Nonlinear Optimization - Assignment 1

Sam Kische, Anja Misimovic, Luka Bajić

November 2025

1 Characterization of Functions

a) $f(\mathbf{x}) = (\mathbf{a}^T\mathbf{x} - d)^2$ where $\mathbf{a} = (-1, 3)^T, d = 2.5$

Because $\mathbf{a}$ and d are constant, we can rewrite f as $f(\mathbf{x}) = (-x_1 + 3x_2 - 2.5)^2$, for which we then compute partial derivatives with respect to x_1 and x_2:

$$\frac{\partial f(\mathbf{x})}{\partial x_1} = -2(-x_1 + 3x_2 - 2.5)$$

$$\frac{\partial f(\mathbf{x})}{\partial x_2} = 6(-x_1 + 3x_2 - 2.5)$$

Simplifying the brackets, we get the gradient:

$$\nabla f(\mathbf{x}) = \begin{bmatrix} 2x_1 - 6x_2 + 5 \\ -6x_1 + 18x_2 - 15 \end{bmatrix}$$

Next we compute partial derivatives of the gradient with respect to x_1 and x_2:

$$\frac{\partial \nabla f(\mathbf{x})}{\partial x_1} = \begin{bmatrix} 2 \\ -6 \end{bmatrix}$$

$$\frac{\partial \nabla f(\mathbf{x})}{\partial x_2} = \begin{bmatrix} -6 \\ 18 \end{bmatrix}$$

which gives us the Hessian matrix:

$$\nabla^2 f(\mathbf{x}) = \begin{bmatrix} -2 & -6 \\ -6 & 18 \end{bmatrix}$$

To find the eigenvalues of the Hessian matrix, we need to solve $\det(\nabla^2 f(\mathbf{x}) - \lambda I_2) = 0$:

$$\det(\nabla^2 f(\mathbf{x}) - \lambda I_2) = \det(\begin{bmatrix} 2-\lambda & -6 \\ -6 & 18-\lambda \end{bmatrix}) = (2-\lambda)(18-\lambda) - ((-6)*(-6)) =$$

$$= 36 - 20\lambda + \lambda^2 - 36 = \lambda^2 - 20\lambda = \lambda(\lambda - 20)$$

Setting $\lambda(\lambda - 20)$ to 0, we get $\lambda_1 = 20$ and $\lambda_2 = 0$, which implies that the Hessian matrix is positive semi-definite, because $\lambda_i \geq 0 \quad \forall i \in 1, \ldots, n$, where n is the dimension of the gradient.

To find the stationary points we have to equate the gradient with a zero vector, i.e. $\nabla f(\mathbf{x}) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$, which gives us the following system of equations:

$$2x_1 - 6x_2 + 5 = 0$$

$$-6x_1 + 18x_2 - 15 = 0$$

We immediately see that if we multiply the first equation with 3 and equate the two equations, we get $6x_1 - 18x_2 + 15 = -6x_1 + 18x_2 - 15$, where both x_1 and x_2 terms cancel out, which implies that the equality holds for any $x_1, x_2 \in \mathbb{R}$. In terms of optimization this means that the entire hyperplane defined by $2x_1 - 6x_2 + 5 = 0$ is a valid solution. And due to the established positive-semidefiniteness of the Hessian matrix, the entire hyperplane is a strict local minimum.

b) $f(\mathbf{x}) = (x_1 - 2)^2 + x_1 x_2^2 - 2$

Computing the partial derivatives of f with respect to x_1 and x_2 we get:

$$\frac{\partial f(\mathbf{x})}{\partial x_1} = 2(x_1 - 2) + x_2^2 = 2x_1 + x_2^2 - 4$$

$$\frac{\partial f(\mathbf{x})}{\partial x_2} = 2x_1 x_2$$

which gives us the gradient:

$$\nabla f(\mathbf{x}) = \begin{bmatrix} 2x_1 + x_2^2 - 4 \\ 2x_1 x_2 \end{bmatrix}$$

Next we compute the partial derivatives of the gradient with respect to x_1 and x_2:

$$\frac{\partial \nabla f(\mathbf{x})}{\partial x_1} = \begin{bmatrix} 2 \\ 2x_2 \end{bmatrix}$$

$$\frac{\partial \nabla f(\mathbf{x})}{\partial x_2} = \begin{bmatrix} 2x_2 \\ 2x_1 \end{bmatrix}$$

which gives us the Hessian matrix:

$$\nabla^2 f(\mathbf{x}) = \begin{bmatrix} 2 & 2x_2 \\ 2x_2 & 2x_1 \end{bmatrix}$$

To find stationary points we have to equate the gradient with the zero vector, i.e. $\nabla f(x) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$, which gives us a system of equations:

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

$$2x_1 x_2 = 0$$

Rearranging the first equation we get $x_2 = \sqrt{-2x_1 + 4}$, which we plug into the second equation to get $2x_1\sqrt{-2x_1 + 4} = 0$. Squaring both sides of the equations and then multiplying with $-\frac{1}{8}$, we get:

$$4x_1^2(-2x_1 + 4) = 0$$

$$-8x_1^3 + 16x_1^2 = 0$$

$$x_1^3 - 2x_1^2 = 0$$

$$x_1^2(x_1 - 2) = 0$$

from which we get three candidate x-values: $x_1, 1 = x_1, 2 = 0$ and $x_1, 3 = 2$. Plugging these three values into the first equation we get:

$$2 \cdot 0 + x_2^2 - 4 = 0$$

$$x^2 = \sqrt{4}$$

which gives us $x_{2,1} = 2$, $x_{2,2} = -2$. And

$$2 \cdot 2 + x_2^2 - 4 = 0$$

$$x_2^2 = 0$$

which gives us $x_{2,3} = 0$. Therefore we have obtained three candidate points for local extrema: $\begin{bmatrix} 0 \\ 2 \end{bmatrix}$, $\begin{bmatrix} 0 \\ -2 \end{bmatrix}$ and $\begin{bmatrix} 2 \\ 0 \end{bmatrix}$.

To check whether these points are minima, maxima or saddle points, we plug each of them into the Hessian matrix and determine the definiteness of the obtained matrix:

- $\mathbf{x} = \begin{bmatrix} 2 \\ 0 \end{bmatrix}$:

$$\nabla^2 f(\begin{bmatrix} 2 \\ 0 \end{bmatrix}) = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$

  The matrix is diagonal, so the eigenvalues are $\lambda_1 = 2 > 0$ and $\lambda_2 = 2 > 0$, therefore the matrix is positive definite, and the point $\begin{bmatrix} 2 \\ 0 \end{bmatrix}$ is a strict local minimum.

- $\mathbf{x} = \begin{bmatrix} 0 \\ 2 \end{bmatrix}$:

$$\nabla^2 f(\begin{bmatrix} 0 \\ 2 \end{bmatrix}) = \begin{bmatrix} 2 & 4 \\ 4 & 0 \end{bmatrix}$$

  TODO

- $\mathbf{x} = \begin{bmatrix} 0 \\ -2 \end{bmatrix}$:

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

  TODO

c) $f(\mathbf{x}) = x_1^2 + x_1\|\mathbf{x}\|^2 + \|\mathbf{x}\|^2$

Using the property $\|\mathbf{x}\|^2 = (\sqrt{x_1^2 + x_2^2})^2 = x_1^2 + x_2^2$, we simplify the function f as follows:

$$f(\mathbf{x}) = x_1^2 + x_1(x_1^2 + x_2^2) + x_1^2 + x_2^2 = x_1^3 + 2x_1^2 + x_2^2 + x_1x_2^2$$

Next we compute the partial derivatives of f with respect to x_1 and x_2:

$$\frac{\partial f(\mathbf{x})}{\partial x_1} = 3x_1^2 + 4x_1 + x_2^2$$

$$\frac{\partial f(\mathbf{x})}{\partial x_2} = 2x_2 + 2x_1x_2$$

which gives us the gradient:

$$\nabla f(\mathbf{x}) = \begin{bmatrix} 3x_1^2 + 4x_1 + x_2^2 \\ 2x_2 + 2x_1x_2 \end{bmatrix}$$

Computing the partial derivatives of the gradient with respect to x_1 and x_2 we get:

$$\frac{\partial \nabla f(\mathbf{x})}{\partial x_1} = \begin{bmatrix} 6x_1 + 4 \\ 2x_2 \end{bmatrix}$$

$$\frac{\partial \nabla f(\mathbf{x})}{\partial x_2} = \begin{bmatrix} 2x_2 \\ 2x_1 + 2 \end{bmatrix}$$

which gives us the Hessian matrix:

$$\nabla^2 f(x) = \begin{bmatrix} 6x_1 + 4 & 2x_2 \\ 2x_2 & 2x_1 + 2 \end{bmatrix}$$

To get the stationary points, we need $\nabla f(\mathbf{x}) = \begin{bmatrix} 3x_1^2 + 4x_1 + x_2^2 \\ 2x_2 + 2x_1x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$, which gives us the following system of equations:

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

$$2x_2 + 2x_1x_2 = 0$$

From the second equation we get:

$$x_2(2x_1 + 2) = 0$$

which gives us $x_2 = 0$ and $2x_1 + 2 = 0 \rightarrow x_1 = -1$. Plugging these two values into the first equation we get:

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

$$x_1(3x_1 + 4) = 0$$

therefore $x_{1,1} = 0$ and $x_{1,2} = -\frac{4}{3}$. And

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

$$x_2^2 = 1$$

therefore $x_{2,1} = -1$ and $x_{2,2} = 1$. In total we have thus obtained four candidate points: $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$, $\begin{bmatrix} -\frac{4}{3} \\ 0 \end{bmatrix}$, $\begin{bmatrix} -1 \\ -1 \end{bmatrix}$ and $\begin{bmatrix} -1 \\ 1 \end{bmatrix}$.

Next we need to determine the definiteness of the Hessian matrix in these points:

- $\mathbf{x} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$:

  $$\nabla^2 f(\begin{bmatrix} 0 \\ 0 \end{bmatrix}) = \begin{bmatrix} 4 & 0 \\ 0 & 2 \end{bmatrix}$$

  Because the matrix is diagonal, the eigenvalues are $\lambda_1 = 4 > 0$ and $\lambda_2 = 2 > 0$, therefore the matrix is positive definite, and the point $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$ is a strict local minimum.

- $\mathbf{x} = \begin{bmatrix} -\frac{4}{3} \\ 0 \end{bmatrix}$:

  $$\nabla^2 f(\begin{bmatrix} -\frac{4}{3} \\ 0 \end{bmatrix}) = \begin{bmatrix} -4 & 0 \\ 0 & -\frac{2}{3} \end{bmatrix}$$

  Again we have a diagonal matrix, where the eigenvalues are $\lambda_1 = -4 < 0$ and $\lambda_2 = -\frac{2}{3} < 0$, therefore the matrix is negative definite, and the point $\begin{bmatrix} -\frac{4}{3} \\ 0 \end{bmatrix}$ is a strict local maximum.

- $\mathbf{x} = \begin{bmatrix} -1 \\ -1 \end{bmatrix}$:

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

  TODO

- $\mathbf{x} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$:

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

  TODO

d) $f(\mathbf{x}) = \alpha x_1^2 - 2x_1 + \beta x_2^2$

Computing the partial derivatives of f with respect to x_1 and x_2 we get:

$$\frac{\partial f(\mathbf{x})}{\partial x_1} = 2\alpha x_1 - 2$$

$$\frac{\partial f(\mathbf{x})}{\partial x_2} = 2\beta x_2$$

which gives us the gradient:

$$\nabla f(\mathbf{x}) = \begin{bmatrix} 2\alpha x_1 - 2 \\ 2\beta x_2 \end{bmatrix}$$

Next we compute the partial derivatives of the gradient with respect to x_1 and x_2:

$$\frac{\partial \nabla f(\mathbf{x})}{\partial x_1} = \begin{bmatrix} 2\alpha \\ 0 \end{bmatrix}$$

which gives us the Hessian matrix:

$$\nabla^2 f(\mathbf{x}) = \begin{bmatrix} 2\alpha & 0 \\ 0 & 2\beta \end{bmatrix}$$

To get the stationary points, we need $\nabla f(\mathbf{x}) = \begin{bmatrix} 2\alpha x_1 - 2 \\ 2\beta x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$, which gives us the following system of equations:

$$2\alpha x_1 - 2 = 0$$

$$2\beta x_2 = 0$$

Rearranging the first equation, we get

$$\alpha x_1 = 1 \rightarrow x_1 = \frac{1}{\alpha}, \quad \alpha \neq 0$$

We clearly see that if $\alpha = 0$ we have no stationary points because $0x_1 = 1$ is not satisfied for any $x_1 \in \mathbb{R}$.

For the second equation we get:

$$x_2 = \begin{cases} c \in \mathbb{R}, & \text{if } \beta = 0 \\ 0, & \text{if } \beta \neq 0 \end{cases}$$

Because the Hessian matrix is diagonal, we know the eigenvalues are $\lambda_1 = 2\alpha$ and $\lambda_2 = 2\beta$. Therefore α and β are the only parameters that determine whether a candidate point is a minima, maxima or a saddle point. More specifically:

- minima: positive definiteness of the Hessian matrix is a sufficient condition, in other words all eigenvalues have to be strictly positive, so it is necessary that $\alpha > 0$ and $\beta > 0$.
- maxima: negative definiteness of the Hessian matrix is a sufficient condition, in other words all eigenvalues have to be strictly negative, so it is necessary that $\alpha < 0$ and $\beta < 0$.
- saddle point: indefiniteness of the Hessian matrix is a sufficient condition, in other words at least one eigenvalue has to be strictly positive and at least one eigenvalue has to be strictly negative, so either $\alpha > 0$ and $\beta < 0$, or $\alpha < 0$ and $\beta > 0$.

TODO what if beta is 0? I assume you can't do anything unless you know the specific point (x1, x2).

2 Matrix Calculus

a) $f(\mathbf{x}) = \frac{1}{4}\|\mathbf{x} - \mathbf{b}\|^4$ **for** $\mathbf{b} \in \mathbb{R}^n$

b) $f(\mathbf{x}) = \sum_{i=1}^{n} g((\mathbf{A}\mathbf{x})_i)$ **with** $g(z) = \frac{1}{2}z^2 + z$ **for** $z \in \mathbb{R}, \mathbf{A} \in \mathbb{R}^{n\times n}$**, subscript** i **denoting the** i**-th element**

c) $f(\mathbf{x}) = (\mathbf{x} \oslash \mathbf{b})^{\mathbf{T}}\mathbf{D}(\mathbf{x} \oslash \mathbf{b})$ **for** $\mathbf{b} \in \mathbb{R}^n, \mathbb{D} \in \mathbb{R}^{n\times n}$

3 Numerical Gradient Verification

7 Computing Gradient using Automatic Differentiation

8 Training the Neural Network

9 Inference