République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
Université A. Mira de Béjaïa
Faculté des Sciences Exactes
Département de Recherche Opérationnelle

Course Material

intended for 3rd year undergraduate students

Mathematical programming

Dr. Aicha Anzi

Béjaia, 2025.

Preface

This course on Mathematical Programming provides a comprehensive introduction to the theory of optimization. Intended for students in applied mathematics and operations research, it aims to equip them with mathematical tools required to model and solve real world optimization problems. It allows to prepare students for advanced topics in nonlinear programming, convex optimization, and numerical analysis.

The content is organized into five chapters. Chapter 1 establishes the mathematical prerequisites, covering differential calculus in multiple dimensions, quadratic forms with Sylvester's criterion, and convex analysis. Chapter 2 develops the theory of unconstrained optimization, presenting existence results via Weierstrass theorem and characterizing optimal solutions through first and second-order conditions. Chapter 3 introduces numerical methods. It describes iterative algorithms like gradient descent and conjugate gradient, and explains how these methods converge to a solution. Chapter 4 extends the framework to constrained optimization, deriving the fundamental Karush-Kuhn-Tucker conditions. Finally, chapter 5 contains practice exercises with solutions covering all course material.

Contents

1 Review and Supplements **1**

1.1 Introduction . . . 1

1.2 Differential Calculus . . . 2

1.3 Quadratic Forms . . . 5

1.3.1 Definite and Semidefinite Quadratic Forms . . . 5

1.3.2 Matrix Associated with a Quadratic Form . . . 5

1.3.3 Sylvester's Criterion for Symmetric Matrices . . . 6

1.4 Convexity . . . 9

1.4.1 Characterization of Convexity . . . 11

1.4.2 Operations preserving Convexity . . . 12

1.4.3 Convexity of quadratic forms . . . 12

2 Optimization of Nonlinear Functions in $\mathbb{R}^n$ **14**

2.1 Existence of Optimal Solutions . . . 15

2.2 Characterization of Optimal Solutions . . . 15

2.3 Sufficient Optimality Conditions . . . 17

2.4 Optimization of Convex Functions . . . 18

3 Numerical Methods for Minimizing a Differentiable Function in $\mathbb{R}^n$ **20**

3.1 Introduction . . . 20

3.2 Gradient Methods . . . 22

3.3 Conjugate Gradient Method . . . 23

3.3.1 Construction of conjugate directions: Quadratic case . . . 23

4 Optimization of Nonlinear Functions under Constraints **26**

4.1 Necessary optimality conditions . . . 26

4.2 Writing the optimality conditions . . . 27

4.3 Optimization of nonlinear functions under equality constraints . . . 29

4.3.1 Necessary optimality conditions . . . 29

4.3.2 Sufficient optimality condition . . . 30

4.4 Optimization of nonlinear functions under inequality constraints . . . 30

4.4.1 Necessary optimality conditions . . . 31

4.4.2 Sufficient optimality condition . . . 32

4.5 Optimization of nonlinear functions under mixed constraints (equality and inequality) . . . 33

4.5.1 Necessary optimality conditions (KKT) 33
4.5.2 Sufficient optimality condition 35
4.5.3 Convex Optimization 35

5 Exercises **37**
5.1 Exercises with solutions 37
5.2 Exercices without solutions 47

Chapter 1

Review and Supplements

The scalar product is defined by $\langle x, y\rangle = x^\top y = \sum_{i=1}^{n} x_i y_i, \quad x \in \mathbb{R}^n,\ y \in \mathbb{R}^n.$

The associated Euclidean norm is defined by $\|x\| = \sqrt{\langle x, x\rangle} = \sqrt{x^\top x} = \sqrt{\sum_{i=1}^{n} x_i^2},\ x \in \mathbb{R}^n.$

1.1 Introduction

In this course, we are interested in the problem of finding a point $x^* \in X$, such that a real function f defined on the set X takes its minimum value. Formally, this optimization problem is presented as:

$$\begin{cases} \min f(x) \\ \quad s.t. \quad x \in X, \end{cases} \tag{1.1}$$

where $f : X \subset \mathbb{R}^n \to \mathbb{R}$, a function called the **objective** function.
X: is the set of **feasible** solutions and $x \in X$ is called a feasible solution.

We have two cases:

1. If $X = \mathbb{R}^n$, we say that (1.1) is an unconstrained optimization problem.
2. If $X \subset \mathbb{R}^n$, (1.1) is a constrained optimization problem. In this case, X can have the general form: $X = \{x \in \mathbb{R}^n : h_i(x) = 0, g_j(x) \leq 0; i = \overline{1,k}, j = \overline{1,m}\}$, where h_i and g_j are called constraint functions.

In particular, if f, h_i and g_j are linear functions, the problem (1.1) is a linear program (LP) and in this case the set X is a polyhedron.
The formulation (1.1) also encompasses maximization problems. Indeed, we have:

$$\max f(x) = -\min(-f)(x),$$

which allows us to transform a minimization problem into a maximization problem and vice versa.
The solution of problem (1.1) is called an optimum or extremum (minimum or maximum).

Definition 1.1 (Global Optimum)**.** A solution of problem (1.1) is a point $x^* \in X$ such that

$$f(x^*) \leq f(x), \quad \forall x \in X.$$

In this case, x^* is said to be a minimum or global optimum.
For a maximization problem, $x^* \in X$ is a global maximum, if

$$f(x^*) \geq f(x), \quad \forall x \in X.$$

Definition 1.2 (Local Optimum)**.** $x^* \in X$ is a local optimum of problem (1.1), if there exists a neighborhood of x^*, $\mathcal{V}(x^*)$[1] such that

$$f(x^*) \leq f(x), \quad \forall x \in X \cap \mathcal{V}(x^*), \text{ for a minimum.}$$

$$f(x^*) \geq f(x), \quad \forall x \in X \cap \mathcal{V}(x^*), \text{ for a maximum.}$$

Every global extremum is local. The converse is false.

When the inequalities are strict, we speak of a strict optimum.

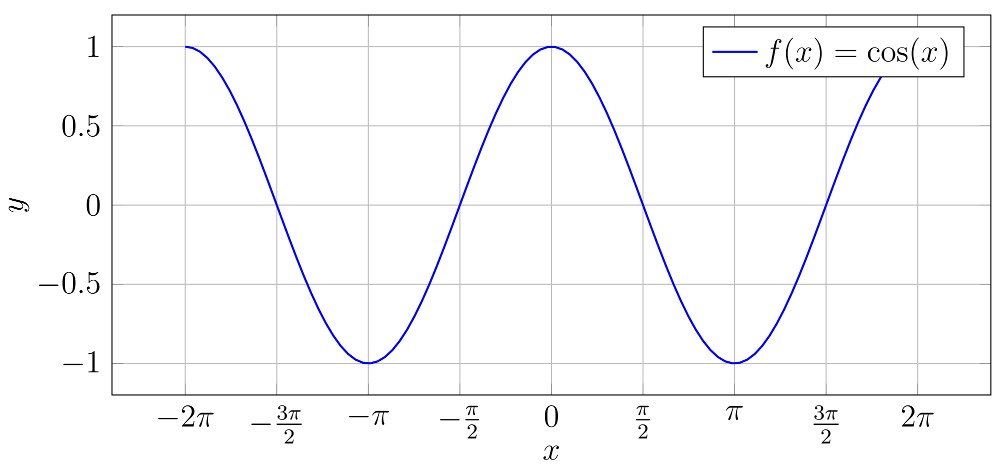


Figure 1.1: Function $f(x) = \cos(x)$ with infinitely many global minima and maxima.

1.2 Differential Calculus

Definition 1.3. We say that $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable at the point $x^* \in \mathbb{R}^n$, if there exists a vector $\nabla f(x^*) \in \mathbb{R}^n$, called the gradient of f at point x^* such that

$$f(x) = f(x^*) + (x - x^*)^\top \nabla f(x^*) + o(\|x - x^*\|), \forall x \in \mathbb{R}^n,$$

where $\lim\limits_{x \to x^*} \frac{o(\|x-x^*\|)}{\|x-x^*\|} = 0$.

[1] $\mathcal{V}(x^*) = \mathcal{B}(x^*, \varepsilon) = \{x \in \mathbb{R}^n / \|x - x^*\| \leq \varepsilon\}, \ \varepsilon > 0$

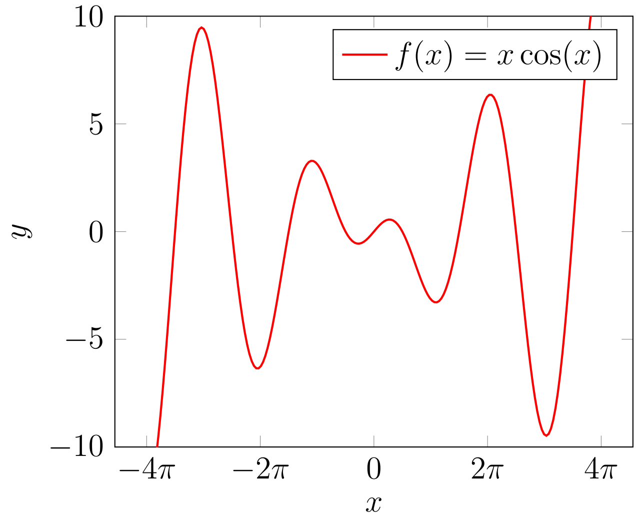


Figure 1.2: Function $f(x) = x\cos(x)$ with infinitely many local minima and maxima.

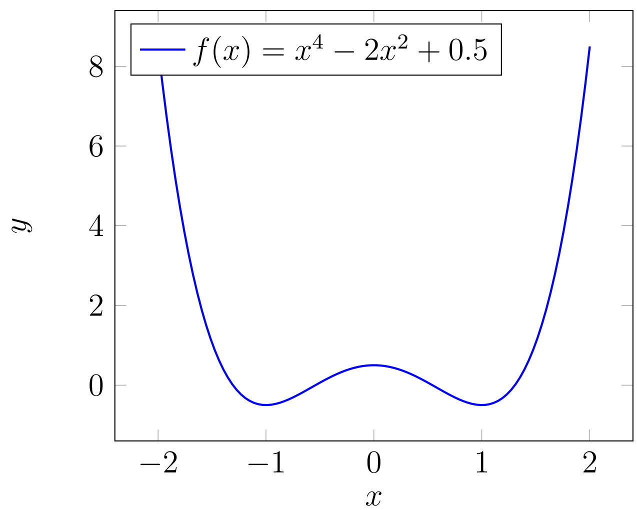


Figure 1.3: Example of a function with a global minimum at $x = \pm 1$ and local maxima at $x = 0$.

The components of the vector $\nabla f(x^*)$ are the partial derivatives of f at x^*, $\frac{\partial f(x^*)}{\partial x_i}$:

$$\begin{aligned}\frac{\partial f(x^*)}{\partial x_i} &= \lim_{h\to 0} \frac{f(x_1^*, x_2^*, \ldots, x_i^* + h, \ldots, x_n^*) - f(x_1^*, x_2^*, \ldots, x_n^*)}{h} \\ &= \lim_{h\to 0} \frac{f(x^* + he_i) - f(x^*)}{h}\end{aligned}$$

where $h \in \mathbb{R}$ and $x^* = (x_1^*, x_2^*, \ldots, x_n^*)^\top \in \mathbb{R}^n$ and $e_i = (0, \ldots, 1, \ldots, 0)^\top$.

If $\nabla f(x^*) : \mathbb{R}^n \to \mathbb{R}$ is continuous, then f is continuously differentiable.

Definition 1.4. We say that $f : \mathbb{R}^n \to \mathbb{R}$ is twice differentiable at the point $x^* \in \mathbb{R}^n$, if in addition to the gradient, there exists a matrix $H(x^*)$, called the Hessian matrix of f at point x^*, such that:

$$f(x) = f(x^*) + (x - x^*)^\top \nabla f(x^*) + \frac{1}{2}(x - x^*)^\top H(x^*)(x - x^*) + o(\|x - x^*\|^2), \forall x \in \mathbb{R}^n,$$

where $\lim\limits_{x \to x^*} \frac{o(\|x-x^*\|^2)}{\|x-x^*\|^2} = 0$.

If the second-order partial derivatives, $\frac{\partial^2 f(x^*)}{\partial x_j \partial x_i}$, $i = \overline{1,n}; j = \overline{1,n}$ exist and are continuous on $\mathbb{R}^n$, then we say that f is of class C^2. In this case, the Hessian is a symmetric matrix

$$H(x^*) = \nabla^2 f(x^*) = \begin{pmatrix} \frac{\partial^2 f(x^*)}{\partial x_1^2} & \frac{\partial^2 f(x^*)}{\partial x_2 \partial x_1} & \cdots & \frac{\partial^2 f(x^*)}{\partial x_n \partial x_1} \\ \frac{\partial^2 f(x^*)}{\partial x_1 \partial x_2} & \frac{\partial^2 f(x^*)}{\partial x_2^2} & \cdots & \frac{\partial^2 f(x^*)}{\partial x_n \partial x_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f(x^*)}{\partial x_1 \partial x_n} & \frac{\partial^2 f(x^*)}{\partial x_2 \partial x_n} & \cdots & \frac{\partial^2 f(x^*)}{\partial x_n^2} \end{pmatrix}$$

Exercise 1.1. Compute the gradient and the Hessian of the functions:

1. $f(x) = f(x_1, x_2) = x_1^2 + x_2^2 - 2x_1x_2 + x_1 + x_2$,
2. $g(x) = g(x_1, x_2, x_3) = x_1^2 + x_2^2 + x_3^2 - 2x_1 - 2x_2 + 1$.

Definition 1.5 (Directional Derivative)**.** The directional derivative of a function f at point x^*, in the direction $d \in \mathbb{R}^n$, denoted $D_f(x^*, d)$, is the limit defined by

$$D_f(x^*, d) = \lim_{h \to 0} \frac{f(x^* + hd) - f(x^*)}{h}.$$

The directional derivative gives us information about the slope of the function in the direction d (just as the derivative gives information about the slope of a single-variable function). Indeed, if

- $D_f(x^*, d) > 0$, then d is a direction of increase for f starting from x^*.
- $D_f(x^*, d) = 0$, nothing can be concluded.
- $D_f(x^*, d) < 0$, then d is a direction of decrease for f starting from x^*. In this case, d is called a *descent* direction of f at x^*. This notion is widely used in optimization to search for the minimum of a function on $\mathbb{R}^n$.

In the case where f is differentiable, we have the following result.

Proposition 1.1. Let $f : \mathbb{R}^n \to \mathbb{R}$ be differentiable and $x^* \in \mathbb{R}^n$. Then for all $d \in \mathbb{R}^n$

$$D_f(x^*, d) = d^\top \nabla f(x^*)$$

1.3 Quadratic Forms

Definition 1.6. A quadratic form of n variables $x_1, x_2, \dots, x_n$ is a real-valued function that can be written as:

$$F(x) = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i x_j \tag{1.2}$$

Letting $x = (x_1, x_2, \dots, x_n)^\top, \quad A = (a_{ij})_{i=\overline{1,n}; j=\overline{1,n}}, \quad (1.2) \Rightarrow F(x) = x^\top A x.$

Example 1.1. Let $x \in \mathbb{R}^3$

$$\begin{aligned} F(x) =& a_{11}x_1^2 + a_{22}x_2^2 + a_{33}x_3^2 + a_{12}x_1x_2 + a_{13}x_1x_3 + a_{21}x_2x_1 + a_{23}x_2x_3 + a_{31}x_3x_1 + a_{32}x_3x_2 \\ =& \begin{pmatrix} x_1, x_2, x_3 \end{pmatrix}^\top \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \\ =& a_{11}x_1^2 + a_{22}x_2^2 + a_{33}x_3^2 + (a_{12} + a_{21})x_1x_2 + (a_{13} + a_{31})x_1x_3 + (a_{23} + a_{32})x_2x_3 \end{aligned}$$

Let $d_{ij} = \frac{a_{ij} + a_{ji}}{2} = d_{ji}, \quad \forall 1 \le i, j \le n.$
The matrix $D = (d_{ij})_{i=\overline{1,n}; j=\overline{1,n}}$ is symmetric and we have

$$F(x) = x^t D x = x^t A x, \quad \forall x \in \mathbb{R}^n.$$

1.3.1 Definite and Semidefinite Quadratic Forms

Let $F(x) = x^\top D x$ be a quadratic form with D symmetric.

- F is **positive definite** if $x^\top D x > 0$ for all $x \in \mathbb{R}^n$, $\|x\| \neq 0$.
- F is **positive semidefinite** if $x^\top D x \ge 0$ for all $x \in \mathbb{R}^n$, $\|x\| \neq 0$.
- F is **negative definite** if $x^\top D x < 0$ for all $x \in \mathbb{R}^n$, $\|x\| \neq 0$.
- F is **negative semidefinite** if $x^\top D x \le 0$ for all $x \in \mathbb{R}^n$, $\|x\| \neq 0$.
- F is **indefinite** if it is positive for some values of x and negative for others.

1.3.2 Matrix Associated with a Quadratic Form

Let D be a symmetric matrix and $F(x) = x^\top D x$ its associated quadratic form. Then

- D is **positive definite** ($D > 0$) if its associated quadratic form is positive definite.
- D is **positive semidefinite** ($D \ge 0$) if its associated quadratic form is positive semidefinite.
- D is **negative definite** ($D < 0$) if its associated quadratic form is negative definite.

- D is **negative semidefinite** ($D \leq 0$) if its associated quadratic form is negative semidefinite.

Example 1.2.

1. $F(x) = x_1^2 + x_2^2$ is positive definite.

$$F(x) = (x_1, x_2) \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = x^\top D x$$

where $D = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} > 0$.

2. $F(x) = x_1^2 - 2x_1x_2 + x_2^2$ is positive semidefinite.

$$F(x) = (x_1 - x_2)^2 \geq 0 \quad \forall x \in \mathbb{R}^2$$

The associated matrix is:

$$D = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \geq 0$$

3. $F(x) = x_1^2 - x_2^2 + 2x_1x_2$ is indefinite.

$$\begin{cases} F(x) = -x_2^2 < 0 & \text{if } x_1 = 0, x_2 \neq 0 \\ F(x) = x_1^2 > 0 & \text{if } x_1 \neq 0, x_2 = 0 \end{cases}$$

The associated matrix is indefinite.

1.3.3 Sylvester's Criterion for Symmetric Matrices

Let a symmetric matrix D of size $n \times n$:

$$D = \begin{pmatrix} d_{11} & d_{12} & \cdots & d_{1n} \\ d_{21} & d_{22} & \cdots & d_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ d_{n1} & d_{n2} & \cdots & d_{nn} \end{pmatrix}$$

Definition 1.7. The minor of order p of the matrix D is given by the determinant of the submatrix of order p, formed by the rows $(i_1, i_2, \ldots, i_p)$ and columns $(j_1, j_2, \ldots, j_p)$ of D and is denoted:

$$D\begin{pmatrix} i_1, i_2, \ldots, i_p \\ j_1, j_2, \ldots, j_p \end{pmatrix} = \begin{vmatrix} d_{i_1j_1} & d_{i_1j_2} & \cdots & d_{i_1j_p} \\ d_{i_2j_1} & d_{i_2j_2} & \cdots & d_{i_2j_p} \\ \vdots & \vdots & \ddots & \vdots \\ d_{i_pj_1} & d_{i_pj_2} & \cdots & d_{i_pj_p} \end{vmatrix}$$

The minor is called *principal* if it is formed from rows and columns with the same indices, i.e., $j_1 = i_1,\ j_2 = i_2 \ldots j_p = i_p$. If furthermore $j_1 = i_1 = 1$, $\forall p \leq n$, then the minor is called *leading principal*. A matrix of order n has exactly n leading principal minors. They are denoted $\Delta_1, \Delta_2, \ldots, \Delta_n$.

Example 1.3. Let D be a symmetric matrix. Then the leading principal minors of D are $\Delta_1 = d_{11},\ \Delta_2 = \begin{vmatrix} d_{11} & d_{12} \\ d_{21} & d_{22} \end{vmatrix},\ \Delta_3 = \begin{vmatrix} d_{11} & d_{12} & d_{13} \\ d_{21} & d_{22} & d_{23} \\ d_{31} & d_{32} & d_{33} \end{vmatrix}$

Its principal minors are $D\begin{pmatrix} 2 \\ 2 \end{pmatrix} = d_{22},\ D\begin{pmatrix} 3 \\ 3 \end{pmatrix} = d_{33},\ D\begin{pmatrix} 2 & 3 \\ 2 & 3 \end{pmatrix} = \begin{vmatrix} d_{22} & d_{23} \\ d_{32} & d_{33} \end{vmatrix}$ and $D\begin{pmatrix} 1 & 3 \\ 1 & 3 \end{pmatrix} = \begin{vmatrix} d_{11} & d_{13} \\ d_{31} & d_{33} \end{vmatrix}$.

However, the minors $D\begin{pmatrix} 1 \\ 2 \end{pmatrix} = d_{12},\ D\begin{pmatrix} 2 & 3 \\ 1 & 2 \end{pmatrix} = \begin{vmatrix} d_{21} & d_{22} \\ d_{31} & d_{32} \end{vmatrix}$ are not principal.

Theorem 1.2 (Sylvester's Criterion)**.** *Let D be a symmetric matrix of order n.*

1. *For D to be positive definite ($D > 0$), it is necessary and sufficient that all leading principal minors be positive:*

$$\Delta_1 > 0, \quad \Delta_2 > 0, \quad \dots, \quad \Delta_n > 0.$$

2. *For D to be positive semidefinite ($D \geq 0$), it is necessary and sufficient that all its principal minors be non-negative:*

$$D\begin{pmatrix} i_1 & i_2 & \dots & i_p \\ j_1 & j_2 & \dots & j_p \end{pmatrix} \geq 0, \quad \begin{matrix} \forall 1 \leq i_1 < i_2 < \dots < i_p \leq n; \\ j_1 = i_1,\ j_2 = i_2 \dots j_p = i_p,\ p = 1, \dots, n. \end{matrix}$$

Remark 1.1. Consider the matrix

$$D = \begin{pmatrix} 0 & 0 \\ 0 & -1 \end{pmatrix} \quad \Delta_1 = 0,\ \Delta_2 = 0$$

The leading principal minors are all non-negative, yet the associated quadratic form $F(x) = x^\top D x = -x_2^2 < 0, \quad \forall x \in \mathbb{R}^2$ is negative definite. So D is negative definite. Consequently, the condition $\Delta_1 \geq 0, \quad \Delta_2 \geq 0, \dots, \Delta_n \geq 0$ is not sufficient for a matrix to be positive semidefinite. The other principal minors of the matrix must be checked.

Remark 1.2. Sylvester's criterion also applies to negative definite and negative semidefinite matrices. For a symmetric matrix D, it is stated as follows:

1. $D < 0 \Leftrightarrow (-1)^p \Delta_p > 0,\ \forall p = 1, \dots, n.$
2. $D \leq 0 \Leftrightarrow (-1)^p D\begin{pmatrix} i_1 & i_2 & \dots & i_p \\ j_1 & j_2 & \dots & j_p \end{pmatrix} \geq 0, \quad \begin{matrix} \forall 1 \leq i_1 < i_2 < \dots < i_p \leq n; \\ j_1 = i_1,\ j_2 = i_2 \dots j_p = i_p,\ p = 1, \dots, n. \end{matrix}$

Example 1.4.

1. $A = \begin{pmatrix} 6 & 3 & 0 \\ 3 & 6 & 9 \\ 0 & 9 & 18 \end{pmatrix}$

$$\Delta_1 = \begin{vmatrix} 6 \end{vmatrix} = 6 > 0, \quad \Delta_2 = \begin{vmatrix} 6 & 3 \\ 3 & 6 \end{vmatrix} = 27 > 0, \quad \Delta_3 = \begin{vmatrix} 6 & 3 & 0 \\ 3 & 6 & 9 \\ 0 & 9 & 18 \end{vmatrix} = 0.$$

According to Sylvester's criterion, A is not positive definite. Is it positive semidefinite? To answer this question, we must compute all principal minors.

Principal minors of order $p = 1$:
$A\begin{pmatrix} 2 \\ 2 \end{pmatrix} = 6 > 0,\ A\begin{pmatrix} 3 \\ 3 \end{pmatrix} = 18 > 0.$

Principal minors of order $p = 2$:
$A\begin{pmatrix} 2 & 3 \\ 2 & 3 \end{pmatrix} = \begin{vmatrix} 6 & 9 \\ 9 & 18 \end{vmatrix} = 27 > 0,\ A\begin{pmatrix} 1 & 3 \\ 1 & 3 \end{pmatrix} = \begin{vmatrix} 3 & 0 \\ 0 & 18 \end{vmatrix} = 108 > 0.$
All principal minors are positive or zero. A is therefore positive semidefinite.

2. $B = \begin{pmatrix} 5 & -1 & 2 \\ -1 & 10 & -2 \\ 2 & -2 & 4 \end{pmatrix}$

$$\Delta_1 = \begin{vmatrix} 5 \end{vmatrix} = 5 > 0, \quad \Delta_2 = \begin{vmatrix} 5 & -1 \\ -1 & 10 \end{vmatrix} = 50-1 = 49 > 0, \quad \Delta_3 = \begin{vmatrix} 5 & -1 & 2 \\ -1 & 10 & -2 \\ 2 & -2 & 4 \end{vmatrix} = 144 > 0$$

$\Delta_p > 0,\ \forall p = \overline{1,3} \Rightarrow B > 0.$

3. $C = \begin{pmatrix} -1 & 0 & 1 \\ 0 & -1 & 0 \\ 1 & 0 & -2 \end{pmatrix}$

$$\Delta_1 = \begin{vmatrix} -1 \end{vmatrix} = -1 < 0, \quad \Delta_2 = \begin{vmatrix} -1 & 0 \\ 0 & -1 \end{vmatrix} = 1 > 0, \quad \Delta_3 = \begin{vmatrix} -1 & 0 & 1 \\ 0 & -1 & 0 \\ 1 & 0 & -2 \end{vmatrix} = -1.$$

We have $(-1)^1(-1) = 1 > 0,\ (-1)^2(1) = 1 > 0,\ (-1)^1(-1) = 1 > 0$. So $C < 0$ (negative definite).

4. $D = \begin{pmatrix} 1 & 2 & -1 \\ 2 & -1 & 0 \\ -1 & 0 & 1 \end{pmatrix}$

$$\Delta_1 = \begin{vmatrix} 1 \end{vmatrix} = 1 > 0, \quad \Delta_2 = \begin{vmatrix} 1 & 2 \\ 2 & -1 \end{vmatrix} = -5 > 0.$$

We can conclude without calculating the other minors that the matrix D is indefinite. Indeed, for Δ_1, we have $(-1)^1(1) = 1 < 0$ (D cannot be negative definite) and since $\Delta_2 < 0$, D is not positive definite. If we consider the associated quadratic form $F(x) = x^\top Dx,\ x \in \mathbb{R}^3$; we have: $F(1,0,0) = 1 > 0$ and $F(0,1,0) = -1 < 0$

Remark 1.3. Sylvester's criterion is only applicable to symmetric matrices. Indeed, consider the matrix $A = \begin{pmatrix} -1 & -2 & -1 \\ 2 & -1 & 0 \\ 1 & 0 & -1 \end{pmatrix}$

$$\Delta_1 = |-1| = -1 < 0,\ \Delta_2 = \begin{vmatrix} 1 & -2 \\ 2 & -1 \end{vmatrix} = 5 > 0,\ \Delta_3 = \begin{vmatrix} -1 & -2 & -1 \\ 2 & -1 & 0 \\ 1 & 0 & -1 \end{vmatrix} = -4 < 0$$

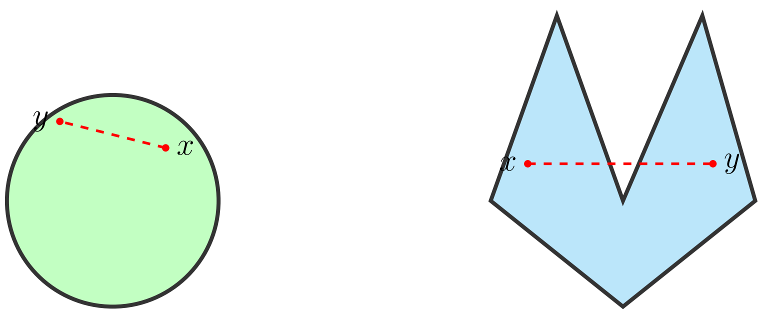
According to remark 1.2, A is negative definite. However, for $x = (1, 0, -1)$, $\|x\| \neq 0$, the associated quadratic form $x^\top Dx = 0$. This incorrect result is due to the non-symmetry of A.

1.4 Convexity

Definition 1.8 (Convex Set)**.** A set $X \subset \mathbb{R}^n$ is said to be convex if, for all $x, y \in X$ and all $\lambda \in [0, 1]$, we have:

$$z = \lambda x + (1 - \lambda)y \in X.$$

In other words, a set X is convex if every segment joining two points of X is contained in X.



A - convex B - non convex

Figure 1.4: Example of convex and non-convex sets

Example 1.5.

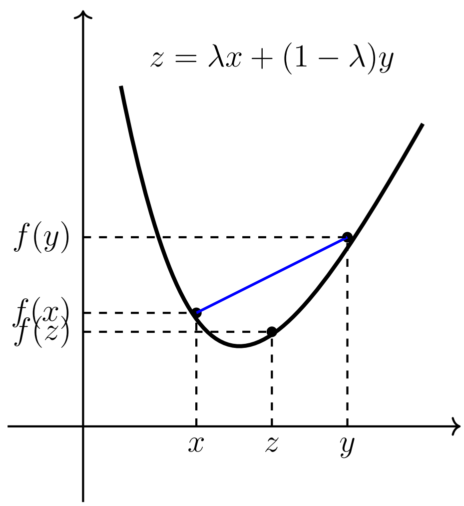
1. Intervals in $\mathbb{R}$ (open, closed, or half-open) are convex.
2. Balls in $\mathbb{R}^n$ ($B(x_0, r) = \{x \in \mathbb{R}^n : \|x - x_0\| \leq r\}$) are convex.
3. Spheres in $\mathbb{R}^n$ ($S(x_0, r) = \{x \in \mathbb{R}^n : \|x - x_0\| = r\}$ are not convex.

Property 1.3.

1. An intersection of convex sets is convex, $X = \bigcap_{i=1}^{n} X_i$.
2. The sum of two convex sets, X_1 and X_2 is convex, $X = X_1 + X_2 = \{x + y; x \in X_1, y \in X_2\}$.
3. The set defined by $\{\lambda x, x \in X, \lambda \geq 0\}$ is convex.

Definition 1.9 (Convex Function)**.** A function $f : X \subset \mathbb{R}^n \to \mathbb{R}$ is said to be convex on the convex set X, if for all $x, y \in \mathbb{R}^n$ and all $\lambda \in [0, 1]$, we have:

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y).$$

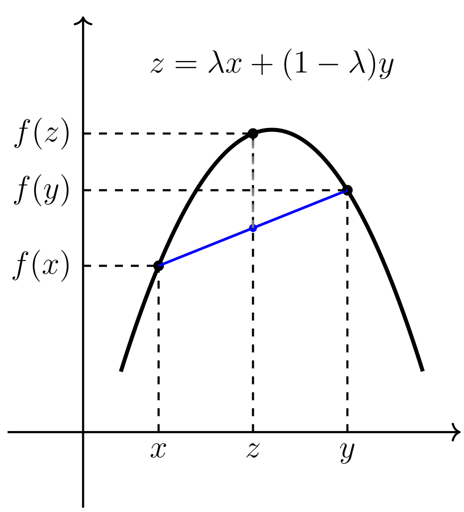


The segment joining the two points $(x, f(x))$ and $(y, f(y))$ is always above the graph of f.

Figure 1.5: Geometric interpretation of convexity

Definition 1.10 (Concave Function)**.** A function $f : X \subset \mathbb{R}^n \to \mathbb{R}$ is said to be concave on the convex set X, if for all $x, y \in \mathbb{R}^n$ and all $\lambda \in [0, 1]$, we have:

$$f(\lambda x + (1 - \lambda)y) \geq \lambda f(x) + (1 - \lambda)f(y).$$



The segment joining the two points $(x, f(x))$ and $(y, f(y))$ is always below the graph of f.

Figure 1.6: Graph of a concave function

Remark 1.4.

Affine functions are both convex and concave.

Property 1.4. A function $f : X \to \mathbb{R}$ is concave on the convex set X if $(-f)$ is convex on X.

1.4.1 Characterization of Convexity

In general, it is difficult to verify convexity using the definition. The following result gives criteria for convexity of differentiable functions.

Proposition 1.5. [3] Let $X \subset \mathbb{R}^n$ be convex and $f : X \to \mathbb{R}$ a function of class C^1. Then:

1. f is convex on $X \Leftrightarrow f(y) \geq f(x) + [\nabla f(x)]^\top (y - x)$, $\forall x, y \in X$.
2. f is convex on $X \Leftrightarrow \nabla f(x)$ is an increasing function on X, i.e.:
$$[\nabla f(y) - \nabla f(x)]^\top (y - x) \geq 0, \forall x, y \in X.$$
3. If moreover $f \in C^2$, then f is convex on $X \Leftrightarrow \forall x \in X, x^\top \nabla^2 f(x) x \geq 0$.

We define below two important concepts for the characterization of convex functions and sets.

Definition 1.11 (Epigraph)**.** The epigraph of a function $f : \mathbb{R}^n \to \mathbb{R}$ is the set:
$$\text{epi}(f) = \{(x, \alpha) \in \mathbb{R}^n \times \mathbb{R} \mid f(x) \leq \alpha\}.$$

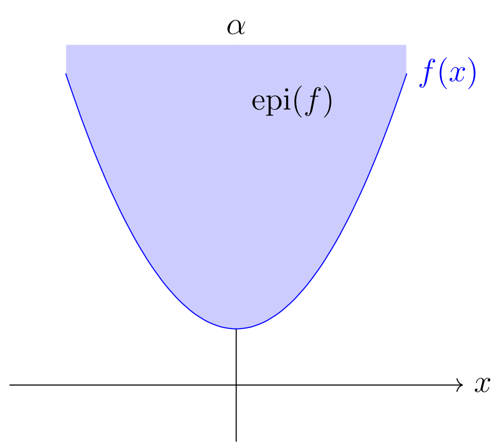


Figure 1.7: Epigraph of a convex function.

Theorem 1.6. *[3] A function $f : \mathbb{R}^n \to \mathbb{R}$ is convex if and only if its epigraph epi(f) is a convex set.*

Definition 1.12 (Level Set)**.** For a function $f : \mathbb{R}^n \to \mathbb{R}$ and a constant $\alpha \in \mathbb{R}$, the level set α is defined by:
$$\mathbf{L}_\alpha(f) = \{x \in \mathbb{R}^n \mid f(x) \leq \alpha\}.$$

Proposition 1.7. If $f : \mathbb{R}^n \to \mathbb{R}$ is a convex function, then for all $\alpha \in \mathbb{R}$, its level set $\mathbf{L}_\alpha(f)$ is either empty or convex.

1.4.2 Operations preserving Convexity

Let $X \subset \mathbb{R}^n$ be convex. Let $f_1, f_2, ..., f_m : X \to \mathbb{R}$ be convex functions on X and $\alpha_1, ..., \alpha_m$ strictly positive constants.

- The function $f = \alpha_1 f_1 + \alpha_2 f_2 + ... + \alpha_m f_m$ is convex on X. If at least one of the functions f_i is strictly convex, then f is strictly convex.
- Let $g : \mathbb{R} \to \mathbb{R}$ be a convex and increasing function. Then the function $g \circ f$ is convex. For example, $g(x) = e^x$ and $f(x) = x^2$.

1.4.3 Convexity of quadratic forms

Property 1.8. Let a quadratic form be $F(x) = x^\top Dx$, where D is symmetric.
Then:

$$F(x) \text{ is convex} \Leftrightarrow D \geq 0$$

Proof. $\Rightarrow$:
Let F be a convex quadratic form and suppose that $D \ngeq 0$. Then $\exists x \in \mathbb{R}^n$, $\|x\| \neq 0$ such that $x^\top Dx < 0$.
Therefore for $y = -x$, we have $y^\top Dy < 0$.
Let us take $z = \frac{1}{2}x + \frac{1}{2}y$.
Since F is convex:

$$F(z) = z^\top Dz = F\left(\frac{1}{2}x + \frac{1}{2}y\right) \leq \frac{1}{2}F(x) + \frac{1}{2}F(y) = F(0) = 0 \quad (\text{since } y = -x)$$

On the other hand:

$$F(z) = x^\top Dx + y^\top Dy < 0$$

Contradiction. Therefore $D \geq 0$.
$\Leftarrow$:
Let D be symmetric with $D \geq 0$ and $\lambda \in [0, 1]$, $x \in \mathbb{R}^n$, $y \in \mathbb{R}^n$. We have:

$$\begin{aligned} F(\lambda x + (1-\lambda)y) &= [\lambda x + (1-\lambda)y]^\top D[\lambda x + (1-\lambda)y] \\ &= \lambda^2 x^\top Dx + (1-\lambda)^2 y^\top Dy + 2\lambda(1-\lambda)y^\top Dx \end{aligned}$$

Since $D \geq 0$:

$$x^\top Dx \geq 0$$
$$y^\top Dy \geq 0$$

We also have:

$$(x-y)^\top D(x-y) \geq 0$$
$$\Rightarrow x^\top Dx + y^\top Dy - 2y^\top Dx \geq 0$$
$$\Rightarrow 2y^\top Dx \leq x^\top Dx + y^\top Dy$$

Therefore:

$$\Rightarrow 2\lambda(1-\lambda)y^\top Dx \leq \lambda(1-\lambda)[x^\top Dx + y^\top Dy]$$

Thus:

$$\begin{aligned}
F(\lambda x + (1-\lambda)y) &= \lambda^2 x^\top Dx + (1-\lambda)^2 y^\top Dy + 2\lambda(1-\lambda)y^\top Dx \\
&\leq \lambda^2 x^\top Dx + (1-\lambda)^2 y^\top Dy + \lambda(1-\lambda)[x^\top Dx + y^\top Dy] \\
&= \lambda^2 x^\top Dx + (1-\lambda)^2 y^\top Dy + \lambda(1-\lambda)x^\top Dx + \lambda(1-\lambda)y^\top Dy \\
&= x^\top Dx[\lambda^2 + \lambda(1-\lambda)] + y^\top Dy[(1-\lambda)^2 + \lambda(1-\lambda)] \\
&= \lambda x^\top Dx[\lambda + (1-\lambda)] + (1-\lambda)y^\top Dy[\lambda + (1-\lambda)] \\
&= \lambda x^\top Dx + (1-\lambda)y^\top Dy \\
&= \lambda F(x) + (1-\lambda)F(y)
\end{aligned}$$

Therefore: $F(\lambda x + (1-\lambda)y) \leq \lambda F(x) + (1-\lambda)F(y) \Rightarrow F$ is convex. □

Remark 1.5. By proceeding in a similar way as for property 1.8, we can prove the following result:

$$F(x) \text{ is strictly convex} \Leftrightarrow D > 0,$$

where D is a symmetric matrix.

Chapter 2

Optimization of Nonlinear Functions in $\mathbb{R}^n$

The problem considered in this chapter is of the form:

$$\min_{x\in\mathbb{R}^n} f(x),$$

with $f : \mathbb{R}^n \to \mathbb{R}$ a differentiable function.

Definition 2.1 (Coercive Function)**.** A function $f : X \subseteq \mathbb{R}^n \to \mathbb{R}$ is said to be **coercive** if:

$$\lim_{\|x\|\to\infty} f(x) = +\infty.$$

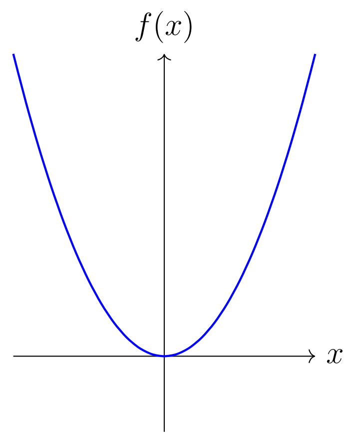


Figure 2.1: Example of a coercive function: $f(x) = x^2$.

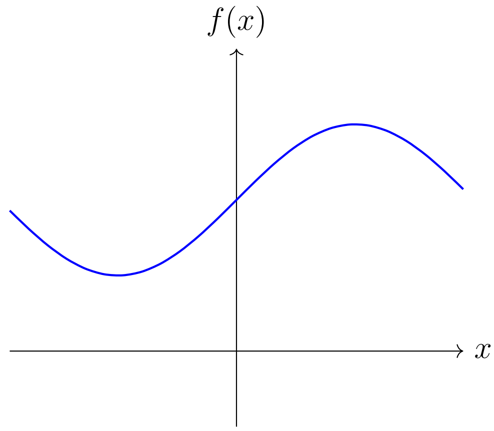


Figure 2.2: Example of a non-coercive function: $f(x) = 2 + \sin(x)$.

2.1 Existence of Optimal Solutions

Theorem 2.1 (Weierstrass)**.** *Let $X \subseteq \mathbb{R}^n$ be non-empty and closed set, and $f : X \subseteq \mathbb{R}^n \to \mathbb{R}$ a lower semicontinuous function on X.*

1. *If X is compact, there exists $x^* \in X$ such that:*
$$f(x^*) = \min_{x \in X} f(x).$$
2. *If f is coercive, there exists $x^* \in X$ such that:*
$$f(x^*) = \min_{x \in X} f(x).$$

Corollary 2.2. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a lower semicontinuous and coercive function. Then there exists $x^* \in \mathbb{R}^n$ such that:
$$\min_{x \in \mathbb{R}^n} f(x) = f(x^*).$$

2.2 Characterization of Optimal Solutions

Theorem 2.3 (Necessary Optimality Condition)**.** *Let $x^* \in \mathbb{R}^n$ be a local minimum point of f, and suppose that $f : \mathbb{R}^n \to \mathbb{R}$ is continuously differentiable on an open subset $S \subseteq \mathbb{R}^n$ containing x^*. Then:*
$$\nabla f(x^*) = 0 \quad (\textit{First order necessary condition}).$$
If moreover, f is twice continuously differentiable on S, then the Hessian matrix is positive semidefinite:
$$\nabla^2 f(x^*) \geq 0 \quad (\textit{Second order necessary condition}).$$

Proof. Suppose that x^* is a local minimum point of f. Then, for all $d \in \mathbb{R}^n$ and $\alpha \in \mathbb{R}$ sufficiently small, we have:

$$f(x^* + \alpha d) \geq f(x^*).$$

Using the first order Taylor expansion around x^*, we obtain:

$$f(x^* + \alpha d) = f(x^*) + \alpha d^\top \nabla f(x^*) + o(\alpha).$$

Thus, for α sufficiently small, we have:

$$f(x^* + \alpha d) - f(x^*) = \alpha d^\top \nabla f(x^*) + o(\alpha) \geq 0.$$

Dividing by α and taking the limit $\alpha \to 0^+$, we obtain:

$$d^\top \nabla f(x^*) \geq 0 \quad \forall d \in \mathbb{R}^n.$$

Similarly, considering $\alpha \to 0^-$, we obtain:

$$d^\top \nabla f(x^*) \leq 0 \quad \forall d \in \mathbb{R}^n.$$

Therefore, we have:

$$d^\top \nabla f(x^*) = 0 \quad \forall d \in \mathbb{R}^n,$$

which implies:

$$\nabla f(x^*) = 0.$$

If f is twice continuously differentiable, we can use the second order Taylor expansion:

$$f(x^* + \alpha d) = f(x^*) + \alpha d^\top \nabla f(x^*) + \frac{1}{2}\alpha^2 d^\top \nabla^2 f(x^*) d + o(\alpha^2).$$

Since $\nabla f(x^*) = 0$, we have:

$$f(x^* + \alpha d) - f(x^*) = \frac{1}{2}\alpha^2 d^\top \nabla^2 f(x^*) d + o(\alpha^2) \geq 0.$$

Dividing by α^2 and taking the limit $\alpha \to 0$, we obtain:

$$d^\top \nabla^2 f(x^*) d \geq 0 \quad \forall d \in \mathbb{R}^n,$$

which means that $\nabla^2 f(x^*)$ is positive semidefinite. □

Definition 2.2. A point $x^* \in \mathbb{R}^n$ satisfying the first order necessary condition ($\nabla f(x^*) = 0$) is called a critical or stationary point.

Remark 2.1. The first and second order necessary conditions are not sufficient for optimality.

Example 2.1. Consider the function $f : \mathbb{R} \to \mathbb{R}$ defined by:

$$f(x) = x^3.$$

The gradient of f is given by:$f'(x) = 3x^2$. Solving $f'(x) = 0$, we find a critical point at $x = 0$.

The second derivative of f is:$f''(x) = 6x$. At the critical point $x = 0$, we have: $f''(0) = 0$. Thus, $x = 0$ satisfies the first and second order necessary conditions, yet it is neither a local minimum nor a local maximum. Indeed:

For $x > 0$, $f(x) > 0$.

For $x < 0$, $f(x) < 0$.

Consequently, $x = 0$ is a saddle point (see figure 2.3).

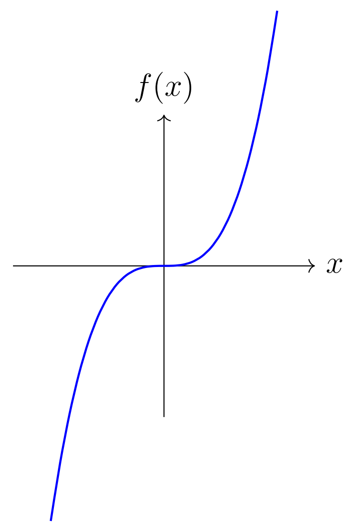


Figure 2.3: Graph of $f(x) = x^3$ with a critical point at $x = 0$.

2.3 Sufficient Optimality Conditions

Theorem 2.4 (Sufficient Optimality Condition)**.** *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a twice continuously differentiable function on an open set $S \subseteq \mathbb{R}^n$. If $x^* \in S$ satisfies:*
1) $\nabla f(x^*) = 0$,
2) $\nabla^2 f(x^*) > 0$ (the Hessian is positive definite),
then x^* is a strict local minimum point.

Proof. Using the second order Taylor expansion, for $d \in \mathbb{R}^n$ and $\alpha > 0$ small enough:

$$f(x^* + \alpha d) = f(x^*) + \alpha d^\top \nabla f(x^*) + \frac{1}{2}\alpha^2 d^\top \nabla^2 f(x^*) d + o(\alpha^2).$$

Since $\nabla f(x^*) = 0$, we have:

$$f(x^* + \alpha d) - f(x^*) = \frac{1}{2}\alpha^2 d^\top \nabla^2 f(x^*) d + o(\alpha^2).$$

As $\nabla^2 f(x^*) > 0$, we have:

$$f(x^* + \alpha d) - f(x^*) > 0 \implies f(x^*) < f(x^* + \alpha d).$$

Thus, x^* is a strict local minimum point. □

2.4 Optimization of Convex Functions

Convexity is of great importance in optimization. First, because in convex optimization, every local minimum is also global. But also, thanks to convexity, the first-order necessary condition is sufficient to characterize this global minimum.

Theorem 2.5 (Global Minimum of Convex Functions). *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function. Then every local minimum point is a global minimum point.*

Proof. Suppose that x^* is a local minimum. By contradiction, suppose there exists $\overline{x}$ such that $f(\overline{x}) < f(x^*)$.

Let $x_\lambda = \lambda\overline{x} + (1-\lambda)x^*$, $\lambda \in [0,1]$. Then:

$$f(x^*) \leq f(x_\lambda).$$

Since f is convex, we have:

$$f(x_\lambda) \leq \lambda f(\overline{x}) + (1-\lambda) f(x^*).$$

Thus:

$$f(x^*) \leq \lambda f(\overline{x}) + (1-\lambda) f(x^*) \implies f(x^*) \leq f(\overline{x}),$$

which is a contradiction. □

Theorem 2.6 (Uniqueness of the Global Minimum). *If $f : \mathbb{R}^n \to \mathbb{R}$ is a strictly convex function, then it admits a unique global minimum point.*

Proof. Suppose that f admits two global minimum points x^* and $\overline{x}$. Then:

$$f(x^*) = f(\overline{x}) = \min_{x \in \mathbb{R}^n} f(x).$$

Let $\lambda \in [0,1]$, $x_\lambda = \lambda x^* + (1-\lambda)\overline{x}$. Since f is strictly convex, we have:

$$f(x_\lambda) < \lambda f(x^*) + (1-\lambda) f(\overline{x}) = f(x^*).$$

Then

$$f(x_\lambda) < f(x^*),$$

which is a contradiction. □

Theorem 2.7 (Necessary and Sufficient Condition). *Let $f : \mathbb{R}^n \to \mathbb{R}$ be convex and continuously differentiable. Then: x^* is a global minimum point if and only if*

$$\nabla f(x^*) = 0.$$

Proof.

- $\Rightarrow$ (see Theorem 2.3).

- $\Leftarrow$

Let $x^* \in \mathbb{R}^n$ such that $\nabla f(x^*) = 0$ and let $x \in \mathbb{R}^n$ be arbitrary.
Since f is convex, we have:

$$f(x) - f(x^*) \geq (x - x^*)^\top \cancelto{0}{\nabla f(x^*)}$$

$$\Rightarrow f(x) - f(x^*) \geq 0$$

$$\Rightarrow f(x) \geq f(x^*) \Rightarrow x^* \text{ global minimum point}$$

□

Remark 2.2. Theorems 2.5, 2.6 and 2.7 remain valid when restricting the domain of f to a convex set $X \subset \mathbb{R}^n$.

Chapter 3

Numerical Methods for Minimizing a Differentiable Function in $\mathbb{R}^n$

3.1 Introduction

Consider the problem

$$\min_{x \in \mathbb{R}^n} f(x) = f(x^*), \tag{3.1}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is a differentiable function. The solutions of this problem are critical points satisfying:

$$\nabla f(x^*) = 0$$

.

Numerical methods for finding solutions to problem (3.1) are iterative algorithms following the scheme:

- Generate a sequence of points $\{x^k\}_{k \in \mathbb{N}}$ such that $f(x^{k+1}) \leq f(x^k)$
- Choose a direction d_k at each iteration k.
- Choose a step size $\theta_k > 0$.

The choice of the direction d_k and the step size θ_k depends on the point x^k and determines the next iteration:

$$x^{k+1} = x^k + \theta_k d_k \tag{3.2}$$

At each iteration of this iterative process, d_k must satisfy the descent condition

$$[\nabla f(x^k)]^\top d < 0$$

If $\nabla f(x^k) = 0$, then the iterative process stops and x^k is a candidate to be an extremum.

Remark 3.1 (Geometric interpretation)**.**
For $f : \mathbb{R}^n \to \mathbb{R}$, $x^* \in \mathbb{R}^n$, $d \in \mathbb{R}^n$, define:

$$x(\theta) = x^* + \theta d, \quad \theta > 0$$

where $x(\theta)$ represents the point reached after moving a step θ in the direction d. Consider the first order Taylor expansion around x^*:

$$f(x^* + \theta d) - f(x^*) = \nabla f(x^*)^\top (\theta d) + o(\theta) \tag{3.3}$$

where $\lim_{\theta \to 0} \frac{o(\theta)}{\theta} = 0$.

$$(3.3) \Rightarrow f(x^* + \theta d) - f(x^*) = \theta \left(\nabla f(x^*)^\top d + \frac{o(\theta)}{\theta} \right)$$

When $\theta \to 0$

$$Sign[f(x^* + \theta d) - f(x^*)] = Sign[\nabla f(x^*)^\top d]$$

Therefore, to have $f(x^* + \theta d) < f(x^*)$ for $\theta > 0$ sufficiently small, it is necessary to have:

$$\nabla f(x^*)^\top d < 0$$

This condition implies that the vectors $\nabla f(x^*)$ and d must be oriented in opposite directions.

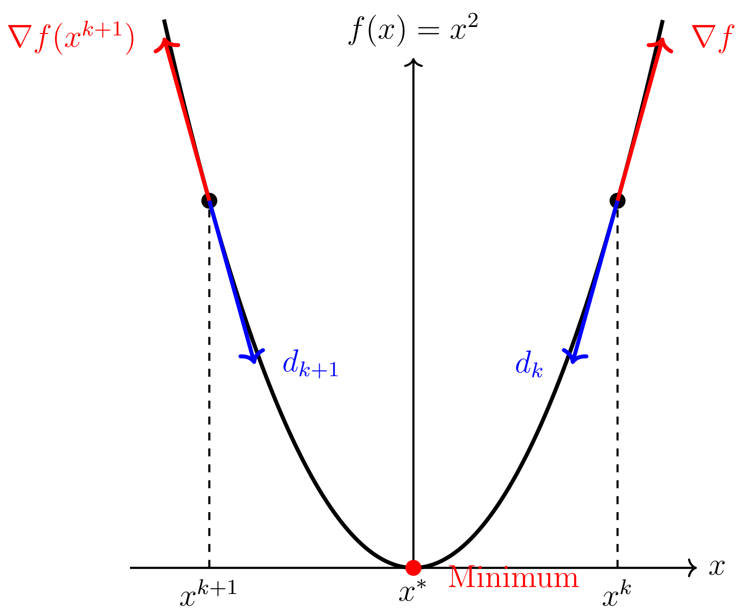


Figure 3.1: Gradient descent for $f(x) = x^2$

Definition 3.1 (Descent Direction)**.** A vector $d \in \mathbb{R}^n$ is called a *descent direction* at point x^* (see figure 3.1) if:

$$[\nabla f(x^*)]^\top d < 0 \tag{3.4}$$

There are several families of algorithms, each differing in the choice of the direction d and the step θ. In this course, we present two families of algorithms: gradient methods and conjugate gradient methods. It should also be noted that there are several rules for choosing the step size. However, this part will not be covered in this course.

3.2 Gradient Methods

Gradient methods constitute a class of iterative algorithms for minimizing differentiable functions. The principle is based on:

- Determining a descent direction $d \in \mathbb{R}^n$ (with $\|d\| = 1$) that maximizes the local rate of decrease of the function along this direction. Starting from a point x^k, the goal is to reach x^{k+1} with $f(x^{k+1}) = f(x^k + \theta d)$ as small as possible (with $\theta \in \mathbb{R}$).
- Since d is a descent direction at x^k, we have $\nabla f(x^k)^\top d < 0$. We then seek $d \in \mathbb{R}^n$ that minimizes this expression.
- Let α be the angle between $\nabla f(x^k)$ and d. We have:

$$\nabla f(x^k)^\top d = \|\nabla f(x^k)\| \cdot \cos\alpha \quad (\|d\| = 1)$$

- This expression is minimized when $\cos\alpha = -1$, i.e., when the direction d is exactly opposite to the gradient:

$$d = -\nabla f(x^k)$$

  or (normalized direction):

$$d = -\frac{\nabla f(x^k)}{\|\nabla f(x^k)\|} \quad (\text{since } \|d\| = 1)$$

This direction, called the *steepest descent direction*, guarantees the maximum local decrease of the function at each iteration. The direction being fixed, the difference between the variants of these algorithms lies in the choice of the step size θ. It can be fixed to a constant or calculated optimally.

Algorithm 1 Gradient Descent with Fixed Step Size

Require: Initial point x^0, step size $\theta > 0$, tolerance ϵ

```
k ← 0
while ‖∇f(x^k)‖ > ε do
    d_k ← −∇f(x^k)
    x^{k+1} ← x^k + θ d_k
    k ← k + 1
end while
return x^k
```

Algorithm 2 Gradient Descent with Optimal Step Size

Require: Initial point x^0, tolerance ϵ

$k \leftarrow 0$
while $\|\nabla f(x^k)\| > \epsilon$ **do**
  $d_k \leftarrow -\nabla f(x^k)$
  Compute θ_k such that:

$$f(x^k + \theta_k d_k) = \min_{\theta \geq 0} f(x^k + \theta d_k)$$

  $x^{k+1} \leftarrow x^k + \theta_k d_k$
  $k \leftarrow k + 1$
end while
return x^k

3.3 Conjugate Gradient Method

The idea of the Conjugate Gradient Method is to iteratively construct mutually conjugate directions that guarantee convergence in a finite number of iterations. Initially designed for convex quadratic functions, it avoids the problem encountered when applying the simple gradient method, which can oscillate without much progress in the case of functions with narrow, elongated valleys. By ensuring that each new search direction is conjugate to the previous ones, the method effectively accelerates convergence compared to the simple gradient approach.

Definition 3.2 (Conjugate Directions)**.** Let A be a symmetric positive definite matrix. Two non-zero vectors d_i and d_j are said to be *conjugate* with respect to A (or A-*conjugate*), if:

$$d_i^\top A d_j = 0 \quad \text{for } i \neq j$$

Definition 3.3. A family of vectors $\{d_0, d_2 \ldots, d_{n-1}\}$ is said to be Q-conjugate, if:

$$d_i^\top A d_j = 0 \quad \forall i \neq j$$

3.3.1 Construction of conjugate directions: Quadratic case

Let the quadratic function

$$F(x) = \frac{1}{2} x^\top A x - b^\top x,$$

with A symmetric positive definite of dimension $n \times n$.
The gradient is written $\nabla F(x) = Ax - b$. We define the residual by $r = \nabla F(x) = Ax - b$.

The conjugate gradient minimizes F in n steps, following n directions $d_0, d_2, \ldots, d_{n-1}$ mutually conjugate with respect to A. Starting with $d_0 = -\nabla F(x^0)$, the method chooses the new direction as a linear combination of the current gradient and the direction from the previous iteration. We then have the following formula:

$$d_{k+1} = -r_{k+1} + \beta_k d_k,$$

where the coefficient β_k is calculated by imposing the conjugacy condition and is given by

$$\beta_k = \frac{r_{k+1}^\top A d_k}{d_k^\top d_k}$$

At each iteration, the optimal step θ_k is calculated to minimize the function $F(x^{k+1})$, with $x^{k+1} = x^k + \theta d_k$. The function F being convex (since $A > 0$), the optimality condition gives

$$\frac{d}{d\theta} F(x^k + \theta d_k) = d_k^\top A(x^k + \theta d_k) - b^\top d_k = 0$$

Expanding, we obtain:

$$\theta_k = -\frac{d_k^\top (Ax^k - b)}{d_k^\top A d_k} = -\frac{d_k^\top r_k}{d_k^\top A d_k},$$

where $r_k = Ax^k - b$.

The conjugate gradient algorithm for a quadratic function is stated as Algorithm 3.

Algorithm 3 Conjugate Gradient - Quadratic Case

Require: Initial point x^0, $k \leftarrow 0$
1: $r_0 \leftarrow -\nabla F(x^0)$
2: $d_0 \leftarrow r_0$
3: **while** $\|r_k\| > \epsilon$ **do**
4: $\quad \theta_k \leftarrow -\frac{d_k^\top r_k}{d_k^\top A d_k}$
5: $\quad x^{k+1} \leftarrow x^k + \theta_k d_k$
6: $\quad r_{k+1} \leftarrow -\nabla F(x^{k+1})$
7: $\quad \beta_k \leftarrow \frac{r_{k+1}^\top A d_k}{d_k^\top A d_k}$
8: $\quad d_{k+1} \leftarrow r_{k+1} + \beta_k d_k$
9: $\quad k \leftarrow k + 1$
10: **end while**

A more practical and less expensive version of this algorithm can be obtained by replacing in the previous algorithm:

- The coefficient β_k by the Fletcher-Reeves formula

$$\beta_k = \frac{r_k^\top r_k}{r_{k-1}^\top r_{k-1}}$$

- The step θ_k by

$$\theta_k = -\frac{r_k^\top r_k}{d_k^\top A d_k}$$

Remark 3.2. These two formulas can be obtained using the property of orthogonality of gradients in the conjugate gradient method.

Case of a general function

The conjugate gradient algorithm for quadratic functions can be generalized to general functions. This gives us the Fletcher-Reeves algorithm presented below:

Algorithm 4 Conjugate Gradient - Fletcher-Reeves

Choose x^0, $k \leftarrow 0$
$d_0 \leftarrow -\nabla f(x^0)$
while $\|\nabla f(x^k)\| > \epsilon$ **do**
    Find optimal θ_k: $\min\limits_{\theta \geq 0} f(x^k + \theta d_k)$
    $x^{k+1} \leftarrow x^k + \theta_k d_k$
    $\beta_k \leftarrow \frac{\|\nabla f(x^{k+1})\|^2}{\|\nabla f(x^k)\|^2}$
    $d_{k+1} \leftarrow -\nabla f(x^{k+1}) + \beta_k d_k$
    $k \leftarrow k + 1$
end while

Chapter 4

Optimization of Nonlinear Functions under Constraints

The objective of this chapter is to minimize a function f on a set X of feasible solutions. This optimization problem is written in the general form:

$$\begin{cases} \min f(x) \\ x \in X \end{cases} \tag{4.1}$$

where $X = \{x \in \mathbb{R}^n / h_i(x) = 0, i = 1, \ldots, k; g_j(x) \leq 0, j = 1, \ldots, m\}$.

When the functions $f,\ h_i,\ i = \overline{1,k}$ and $g_j,\ j = \overline{1,m}$ are linear, the problem (4.1) is a linear programming (LP) problem. In this case, the optimal solution is an extreme point of the polyhedron of feasible solutions. This is no longer true in the nonlinear case where the optimal solution can also be inside or on the boundary of the set X. The following example provides a concrete illustration of this property.

Example 4.1. Let the set X of $\mathbb{R}^2$ defined by

$$X = \{x = (x_1, x_2) \in \mathbb{R}^2 \ / \ x_1 + x_2 \geq 1,\ 2x_1 + 3x_2 \leq 12,\ x_1 \geq 0,\ x_2 \geq 0\}.$$

Consider the minimization problems of the functions $f_1,\ f_2$ and f_3 on the polyhedron X, with

$f_1(x) = (x_1 - 4)^2 + (x_2 - 6)^2$
$f_2(x) = (x_1 - 8)^2 + x_2^2$
$f_3(x) = (x_1 - 4)^2 + (x_2 - 1)^2$

As shown in figure 4.1, the minimum x_1^* of f_1 lies on the boundary of X, the minimum x_2^* of f_2 is an extreme point of X and the minimum x_3^* of f_3 is an interior point of X.

4.1 Necessary optimality conditions

Consider the general form optimization problem (4.1). Before presenting the optimality conditions in this case, let's start by presenting some definitions.

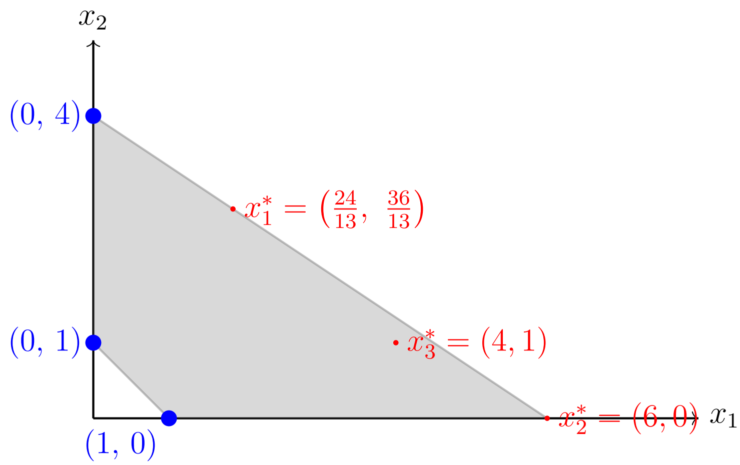


Figure 4.1: Optimal solutions

Definition 4.1. A feasible solution $x^* \in X$ is said to be an optimal solution (minimum point) of problem (4.1), if

$$f(x^*) = \min_{x \in X} f(x), \text{ i.e. } f(x^*) \leq f(x), \ \forall x \in X.$$

Remark 4.1. Such a solution is said to be global, as opposed to local minima which satisfy this definition only in a neighborhood of x^*. Formally, $x^* \in X$ is a local optimal solution of problem (4.1), if

$$\exists \varepsilon > 0, \ f(x^*) \leq f(x), \ \forall x \in X \cap \mathcal{B}(x^*, \varepsilon),$$

where $\mathcal{B}(x^*, \varepsilon) = \{x \in \mathbb{R}^n / \|x - x^*\| \leq \varepsilon\}$.

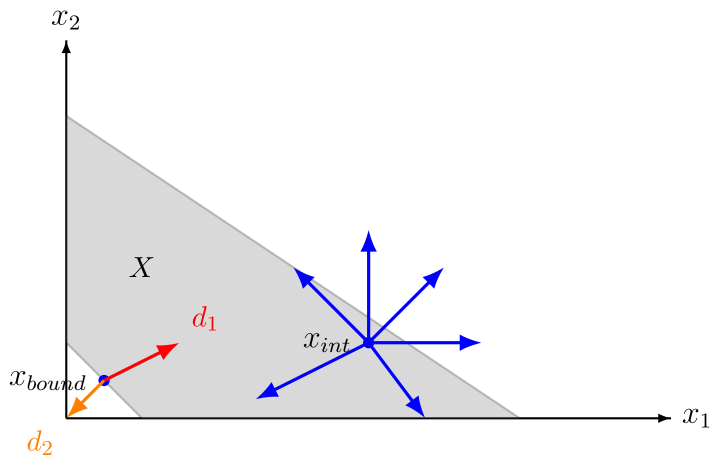
Definition 4.2. Let $x \in X$. A vector $d \in \mathbb{R}^n$ is said to be a feasible direction at x, if there exists a number $\bar{\alpha} > 0$ such that

$$x + \alpha d \in X, \ \forall \alpha \in [0, \bar{\alpha}].$$

If x is an interior point of X, any direction d is feasible.

4.2 Writing the optimality conditions

In unconstrained optimization ($X = \mathbb{R}^n$), we looked for *descent* directions that allowed us to decrease the function f until reaching its minimum. In the framework of constrained optimization, we look for *feasible* directions that allow us to minimize f while remaining within the domain of feasible solutions.



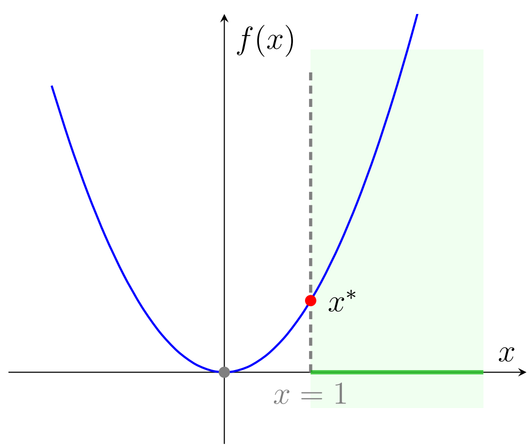
For the point x_{int}, all directions are feasible.
For the point x_{bound}:
- d_1 is a feasible direction.
- d_2 is a non-feasible direction.

Figure 4.2: Feasible directions

Let's now return to optimality conditions and consider the following optimization problem.

Example 4.2.

$$\begin{cases} \min f(x) = x^2, \\ \quad s.t. \quad x \geq 1 \end{cases}$$



The function $f(x) = x^2$ reaches its minimum at $x^* = 1$, with $f(x^*) = 1$. However, the gradient of f does not vanish at $x^* = 1$; $\nabla f(1) = f^{'}(1) = 2 \neq 0$.

Consequently, the condition $\nabla f(x) = 0$ is no longer valid to characterize optimality in the constrained case. However, the formulation of the optimality conditions is based on the same principle through the introduction of a function called the Lagrangian (or Lagrange function). This function is the combination of the objective function f and the constraint functions h_i and g_j. The search for extrema of f on X is then reduced to the search for unconstrained extrema of this same Lagrange function. Theorem 4.1 states a generic necessary optimality condition for any form of the feasible solution set X.

Theorem 4.1. *Let $f : \mathbb{R}^n \to \mathbb{R}$, of class C^1. If x^* is a local minimum point of problem* (4.1)*, then for every feasible direction $d \in \mathbb{R}^n$ at x^*, we have*

$$[\nabla f(x^*)]^\top d \geq 0$$

This condition means that at a minimum point, no direction can decrease the value of the function f while maintaining feasibility (staying in the set X). For constrained problems, this condition is equivalent to the first-order Karush–Kuhn–Tucker (KKT) optimality condition which will be studied in the following sections.

4.3 Optimization of nonlinear functions under equality constraints

Consider the optimization problem under equality constraints:

$$\begin{cases} \min f(x) \\ \quad s.t.\ h(x) = 0, \end{cases} \tag{4.2}$$

where:

$$h : \mathbb{R}^n \to \mathbb{R}^k, \quad h(x) = \begin{pmatrix} h_1(x) \\ \vdots \\ h_k(x) \end{pmatrix}$$

The Lagrange function associated with problem (4.2)

$$\mathcal{L}(x, \mu) = f(x) + \mu^\top h(x) = f(x) + \sum_{i=1}^{k} \mu_i h_i(x)$$

where $\mu = (\mu_1, \ldots, \mu^k)^\top$ is the vector of Lagrange multipliers.

4.3.1 Necessary optimality conditions

Assume that f and $h_i, \quad i = \overline{1, k}$ are of class C^1.

Theorem 4.2 (First order necessary condition)**.** *Let $x^* \in X$ be a minimum point of problem* (4.2) *and assume that*

$$\textit{the vectors } \nabla h_1(x^*), \nabla h_2(x^*), \ldots, \nabla h_k(x^*) \textit{ are linearly independent.} \tag{4.3}$$

Then, there exists a Lagrange multiplier vector μ^* such that:

$$\nabla_x \mathcal{L}(x^*, \mu^*) = \frac{\partial \mathcal{L}}{\partial x}(x^*, \mu^*) = 0 \tag{4.4}$$

$$\nabla_\mu \mathcal{L}(x^*, \mu^*) = \frac{\partial \mathcal{L}}{\partial \mu}(x^*, \mu^*) = 0. \tag{4.5}$$

Remark 4.2.

1. The relations (4.4) are called the Lagrange stationarity conditions.

2. The condition (4.3) is called the constraint qualification condition.

Remark 4.3. The constraint qualification condition (4.3) holds, in particular, if constraints are affine.

Definition 4.3. A vector $x^* \in X$ is a stationary point of problem (4.2), if there exists a vector μ^* such that the pair (x^*, μ^*) satisfies the relations (4.4).

Thus, the search for stationary points of problem (4.2) is reduced to solving the system (4.4).

Theorem 4.3 (Second order necessary condition)**.** *Assume that the functions f and h_i for $i = 1, \ldots, k$ are of class C^2. If $x^* \in X$ is a solution of problem* (4.2) *satisfying the qualification condition* (4.3) *and μ^* is the corresponding Lagrange multiplier vector, then:*

$$d^\top \frac{\partial^2 L}{\partial x^2}(x^*, \mu^*) d \geq 0, \tag{4.6}$$

$$\forall d \in \mathbb{R}^n \text{ such that } d^\top \nabla h_i(x^*) = 0, \ i = \overline{1, k}.$$

Definition 4.4. A solution $x^* \in X$ of problem (4.2) is said to be *regular* if it satisfies the constraint qualification (4.3).

4.3.2 Sufficient optimality condition

Assume that the functions f and h_i for $i = 1, \ldots, k$ are of class C^2.

Theorem 4.4. *A stationary point x^* of problem* (4.2) *is a locally optimal solution, if there exists a Lagrange multiplier vector μ^* such that*

$$d^\top \frac{\partial^2 L}{\partial x^2}(x^*, \mu^*) d > 0, \tag{4.7}$$

$$\forall d \in \mathbb{R}^n, \|d\| \neq 0 \text{ such that } d^\top \nabla h_i(x^*) = 0, \ i = \overline{1, k}.$$

4.4 Optimization of nonlinear functions under inequality constraints

Consider the optimization problem under inequality constraints:

$$\begin{cases} \min f(x) \\ g(x) \leq 0, \end{cases} \tag{4.8}$$

where:

$$g : \mathbb{R}^n \to \mathbb{R}^m, g(x) = \begin{pmatrix} g_1(x) \\ \vdots \\ g_m(x) \end{pmatrix}$$

We will denote $I = \{1, \ldots, m\}$.

Definition 4.5 (Active/passive constraint)**.**
A constraint $g_j(x) \leq 0$ is:

- Active at point x if $g_j(x) = 0$
- Passive at point x if $g_j(x) < 0$

We will denote by $I_a(x) = \{j \in I \mid g_j(x) = 0\}$ the set of indices of active constraints.

Definition 4.6. A vector $d \in \mathbb{R}^n$ is said to be a feasible direction at point x^* with respect to the constraint $g_j(x) \leq 0$, if:

$$\begin{cases} d^\top \nabla g_j(x^*) < 0 & \text{for } g_j(x^*) = 0 \\ d \text{ arbitrary} & \text{for } g_j(x^*) < 0 \end{cases}$$

Definition 4.7. A vector $d \in \mathbb{R}^n$ is a feasible direction at point x^* with respect to the constraints $g_j(x^*) \leq 0,\ j = 1, \ldots, m$ if it is a feasible direction with respect to each of these constraints.

Remark 4.4. To find the feasible directions with respect to the constraints of problem (4.8), it suffices to solve the system:

$$d^\top \nabla g_j(x^*) < 0, \quad \text{for } j \in I_a(x^*)$$

4.4.1 Necessary optimality conditions

Based on the notions of feasible direction and descent direction (see chapter 1) for the function f, we give the following necessary optimality condition.

Theorem 4.5. *Suppose that x^* is an optimal solution of problem* (4.8)*. Then, there exists no vector $d \in \mathbb{R}^n$ satisfying the following system:*

$$\begin{cases} d^\top \nabla f(x^*) < 0, \\ d^\top \nabla g_j(x^*) < 0 \textit{ for all } j \in I_a(x^*) \end{cases}$$

This theorem characterizes optimality using the notions of feasible and descent directions: at the optimum no direction can improve the value of f without violating the (active) constraints. If such a direction existed, it would contradict optimality. This optimality condition is none other than the geometric formulation of the Karush-Kuhn-Tucker (KKT) optimality conditions which we will present below.

Let us first introduce the Lagrange function associated with problem (4.8):

$$\mathcal{L}(x, \lambda) = f(x) + \lambda^\top g(x) = f(x) + \sum_{j=1}^{k} \lambda_j g_j(x)$$

where $0 \leq \lambda = (\lambda_1, \ldots, \lambda_m)^\top$ is the vector of Lagrange multipliers.

Theorem 4.6 (Karush-Kuhn-Tucker necessary conditions)**.** *Let x^* be a minimum of problem* (4.8) *such that the vectors $\nabla g_j(x^*)$ for $j \in I_a(x^*)$ are linearly independent. Then there exists a unique vector λ^* such that:*

$$\begin{aligned} \nabla_x \mathcal{L}(x^*, \lambda^*) &= \nabla f(x^*) + \sum_{j=1}^{k} \lambda_j^* \nabla g_j(x^*) 0 && (4.9) \\ (\lambda^*)^\top g(x^*) &= 0 && (4.10) \\ \lambda^* &\geq 0 && (4.11) \end{aligned}$$

The relations (4.10) are called complementarity conditions and the relation (4.11) is the positivity condition of the Lagrange multipliers.

Definition 4.8. A feasible solution $x^* \in X$ is said to be *regular* if the vectors $\nabla g_j(x^*)$ for $i \in I_a(x^*)$ are linearly independent.

Definition 4.9 (Strong/weak constraint)**.** A constraint $g_j(x^*) \leq 0$ is said to be *strongly* active at point x^* if $\lambda_j^* > 0$ and *weakly* active if $\lambda_j^* = 0$.

Define the set of indices of strongly active constraints at point x^* by

$$I_a^+(x^*) = \{j \in I \mid g_j(x) = 0,\ \lambda^* > 0\},$$

and the set of indices of weakly active constraints by

$$I_a^-(x^*) = \{j \in I \mid g_j(x) = 0,\ \lambda^* = 0\}$$

Theorem 4.7 (Second order necessary condition)**.** *Assume that the functions f and g_j for $j = \overline{1, m}$ are of class C^2. If x^* is a regular minimum point of problem* (4.8) *and λ^* is the corresponding Lagrange multiplier. Then:*

$$y^\top \frac{\partial^2 \mathcal{L}}{\partial x^2}(x^*, \lambda^*) y \geq 0$$

for all $y \in \mathbb{R}^n$ satisfying:

$$\begin{cases} y^\top \nabla g_j(x^*) = 0 & \forall i \in I_a^+(x^*) \\ y^\top \nabla g_j(x^*) \leq 0 & \forall i \in I_a^-(x^*) \end{cases}$$

4.4.2 Sufficient optimality condition

Definition 4.10 (Pseudo-stationary solution)**.** A point $x^* \in X$ is said to be a *pseudo-stationary solution* of problem (4.8) if there exists a Lagrange multiplier vector $\lambda^* \geq 0$ such that

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial x}(x^*, \lambda^*) &= 0 && (4.12) \\ (\lambda^*)^\top g(x^*) &= 0 && (4.13) \end{aligned}$$

Theorem 4.8. *Assume that the functions f and g_j for $j = \overline{1,m}$ are of class C^2. For a pseudo-stationary solution x^* to be locally optimal, it is sufficient that:*

$$y^\top \frac{\partial^2 \mathcal{L}}{\partial x^2}(x^*, \lambda^*) y > 0$$

for all $y \in \mathbb{R}^n, \|y\| \neq 0$ satisfying:

$$\begin{cases} y^\top \nabla g_j(x^*) = 0 & \forall i \in I_a^+(x^*) \\ y^\top \nabla g_j(x^*) \leq 0 & \forall i \in I_a^-(x^*) \end{cases}$$

4.5 Optimization of nonlinear functions under mixed constraints (equality and inequality)

Consider the following optimization problem:

$$\begin{cases} \min f(x) \\ h_i(x) = 0 & i = 1, \dots, k \\ g_j(x) \leq 0 & j = 1, \dots, m \end{cases} \tag{4.14}$$

where:

$$\begin{aligned} f &: \mathbb{R}^n \to \mathbb{R} \\ h_i &: \mathbb{R}^n \to \mathbb{R} \quad (i = 1, \dots, k) \\ g_j &: \mathbb{R}^n \to \mathbb{R} \quad (j = 1, \dots, m) \end{aligned}$$

The associated Lagrange function is defined by:

$$\mathcal{L}(x, \mu, \lambda) = f(x) + \sum_{i=1}^{k} \mu_i h_i(x) + \sum_{j=1}^{m} \lambda_j g_j(x)$$

where $\mu \in \mathbb{R}^k$ and $\lambda \in \mathbb{R}^m$ are the Lagrange multipliers.

4.5.1 Necessary optimality conditions (KKT)

Theorem 4.9. *Let x^* be a local minimum of problem* (4.14)*. Assume that the gradient vectors of the active constraints:*

$$\{\nabla h_i(x^*) \mid i = 1, \dots, k\} \cup \{\nabla g_j(x^*) \mid j \in I_a(x^*)\} \quad \textit{are linearly independent.}$$

Then, there exist Lagrange multipliers $\mu^* \in \mathbb{R}^k$ and $\lambda^* \in \mathbb{R}^m$ such that:

$$\nabla_x \mathcal{L}(x^*, \mu^*, \lambda^*) = 0 \tag{4.15}$$

$$h_i(x^*) = 0 \quad \forall i = 1, \dots, k \tag{4.16}$$

$$\lambda_j^* g_j(x^*) = 0 \quad \forall j = 1, \dots, m \tag{4.17}$$

$$\lambda_j^* \geq 0 \quad \forall j = 1, \dots, m. \tag{4.18}$$

Example 4.3. Consider the following optimization problem:

$$\begin{aligned} &\min x^2 + y^2 + z^2 \\ &\text{subject to} \\ &\quad x + y + z = 3 \\ &\quad 2x - y + z \leq 5 \end{aligned}$$

Since the constraints are affine, we can directly apply the (KKT) conditions (see remark 4.3 for constraint qualification condition): at a minimum (x, y, z), $\exists \mu \in \mathbb{R}$, $\lambda \in \mathbb{R}_+$ such that $\nabla f(x, y, z) + \mu \nabla h(x, y, z) + \lambda \nabla g(x, y, z) = 0$:

$$\begin{aligned} (i) &\quad \begin{cases} 2x + \mu + 2\lambda = 0 \\ 2y + \mu - \lambda = 0 \\ 2z + \mu + \lambda = 0 \end{cases} \\ (ii) &\quad x + y + z = 3 \\ (iii) &\quad \lambda(2x - y + z - 5) = 0 \\ (iv) &\quad \lambda \geq 0 \end{aligned}$$

Case 1: Assume that $\lambda \neq 0 \implies 2x - y + z = 5$

$$(i) \implies \begin{cases} x = (-\mu - 2\lambda)/2 \\ y = (-\mu + \lambda)/2 \\ z = (-\mu - \lambda)/2 \end{cases}$$

$$(ii) \implies -\mu - 2\lambda - \mu + \lambda - \mu - \lambda = 6 \implies 3\mu + 2\lambda = -6 \quad \text{(a)}$$

$$2x - y + z = 5 \implies 2(-\mu - 2\lambda) - (-\mu + \lambda) + (-\mu - \lambda) = 10 \implies \mu + 3\lambda = -5 \quad \text{(b)}$$

By computing (a) - 3(b) we obtain $2\lambda - 9\lambda = -6 + 15$, i.e., $\lambda = -9/7 < 0$ which contradicts (iv)

Case 2: $\lambda = 0$. Therefore (i) becomes:

$$\begin{cases} 2x + \mu = 0 \\ 2y + \mu = 0 \\ 2z + \mu = 0 \end{cases} \implies 2(x + y + z) + 3\mu = 0 \implies 2 \times 3 + 3\mu = 0 \implies \mu = -2$$

$$\implies x = y = z = 1.$$

We obtain the stationary point $(x, y, z) = (1, 1, 1)$.

4.5.2 Sufficient optimality condition

Theorem 4.10. *Assume that $f, h_i, g_j \in C^2$. If for $x^* \in X$ there exist multipliers $\mu^* \in \mathbb{R}^k$ and $\lambda^* \in \mathbb{R}^m$ satisfying the system* (4.15) *and for all $y \in \mathbb{R}^n$, $\|y\| \neq 0$, such that:*

$$\begin{cases} y^\top \nabla h_i(x^*) = 0 & \forall i \in I_a^+(x^*), \\ y^\top \nabla h_i(x^*) \leq 0 & \forall i \in I_a^-(x^*), \\ y^\top \nabla g_j(x^*) = 0 & \forall j = 1, \ldots, m, \end{cases}$$

we have:

$$y^\top \frac{\partial^2 \mathcal{L}}{\partial x^2}(x^*, \mu^*, \lambda^*) y > 0,$$

then x^* is a local minimum of problem (4.14)*.*

4.5.3 Convex Optimization

We refer to convex optimization if, in problem (4.14), we have:

1. f convex,
2. h_i affine $\forall i = 1, \ldots, k$,
3. g_j convex $\forall j = 1, \ldots, m$.

In this case, the KKT conditions are also sufficient for optimality. Furthermore, any solution to the problem is a global minimum.

Theorem 4.11. *(KKT, necessary and sufficient condition) If in problem* (4.14)*, we have*

1. *f is convex and differentiable,*
2. *h_i, $\forall i = 1, \ldots, k$ are affine,*
3. *g_j, $\forall j = 1, \ldots, m$ are convex and differentiable,*

Then a necessary and sufficient condition for $x^* \in X$ to be a global minimum of the problem is that there exist Lagrange multipliers $\mu^* \in \mathbb{R}^k$ and $\lambda^* \in \mathbb{R}^m$ such that

$$\nabla f(x^*) + \mu^* \nabla h(x^*) + \lambda^* \nabla g(x^*) = 0.$$

Example 4.4. In example 4.3, the studied problem is convex. Indeed, we have

1. The objective function is strictly convex, since its Hessian defined by

$$\nabla^2 f = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix}$$

is positive definite on $\mathbb{R}^3$.
2. The feasible set $X = \{(x, y, z) \mid x + y + z = 3, 2x - y + z \leq 5\}$ is also convex, because:

- The equality constraint $h(x, y, z) = x + y + z - 3 = 0$ defines an affine set (hence convex)
- The inequality constraint $g(x, y, z) = 2x - y + z - 5 \leq 0$ defines a half-space (convex set)
- The intersection of convex sets is convex.

Therefore, according to theorem 4.11, the stationary point found previously is a strict global minimum

Chapter 5

Exercises

5.1 Exercises with solutions

Exercise 5.1. Let the set $C = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 16\}$.
Show that C is convex using
a) the definition. b) the properties of convex functions.

Solution 5.1. a) Using the definition of convex sets

A set C is convex if:

$$\forall P_1, P_2 \in C,\ \forall \lambda \in [0, 1], \quad P = \lambda P_1 + (1 - \lambda)P_2 \in C.$$

Let $P_1 = (x_1, y_1)$, $P_2 = (x_2, y_2) \in C$ and $\lambda \in [0, 1]$.
We have:

$$P = (u, v) = \lambda(x_1, y_1) + (1 - \lambda)(x_2, y_2)$$
$$\Rightarrow \begin{cases} u = \lambda x_1 + (1 - \lambda)x_2 \\ v = \lambda y_1 + (1 - \lambda)y_2 \end{cases}$$

Hence

$$\begin{aligned} u^2 + v^2 &= [\lambda x_1 + (1 - \lambda)x_2]^2 + [\lambda y_1 + (1 - \lambda)y_2]^2 \\ &= \lambda^2 x_1^2 + (1 - \lambda)^2 x_2^2 + 2\lambda(1 - \lambda)x_1 x_2 + \lambda^2 y_1^2 + (1 - \lambda)^2 y_2^2 + 2\lambda(1 - \lambda)y_1 y_2 \end{aligned}$$

Using $(x_1 - x_2)^2 \geq 0 \Rightarrow 2x_1x_2 \leq x_1^2 + x_2^2$ and similarly $(y_1 - y_2)^2 \geq 0 \Rightarrow 2y_1y_2 \leq y_1^2 + y_2^2$, we get:

$$u^2 + v^2 \leq \lambda^2 x_1^2 + (1 - \lambda)^2 x_2^2 + \lambda(1 - \lambda)(x_1^2 + x_2^2) + \lambda^2 y_1^2 + (1 - \lambda)^2 y_2^2 + \lambda(1 - \lambda)(y_1^2 + y_2^2)$$

Simplifying the expression yields:

$$u^2 + v^2 \leq \lambda(x_1^2 + y_1^2) + (1 - \lambda)(x_2^2 + y_2^2)$$

Since $(x_1, y_1), (x_2, y_2) \in X$, we have $x_1^2 + y_1^2 \leq 16$ and $x_2^2 + y_2^2 \leq 16$. Therefore:

$$u^2 + v^2 \leq \lambda \cdot 16 + (1 - \lambda) \cdot 16 = 16$$

$$\Rightarrow u^2 + v^2 \leq 16 \Rightarrow (u, v) \in C$$

It follows that C is convex.

b) Using properties of convex functions

C is convex if it represents the level set of a convex function (see proposition 1.7). We can write $C = \{(x, y) \in \mathbb{R}^2 : f(x, y) \leq 16\}$. Then $C = \mathbf{L}_{16}(f)$ is the level set of the function defined on $\mathbb{R}^2$ by $f(x, y) = x^2 + y^2$. Furthermore, the Hessian matrix of f $\nabla^2 f(x, y) = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$ is positive definite on $\mathbb{R}^2$. Consequently, C is a convex set.

Exercise 5.2. Study the convexity of $f(x) = \frac{1}{x},\ x > 0$.

Solution 5.2. We want to determine whether f is convex on $\mathbb{R}_+^*$, i.e., whether:

$$\forall x, y \in \mathbb{R}_+^*,\ \forall \lambda \in [0, 1], \quad f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda) f(y)$$

Let $x, y \in \mathbb{R}_+^*$, $\lambda \in [0, 1]$. Consider the difference:

$$\begin{aligned}
\alpha &= f(\lambda x + (1 - \lambda)y) - \lambda f(x) - (1 - \lambda) f(y) \\
&= \frac{1}{\lambda x + (1 - \lambda)y} - \frac{\lambda}{x} - \frac{1 - \lambda}{y} \\
&= \frac{xy - [\lambda x + (1 - \lambda)y][\lambda y + (1 - \lambda)x]}{xy(\lambda x + (1 - \lambda)y)} \\
&= \frac{xy - [\lambda^2 xy + \lambda(1 - \lambda)x^2 + \lambda(1 - \lambda)y^2 + (1 - \lambda)^2 xy]}{xy(\lambda x + (1 - \lambda)y)} \\
&= \frac{xy - \lambda^2 xy - \lambda(1 - \lambda)x^2 - \lambda(1 - \lambda)y^2 - (1 - \lambda)^2 xy}{xy(\lambda x + (1 - \lambda)y)} \\
&= \frac{[1 - \lambda^2 - (1 - \lambda)^2]xy - \lambda(1 - \lambda)(x^2 + y^2)}{xy(\lambda x + (1 - \lambda)y)} \\
&= \frac{2\lambda(1 - \lambda)xy - \lambda(1 - \lambda)(x^2 + y^2)}{xy(\lambda x + (1 - \lambda)y}
\end{aligned}$$

Thus:

$$\alpha = \frac{-\lambda(1 - \lambda)(x - y)^2}{xy(\lambda x + (1 - \lambda)y)}$$

For $x, y > 0,\ x \neq y$, $\lambda \in]0, 1[$, we have:

- $\lambda(1 - \lambda) > 0$

- $(x-y)^2 > 0$
- $xy > 0$
- $\lambda x + (1-\lambda)y > 0$

Therefore $\alpha < 0$,
We conclude that for $x, y > 0,\ x \neq y,\ \lambda \in]0,1[$:

$$f(\lambda x + (1-\lambda)y) < \lambda f(x) + (1-\lambda)f(y)$$

The function $f(x) = \frac{1}{x}$ is strictly convex on $\mathbb{R}_+^*$.

Exercise 5.3. Let $f(x) = x_1^2 + 2x_2^2 - 2x_1x_2 + e^{x_1+x_2}$

(a) Compute the gradient $\nabla f(x)$

(b) Compute the Hessian matrix $\nabla^2 f(x)$

(c) Verify if $\nabla^2 f(0,0)$ is positive definite.

Solution 5.3. (a) The gradient is computed as:

$$\frac{\partial f}{\partial x_1} = 2x_1 - 2x_2 + e^{x_1+x_2}$$
$$\frac{\partial f}{\partial x_2} = 4x_2 - 2x_1 + e^{x_1+x_2}$$
$$\nabla f(x) = \begin{pmatrix} 2x_1 - 2x_2 + e^{x_1+x_2} \\ 4x_2 - 2x_1 + e^{x_1+x_2} \end{pmatrix}$$

(b) The Hessian matrix is:

$$\frac{\partial^2 f}{\partial x_1^2} = 2 + e^{x_1+x_2}$$
$$\frac{\partial^2 f}{\partial x_2^2} = 4 + e^{x_1+x_2}$$
$$\frac{\partial^2 f}{\partial x_1 \partial x_2} = \frac{\partial^2 f}{\partial x_2 \partial x_1} = -2 + e^{x_1+x_2}$$
$$\nabla^2 f(x) = \begin{pmatrix} 2 + e^{x_1+x_2} & -2 + e^{x_1+x_2} \\ -2 + e^{x_1+x_2} & 4 + e^{x_1+x_2} \end{pmatrix}$$

(c) At point (0,0):

$$\nabla^2 f(0,0) = \begin{pmatrix} 2+1 & -2+1 \\ -2+1 & 4+1 \end{pmatrix} = \begin{pmatrix} 3 & -1 \\ -1 & 5 \end{pmatrix}$$

Check Sylvester's criterion:

- First leading principal minor: $\Delta_1 = 3 > 0$
- Second leading principal minor: $\Delta_2 = \begin{vmatrix} 3 & -1 \\ -1 & 5 \end{vmatrix} = 15 - 1 = 14 > 0$

Since all leading principal minors are positive, $\nabla^2 f(0,0)$ is positive definite.

Exercise 5.4. Let $f(x) = x_1^2 + x_2^2 + x_3^2 + x_1x_2 + x_2x_3$

(a) Find all stationary points.

(b) Determine the nature of these points.

(c) Does the function have global extrema?

Solution 5.4. (a) Compute the gradient:

$$\nabla f(x) = \begin{pmatrix} 2x_1 + x_2 \\ 2x_2 + x_1 + x_3 \\ 2x_3 + x_2 \end{pmatrix}$$

Set $\nabla f(x) = 0$:

$$2x_1 + x_2 = 0 \tag{5.1}$$
$$x_1 + 2x_2 + x_3 = 0 \tag{5.2}$$
$$x_2 + 2x_3 = 0 \tag{5.3}$$

From (5.1): $x_2 = -2x_1$
From (5.3): $2x_3 = -x_2 = 2x_1 \Rightarrow x_3 = x_1$
Substitute in (5.2): $x_1 + 2(-2x_1) + x_1 = x_1 - 4x_1 + x_1 = -2x_1 = 0 \Rightarrow x_1 = 0$
Therefore: $x_2 = 0$, $x_3 = 0$

The only stationary point is $x^* = (0, 0, 0)$.

(b) Compute the Hessian matrix:

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

Check leading principal minors:

- $\Delta_1 = 2 > 0$
- $\Delta_2 = \begin{vmatrix} 2 & 1 \\ 1 & 2 \end{vmatrix} = 4 - 1 = 3 > 0$

- $\Delta_3 = \begin{vmatrix} 2 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 2 \end{vmatrix} = 2(4-1) - 1(2-0) + 0 = 6 - 2 = 4 > 0$

  All leading principal minors are positive, so the Hessian is positive definite. Therefore, $x^* = (0, 0, 0)$ is a strict local minimum.

(c) Since the Hessian matrix is constant and positive definite everywhere, the function is strictly convex on $\mathbb{R}^3$. Thus, $x^* = (0, 0, 0)$ is a global minimum.

Exercise 5.5. Apply 3 iterations of the gradient method with optimal step size to minimize:

$$f(x) = x_1^2 + 2x_2^2$$

starting from point $x^0 = (2, 1)$. Show all calculations.

Solution 5.5. $\nabla f(x) = (2x_1, 4x_2)^\top$

Iteration 1: $x^0 = (2, 1)^\top$, $d_0 = -\nabla f(x^0) = (-4, -4)^\top$
$x^1 = x^0 + \theta_0 d_0 = (2 - 4\theta_0, 1 - 4\theta_0)^\top$
$\phi(\theta) = f(x^0 + \theta d_0) = (2 - 4\theta)^2 + 2(1 - 4\theta)^2 = 6 - 32\theta + 48\theta^2$
$\phi'(\theta) = -32 + 96\theta = 0 \Rightarrow -32 + 96\theta_0 = 0 \Rightarrow \theta_0 = \frac{1}{3}$
$x^1 = x^0 + \theta_0 d_0 = (\frac{2}{3}, -\frac{1}{3})^\top$.

Iteration 2: $x^1 = (\frac{2}{3}, -\frac{1}{3})^\top$, $d_1 = -\nabla f(x^1) = (-\frac{4}{3}, \frac{4}{3})^\top$
$x^2 = x^1 + \theta_1 d_1 = (\frac{2}{3} - \frac{4}{3}\theta_1, -\frac{1}{3} + \frac{4}{3}\theta_1)^\top$
$\phi(\theta) = f(x^1 + \theta d_1) = (\frac{2}{3} - \frac{4}{3}\theta)^2 + 2(-\frac{1}{3} + \frac{4}{3}\theta)^2 = \frac{2}{3} - \frac{32}{9}\theta + \frac{16}{3}\theta^2$
$\phi'(\theta) = -\frac{32}{9} + \frac{96}{9}\theta = 0 \Rightarrow -\frac{32}{9} + \frac{96}{9}\theta_1 = 0 \Rightarrow \theta_1 = \frac{1}{3}$
$x^2 = x^1 + \theta_1 d_1 = (\frac{2}{9}, \frac{1}{9})^\top$.

Iteration 3: $x^2 = (\frac{2}{9}, \frac{1}{9})^\top$, $d_2 = -\nabla f(x^2) = (-\frac{4}{9}, -\frac{4}{9})^\top$
$x^3 = x^2 + \theta_2 d_2 = (\frac{2}{9} - \frac{4}{9}\theta_2, \frac{1}{9} - \frac{4}{9}\theta_2)^\top$
$\phi(\theta) = f(x^2 + \theta d_2) = (\frac{2}{9} - \frac{4}{9}\theta)^2 + 2(\frac{1}{9} - \frac{4}{9}\theta)^2 = \frac{2}{27} - \frac{32}{81}\theta + \frac{16}{27}\theta^2$
$\phi'(\theta) = -\frac{32}{81} + \frac{96}{81}\theta = 0 \Rightarrow -\frac{32}{81} + \frac{96}{81}\theta_2 = 0 \Rightarrow \theta_2 = \frac{1}{3}$
$x^3 = x^2 + \theta_2 d_2 = (\frac{2}{27}, \frac{1}{27})^\top$.

Note that for this convex function, the global mimimum is $x^* = (0, 0)$ with $f(x^*) = 0$.

Exercise 5.6. Consider the minimization problem of the quadratic form

$$F(x) = x_1^2 + 4x_2^2 - 4x_1 - 8x_2 \longrightarrow \min$$

1. Write the first-order optimality conditions. Are these conditions sufficient?
2. Solve these conditions and find the minimum x^* of F on $\mathbb{R}^2$
3. Find the formula of the optimal step size for minimizing a quadratic form.
4. Apply the gradient algorithm with optimal step size to minimize F, starting from the initial approximation $x^0 = (1, 1)^\top$.

Solution 5.6.
1. First-order optimality conditions:

The gradient is:

$$\nabla F(x) = \begin{pmatrix} 2x_1 - 4 \\ 8x_2 - 8 \end{pmatrix}.$$

$$\nabla F(x) = 0 \Rightarrow \begin{cases} l2x_1 - 4 = 0 & \Rightarrow \quad x_1 = 2, \\ 8x_2 - 8 = 0 & \Rightarrow \quad x_2 = 1. \end{cases} .$$

So the unique stationary point is $x^* = (2, 1)^\top$.
For sufficiency, the Hessian matrix:

$$\nabla^2 F(x) = \begin{pmatrix} 2 & 0 \\ 0 & 8 \end{pmatrix},$$

is positive definite (Leading principal minors: $\Delta_1 > 0$ and $\Delta_2 > 0$).

Hence, F is strictly convex, so the first-order conditions are sufficient for a global minimum.

2. Solution of the minimization problem
From question 1., we have: $x^* = (2, 1)^\top$ with $F^* = -8$.

3. Optimal step size for a quadratic form
Let $F(x) = \frac{1}{2} x^\top D x - C^\top x$, with

$$A = \begin{pmatrix} 2 & 0 \\ 0 & 8 \end{pmatrix}, \quad b = \begin{pmatrix} 4 \\ 8 \end{pmatrix}.$$

Let x^k be given and the direction $d_k = -\nabla F(x^k)$, then the next point is given by:

$$x^{k+1} = x^k + \theta_k d_k,$$

where θ_k is solution of the optimization problem:

$$\min_{\theta \geq 0} F(x^k + \theta d_k)$$

Minimize F with respect to $\theta \Rightarrow \theta$ cancels its derivative.

Therefore:

$$\frac{\partial F(x^k+\theta d_k)}{\partial \theta} = d_k^\top \cdot \nabla F(x^k+\theta d_k) = 0$$

$$\Rightarrow d_k^\top[D(x^k+\theta d_k)+C] = 0$$

$$\Rightarrow d_k^\top D x^k + \theta d_k^\top D d_k + d_k^\top C = 0$$

$$\Rightarrow \theta = -\frac{d_k^\top D x^k + d_k^\top C}{d_k^\top D d_k}$$

$$\Rightarrow \theta_k = -\frac{d_k^\top [D x^k + C]}{d_k^\top D d_k}$$

$$\Rightarrow \theta_k = \frac{-d_k^\top \nabla F(x^k)}{d_k^\top D d_k}$$

4. Gradient with optimal step size for minimizing F:

$$x^0 = (1,1)^T, \quad d_0 = -\nabla F(x^0) = \begin{pmatrix} 2 \\ 0 \end{pmatrix}$$

$$\theta_0 = -\frac{d_0^T \nabla F(x^0)}{d_0^T D d_0} = -\frac{(2,0)\begin{pmatrix} 2 \\ 0 \end{pmatrix}}{(2,0)\begin{pmatrix} 2 & 0 \\ 0 & 8 \end{pmatrix}\begin{pmatrix} 2 \\ 0 \end{pmatrix}} = \frac{4}{8} = \frac{1}{2}$$

$$x^1 = x^0 + \theta_0 d_0 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \frac{1}{2}\begin{pmatrix} 2 \\ 0 \end{pmatrix} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$$

$$\nabla F(x^1) = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \Rightarrow \text{stop}$$

$x^* = (2,1)^T$ is the unique global minimum of F. Convergence in one iteration.

Exercise 5.7. Consider the optimization problem:

$$\begin{cases} \min f(x) & = -3x_1^2 x_2 \\ \text{s.c. } \; h(x) & = 6x_1^2 + 6x_1 x_2 - 12 = 0 \end{cases}$$

1. Prove that the constraints qualification holds.
2. Write the Lagrange conditions.
3. Find the solution of the problem.

Solution 5.7. 1. Qualification of constraints:
We have:

$$\nabla h(x) = \begin{pmatrix} 12x_1 + 6x_2 \\ 6x_1 \end{pmatrix}$$

The constraint is qualified if $\nabla h(x) \neq 0$ (a vector is free if it is not null).

$$\nabla h(x) = 0 \Rightarrow \begin{pmatrix} 12x_1 + 6x_2 \\ 6x_1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \Rightarrow x_1 = x_2 = 0$$

Therefore $\nabla h = 0$ for $x = (0,0)$. However, $(0,0) \notin X = \{x \in \mathbb{R}^2 \mid 6x_1^2 + 6x_1x_2 = 12\}$
Therefore the constraints qualification holds for all $x \in X$, and the Lagrange conditions are necessary for optimality.

2. The Lagrange function:

$$L(x, \mu) = -3x_1^2x_2 + \mu(6x_1^2 + 6x_2x_1 - 12)$$

$$\frac{\partial L}{\partial x_1} = -12x_1x_2 + 12\mu x_1 + 6\mu x_2 = 0 \tag{5.4}$$

$$\frac{\partial L}{\partial x_2} = -3x_1^2 + 6\mu x_1 = 0 \tag{5.5}$$

$$\frac{\partial L}{\partial \mu} = 6x_1^2 + 6x_1x_2 - 12 = 0 \tag{5.6}$$

where $\mu \in \mathbb{R}$.
From (5.5): $x_1 = 2\mu$ and from (5.6): $\mu = \pm 1$.

$$\begin{cases} \mu = 1 \Rightarrow x_1 = 2,\ x_2 = 4 \\ \mu = -1 \Rightarrow x_1 = -2,\ x_2 = -4 \end{cases}$$

Therefore the stationary points are: $(x_1^*, x_2^*, \mu^*) \in \{(2,4,1), (-2,-4,-1)\}$.
The Hessian of Lagrange function

$$\frac{\partial^2 L}{\partial x^2} = \begin{pmatrix} -6x_2 + 12\mu & -6x_1 + 6\mu \\ -6x_1 + 6\mu & 0 \end{pmatrix}$$

- $\frac{\partial^2 L}{\partial x^2}(2,4,1) = \begin{pmatrix} -12 & -6 \\ -6 & 0 \end{pmatrix}$, $\Delta_1 = -12$ and $\Delta_2 = -36$.

Therefore the Hessian at point $(2,4,1)$ is not defined on $\mathbb{R}^2$.
We then search for its nature on the set defined by:

$$H = \left\{ d \in \mathbb{R}^2 \mid d^\top \nabla h(x^*) = 0 \right\}$$

We have: $\nabla h(2,4) = \begin{pmatrix} 48 \\ 12 \end{pmatrix}$. Then

$$\begin{aligned} H &= \left\{ (d_1, d_2) \in \mathbb{R}^2 \mid (d_1, d_2)^\top \begin{pmatrix} 48 \\ 12 \end{pmatrix} = 0 \right\} \\ &= \left\{ d = \alpha(1, -4), \alpha \in \mathbb{R}^+ \right\} \end{aligned}$$

Let $d \in H$, $\|d\| \neq 0$, $d = \alpha(1, -4)$, $\alpha \in \mathbb{R}^+$:

$$d^\top \frac{\partial^2 L}{\partial x^2}(2,4,1) \cdot d = \alpha(1,-4) \begin{pmatrix} -12 & -6 \\ -6 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ -4 \end{pmatrix} = \alpha^2(12,-6) \begin{pmatrix} 1 \\ -4 \end{pmatrix} = 36\alpha^2 > 0, \forall \alpha \in \mathbb{R}^*$$

Therefore $d^\top \frac{\partial^2 L}{\partial x^2}(2,4,1) d > 0 \Rightarrow \frac{\partial^2 L}{\partial x^2}(2,4,1)$ is positive definite.

We then conclude that $(2,4)$ is a local minimum.

Likewise, we conclude that the point $(-2,-4)$ is a local maximum.

Exercise 5.8. Solve the following optimization problem:

$$\begin{cases} \min & f(x) = -3x_1 + x_2 - x_3^2 \\ \text{s.t.} & g(x) = x_1 + x_2 + x_3 \leq 0 \\ & h(x) = -x_1 + 2x_2 + x_3^2 = 0 \end{cases}$$

Solution 5.8. The Lagrangian function:

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

KKT system:

First-order conditions:

$$\begin{cases} \frac{\partial L}{\partial x_1} = -3 + \lambda - \mu = 0 \\ \frac{\partial L}{\partial x_2} = 1 + \lambda + 2\mu = 0 \\ \frac{\partial L}{\partial x_3} = -2x_3 + \lambda + 2\mu x_3 = 0 \end{cases}$$

Complementary slackness and positivity constraints:

$$\begin{cases} \lambda(x_1 + x_2 + x_3) = 0 \\ -x_1 + 2x_2 + x_3^2 = 0 \\ \lambda \geq 0 \end{cases}$$

Solving the system:

Case 1: $\lambda = 0$, then from the first two equations:

$$\begin{cases} \lambda = 3 \\ \lambda = -1 \end{cases}$$

Contradiction.

Case 2: $\lambda > 0$, then from the first two equations:

$$\begin{cases} \lambda - \mu = 3 \\ \lambda + 2\mu = -1 \end{cases} \Rightarrow \begin{cases} \lambda = \frac{5}{3} \\ \mu = -\frac{4}{3} \end{cases}$$

Since $\lambda = \frac{5}{3} > 0$, the constraint $g(x)$ is active:

$$x_1 + x_2 + x_3 = 0$$

From the third derivative:

$$-2x_3 + \frac{5}{3} + 2\left(-\frac{4}{3}\right) x_3 = 0 \Rightarrow -2x_3 + \frac{5}{3} - \frac{8}{3}x_3 = 0 \Rightarrow -\frac{14}{3}x_3 + \frac{5}{3} = 0 \Rightarrow x_3 = \frac{5}{14}$$

Using the equality constraint:

$$-x_1 + 2x_2 + \left(\frac{5}{14}\right)^2 = 0 \Rightarrow -x_1 + 2x_2 + \frac{25}{196} = 0$$

Using the active inequality constraint:

$$x_1 + x_2 + \frac{5}{14} = 0$$

Solving the system:

$$\begin{cases} x_1 + x_2 = -\frac{5}{14} \\ -x_1 + 2x_2 = -\frac{25}{196} \end{cases} \Rightarrow \begin{cases} x_1 = -\frac{115}{588} \\ x_2 = -\frac{95}{588} \end{cases}$$

$$x_1 = -\frac{115}{588}, \quad x_2 = -\frac{95}{588}, \quad x_3 = \frac{5}{14}, \quad \lambda = \frac{5}{3}, \quad \mu = -\frac{4}{3}$$

Hence, the unique stationary point $x^* = \left(-\frac{115}{588}, \frac{-95}{588}, \frac{5}{14}\right)$, with $(\lambda^*, \mu^*) = (\frac{5}{3}, \frac{4}{3})$
The Hessian Matrix of the Lagrangian

$$\frac{\partial^2 L}{\partial x^2}(x^*, \lambda^*, \mu^*) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -2 + 2\mu^* \end{pmatrix}$$

Substituting $\mu * = -4/3$,

$$\frac{\partial^2 L}{\partial x^2}(x^*, \lambda^*, \mu^*) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -\frac{14}{3} \end{pmatrix}$$

Following Sylvester criterion, $\frac{\partial^2 L}{\partial x^2}(x^*, \lambda^*, \mu^*) \geq 0$, on $\mathbb{R}^3$
We then chek its nature on the Hyperplane

$$H = \{y \in \mathbb{R}^3 \mid y^\top \nabla h(x^*) = 0;\; y^\top \nabla g(x^*) = 0 (\text{g active})\}$$

We have $g(x^*) = 0$ and $\lambda^* > 0$, then g is strongly active.

We calculate

$$\nabla g(x^*) = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \text{ and } \nabla h(x^*) = \begin{pmatrix} -1 \\ 2 \\ 5/7 \end{pmatrix}$$

Then

$$H = \left\{ y = \alpha(-3/7, -4/7, 1)^\top, \alpha \in \mathbb{R} \right\}$$

Study of the Quadratic Form on H

$$y^\top \frac{\partial^2 L}{\partial x^2}(x^*, \lambda^*, \mu^*) y = \alpha^2 \begin{pmatrix} -\frac{3}{7} & -\frac{4}{7} & 1 \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -\frac{14}{3} \end{pmatrix} \begin{pmatrix} -\frac{3}{7} \\ -\frac{4}{7} \\ 1 \end{pmatrix}$$

$$= \alpha^2 \left(1 \cdot \left(-\frac{14}{3} \right) \cdot 1 \right) = \alpha^2 \left(-\frac{14}{3} \right) < 0 \quad (\text{for } \alpha \neq 0)$$

Hence, x^* is a local maximum of the problem.

5.2 Exercices without solutions

Exercise 5.9. Let m convex functions $f_i(x)$, $i = \overline{1,m}$ defined on a convex set $X \subset \mathbb{R}^n$. Show that the following functions are convex:

$$f(x) = \sum_{i=1}^{m} \alpha_i f_i(x), \quad \alpha_i \geq 0, \quad i = \overline{1,m}$$

$$g(x) = \max_{1 \leq i \leq m} f_i(x)$$

Exercise 5.10. Find the extrema of the following functions

- $f(x) = 2x_1^3 - 3x_2^3 - 6x_1 + 36x_2$
- $g(x) = x_1^2 + x_2^2 + x_3^2 - 2x_1x_2x_3$
- $k(x) = x_1^4 + x_2^4 - 2x_1^2 - 2x_2^2 - 4x_1x_2$.

Exercise 5.11. Consider the following quadratic form

$$F(x) = F(x_1, x_2, x_3) = -x_1^2 - x_2^2 - x_3^2 - 2x_1x_3$$

1. Write F in the form $\frac{1}{2}x^TQx + q^Tx$.
2. What is the nature of the matrix Q?
3. Is the function F convex?
4. Find the stationary points of F.
5. What is the nature of these stationary points?

Exercise 5.12. Consider the function $f : \mathbb{R}^3 \to \mathbb{R}$ defined by

$$f(x, y, z) = e^{x-y} + e^{y-x} + z^2.$$

1. Determine whether f is convex on $\mathbb{R}^3$.
2. Find all critical points of f.
3. Deduce the minimum.
4. Is this minimum global? Justify your answer.

Exercise 5.13. Use the Fletcher-Reeves conjugate gradient algorithm to minimize the function

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

on $\mathbb{R}^2$ starting from $x^0 = (1, 2)^t$.

Exercise 5.14. Consider the problem:

$$\begin{cases} \min f(x) = x_1^2 + x_2^2 \\ \text{s.t. } x_1^2 + x_2^2 - 1 = 0 \end{cases}$$

Find the stationary points and determine their nature.

Exercise 5.15. Solve the following optimization problem:

$$\begin{cases} \min f(x) = x_1^2 + x_2^2 \\ \text{s.t. } x_1 + x_2 \geq 1 \\ x_1, x_2 \geq 0, \end{cases}$$

using KKT conditions.

Exercise 5.16. Solve the problem:

$$\begin{cases} \min f(x) = (x_1 - 2)^2 + (x_2 - 1)^2 \\ \text{s.t. } x_1^2 - x_2 \leq 0 \\ x_1 + x_2 \leq 2 \end{cases}$$

Bibliography

[1] M. S. Bazaraa, H. D. Sherali, and C. M. Shetty. *Nonlinear Programming: Theory and Algorithms*. Wiley, 2006.

[2] D. P. Bertsekas. *Nonlinear Programming*. Athena Scientific, 1999.

[3] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

[4] R. Fletcher. *Practical Methods of Optimization*. 2000.

[5] D. G. Luenberger and Y. Ye. *Linear and Nonlinear Programming*. Springer, 2008.

[6] O.L. Mangasarian. *Nonlinear programming*. SIAM, 1994.

[7] J. Nocedal and S. J. Wright. *Numerical Optimization*. Springer, 2006.

[8] M.S. Radjef. *A course on mathematical programming*. University A.Mira of Bejaia, Algeria, 2004-2005.