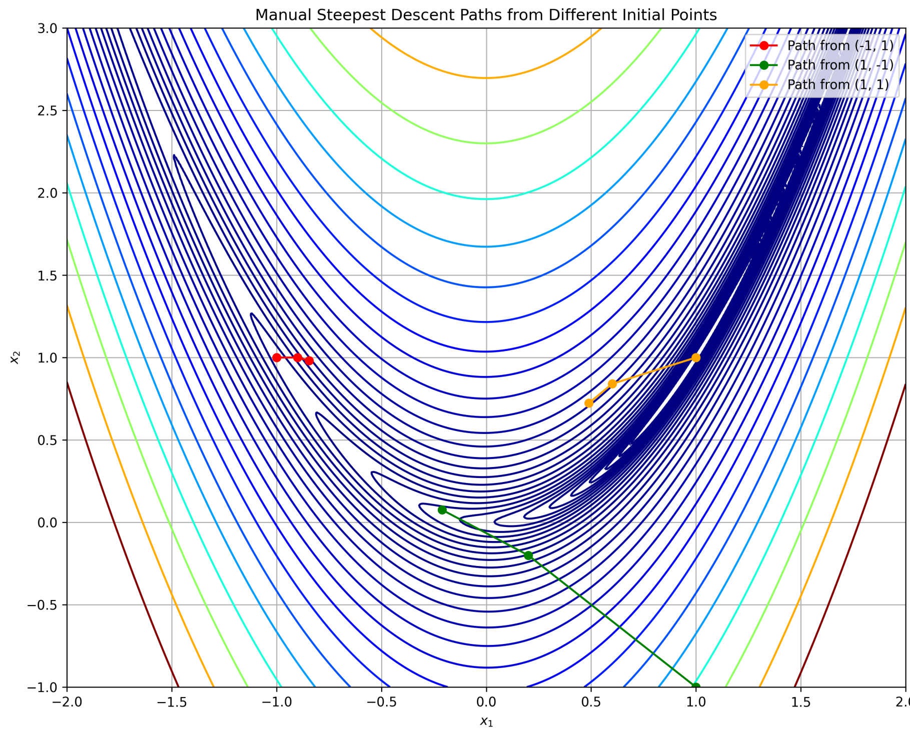

Figure 9: 2D Contour Plot Showing Manually Computed Steepest Descent Paths from Multiple Initial Points

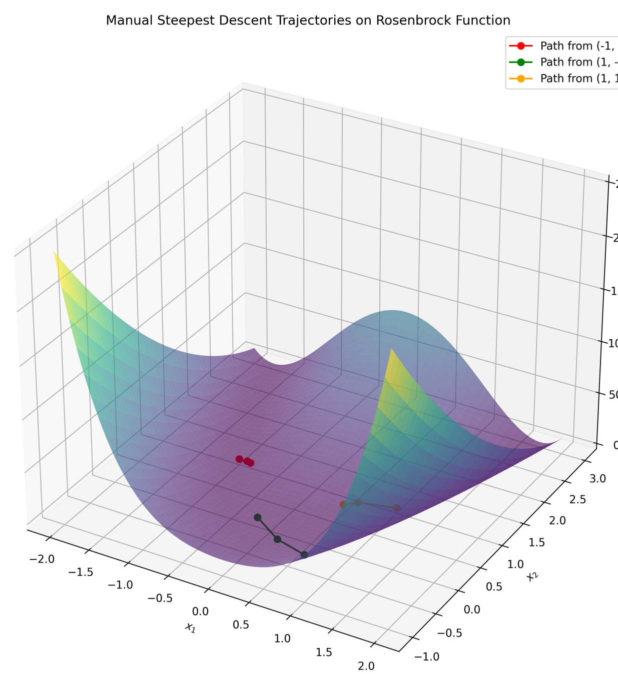

Figure 10: 3D Surface Plot Showing Manually Computed Steepest Descent Trajectories on Rosenbrock Function

Comment: The manually calculated descent paths exhibit distinct behaviors depending on their starting positions.

- The path starting from $(1, -1)$ exhibits a sharp gradient initially and moves quickly toward the curved valley, indicating faster convergence in early iterations.
- The path starting from $(-1, 1)$ moves more slowly, as it begins in a region with less favorable curvature.
- The path from $(1, 1)$ also progresses gradually, following the bowl shape of the Rosenbrock surface.

These results confirm that steepest descent is highly sensitive to both the initial location and the local geometry of the objective function. The visual alignment between the manual updates and the expected curvature of the Rosenbrock function supports the correctness of the analytical calculations.