Université de Ain Témouchent — Belhadj Bouchaib
Faculté des Sciences et de la Technologie
Département de Mathématiques et Informatiques

Polycopié pédagogique

<u>Titre</u>

Optimisation sans contraintes
Cours et Exercices corrigés

(Unconstrained Optimization: Lessons and Solved Exercises)

Cours destiné aux étudiants de

L3 Mathématiques

Rédigé par :

M. Bentout Soufiane

Année universitaire : 2024-2025

Contents

Introduction

This document is the result of teaching this subject at Ain Temouchent University, Department of Mathematics and Computer Science. It is intended **for third-year Mathematics LMD students**.

Optimization: to make something optimal, to give something the best possible conditions for use, functioning, or performance under certain circumstances.

Optimization is a fundamental branch of both mathematics and computer science, aimed at modeling, analyzing, and solving problems that determine the optimal solution while adhering to specific constraints. Whether in daily life—when organizing a desk or arranging furniture—or in more complex industrial settings like task scheduling, optimization problems are ubiquitous. These challenges can often be expressed as a "general optimization problem see for examples [1, 2, 3].

Optimization plays a key role in several disciplines, including operations research, which lies at the intersection of computer science, mathematics, and economics . It is also essential in applied mathematics, crucial to industry and engineering, and has important applications in numerical analysis, where it helps in problems like maximum likelihood estimation in statistics, and in game theory, control theory, and command systems .

Various methods are used to tackle optimization problems, including unconstrained optimization techniques, which are a central topic in many academic courses (see [4, 7, 5]. These methods help researchers and practitioners solve complex mathematical problems and design efficient algorithms.

An optimization problem consists of finding an element $x^* \in D$ (if it exists) for which $f(x^*)$ is the smallest (or largest, respectively) value of f over D, and we write:

$$\min_{x \in D} f(x) = f(x^*), \quad \text{(respectively, } \max_{x \in D} f(x) = f(x^*)).$$

The primary goal of this course is to build a strong foundation in optimization principles, including the understanding of objective functions, constraints, and feasible regions. Additionally, it aims to clarify the distinctions between various types of optimization problems, such as linear versus nonlinear and convex versus non-convex, enabling students to approach and solve these problems with a clear conceptual framework.

We begin with basic reminders of differential calculus and the notion of convexity **in the first chapter.**

In the second chapter, we present some theoretical results on unconstrained optimization.

In the third chapter, we introduce classical algorithms for numerical optimization.

Various examples and exercises accompany this document to help assimilate the more theoretical concepts covered in the course.

Chapter 1

Basic Reminders of Differential Calculus and Convexity

1.1 Differential Calculus

1.1.1 Directional Derivative

In this section, we will focus to give some notions in the differential calculus.

Definition 1.1.1. *For every $n \in \mathbb{N}$, $\mathbb{R}^n$ denotes the Euclidean space $\mathbb{R} \times \mathbb{R} \times \cdots \times \mathbb{R}$ (the Cartesian product of n copies of $\mathbb{R}$). A vector $x \in \mathbb{R}^n$ is typically written as $x = (x_1, x_2, \ldots, x_n)^T$ (column vector).*

1.1.2 Canonical Basis and Norms

Let $e_1, e_2, \ldots, e_n$ denote the elements of the canonical basis of $\mathbb{R}^n$, where e_i is the vector in $\mathbb{R}^n$ given by:

$$(e_i)_j = \delta_{ij} = \begin{cases} 0 & \text{if } j \neq i, \\ 1 & \text{if } j = i, \end{cases}$$

for all $i, j = 1, 2, \ldots, n$ (where δ_{ij} is the **Kronecker** symbol).

1.1.3 Dot Product

For any $x, y \in \mathbb{R}^n$, we consider $\langle x, y \rangle \in \mathbb{R}$ the dot product of x and y, which is given by:

$$\langle x, y \rangle = \sum_{i=1}^{n} x_i y_i.$$

Two vectors $x, y \in \mathbb{R}^n$ are orthogonal (denoted $x \perp y$) if $\langle x, y \rangle = 0$.

1.1.4 Euclidean Norm

For any $x \in \mathbb{R}^n$, we denote by $\|x\| \geq 0$ the **Euclidean norm** of x, given by:

$$\|x\| = \sqrt{\langle x, x \rangle} = \sqrt{\sum_{i=1}^{n} x_i^2}.$$

Recall the properties of a norm (and thus also of the Euclidean norm):

i) $\|\alpha x\| = |\alpha|\|x\|$ for all $\alpha \in \mathbb{R}$ and $x \in \mathbb{R}^n$.

ii) $\|x+y\| \leq \|x\| + \|y\|$ for all $x, y \in \mathbb{R}^n$.

iii) $\|0\| = 0$ and $\|x\| > 0$ if $x \in \mathbb{R}^n \setminus \{0\}$.

Definition 1.1.2 (The open ball centered)**.** *For every $x \in \mathbb{R}^n$ and $r > 0$, we denote by $B(x;r)$ the open ball centered at x with radius r, given by:*

$$B(x;r) = \{y \in \mathbb{R}^n \mid \|y-x\| < r\}.$$

Definition 1.1.3. *If $x^{(k)}$ for $k \in \mathbb{N}$ is a sequence in $\mathbb{R}^n$ and x is an element of $\mathbb{R}^n$, we say that $x^{(k)}$ converges to x (denoted $x^{(k)} \to x$) if $\|x^{(k)} - x\| \to 0$.*

meaning that $x^{(k)} \to x$ if and only if $x_i^{(k)} \to x_i$ in $\mathbb{R}$, where $x_i^{(k)}$ (respectively x_i) is the i-th component of $x^{(k)}$ (respectively x).

Definition 1.1.4. *Let $U \subset \mathbb{R}^n$.*

1. *We define the interior of U as the set of elements $x \in U$ for which there exists $r > 0$ such that $B(x;r) \subset U$.*
2. *We say that U is open if for every $x \in U$, there exists $r > 0$ such that $B(x;r) \subset U$.*
3. *We say that U is closed if for every sequence $\{x^{(k)}\} \subset U$ such that $x^{(k)} \to x \in \mathbb{R}^n$, we have $x \in U$.*

Definition 1.1.5. *If $a, b \in \mathbb{R}^n$, we denote by $[a,b]$ the subset of $\mathbb{R}^n$ given by:*

$$[a,b] = \{a + t(b-a) \mid t \in [0,1]\}.$$

The set $[a,b]$ is also called the segment connecting a to b.

Definition 1.1.6 (The Cauchy-Schwarz inequality)**.**

$$|\langle x, y\rangle| \leq \|x\| \cdot \|y\| \quad \textit{for all } x, y \in \mathbb{R}^n.$$

Definition 1.1.7. *We consider the following function as $f : \mathbb{R}^n \to \mathbb{R}$,and let $x_0 \in \mathbb{R}^n$ be a point where $f(x_0)$ is defined. Then the directional derivative of f at x_0 in the direction $d \in \mathbb{R}^n$ is defined by:*

$$f'_d(x_0) = \lim_{t \to 0^+} \frac{f(x_0 + td) - f(x_0)}{t}$$

if it exists.
This derivative gives the rate of change of f at x_0 in the direction d.

Definition 1.1.8 (Fréchet Differentiability)**.** *A function f is said to be* ***Fréchet differentiable (F-differentiable)*** *at $x_0 \in \mathbb{R}^n$ if there exists a continuous linear map $L(x_0)$ from $\mathbb{R}^n$ to $\mathbb{R}$ such that:*

$$\lim_{d \to 0} \frac{f(x_0 + d) - f(x_0) - L(x_0) \cdot d}{\|d\|} = 0.$$

The map $L(x_0)$ is called the derivative of f at x_0.

Example 1.1.1. *We consider the following function* $f:\mathbb{R}^2\to\mathbb{R}$*, be a function defined by:* $f(x_1,x_2)=x_1-x_2^2$*. For any* $d\in\mathbb{R}^2$*, we get*

$$\lim_{t\to 0^+}\frac{f(x+td)-f(x)}{t}=\lim_{t\to 0^+}\frac{f(x_1+td_1,x_2+td_2)-f(x_1,x_2)}{t}$$
$$=(1,-2x_2)\begin{pmatrix}d_1\\ d_2\end{pmatrix}=f'(x)\cdot d.$$

Example 1.1.2. *We consider* $f:\mathbb{R}^2\setminus\{(0,0)\}\to\mathbb{R}$*, be a function defined by:*

$$f(x_1,x_2)=\begin{cases}\frac{x_1^2x_2}{x_1^2+x_2^2}, & (x_1,x_2)\neq(0,0),\\ 0, & (x_1,x_2)=(0,0).\end{cases}$$

Obviously that The function f *is continuous at* $(0,0)$*. Indeed, for* $x_1=r\cos\theta$*,* $x_2=r\sin\theta$*, with* $r>0$ *and* $\theta\in(0,2\pi)$*, we have:*

$$\lim_{(x_1,x_2)\to(0,0)}f(x_1,x_2)=\lim_{r\to 0^+}\frac{r^3\cos^2\theta\sin\theta}{r^2}=0=f(0,0).$$

Furthermore, f *admits partial derivatives at* $(0,0)$ *since:*

$$\frac{\partial f}{\partial x_1}(0,0)=\lim_{d_1\to 0}\frac{f(d_1,0)-f(0,0)}{d_1}=0,$$
$$\frac{\partial f}{\partial x_2}(0,0)=\lim_{d_2\to 0}\frac{f(0,d_2)-f(0,0)}{d_2}=0.$$

However, f *is not Fréchet differentiable (and so not differentiable) at* $(0,0)$ *because:*

$$\lim_{(d_1,d_2)\to(0,0)}\frac{f(d_1,d_2)-f(0,0)-\left(\frac{\partial f}{\partial x_1}(0,0),\frac{\partial f}{\partial x_2}(0,0)\right)\begin{pmatrix}d_1\\ d_2\end{pmatrix}}{\|(d_1,d_2)\|}=\lim_{r\to 0}(\cos^2\theta\sin\theta)$$

does not exist.

1.2 Convexity

In this section, we provide the essential definitions needed to understand the concepts of convexity and concavity in functions. In mathematics, the term "convex" is used to describe two distinct yet related concepts: when it refers to a geometric class or a set of points, it relates to the concept of a convex set. In this section, we introduce the notions of convex sets and convex functions and demonstrate their main geometric and topological properties.

Definition 1.2.1 (Gradient)**.** *We define by*

$$(\nabla f(x))^T=\left[\frac{\partial f}{\partial x_1},\frac{\partial f}{\partial x_2},\dots,\frac{\partial f}{\partial x_n}\right](x),$$

the ***gradient*** *of* f *at the point* $x=(x_1,\dots,x_n)$*. The gradient will play an essential role in the development and analysis of optimization algorithms.*

Example 1.2.1. *We Consider* $f(x_1,x_2,x_3) = e^{x_1} + \frac{x_2^1}{x_3} - x_1x_2x_3$. *Therefore, the gradient of* f *is written as follows*

$$\nabla f(x_1,x_2,x_3) = \begin{bmatrix} e^{x_1} + 2x_1x_3 - x_2x_3 \\ -x_1x_3 \\ \frac{x_2^1}{x_3} - x_1x_2 \end{bmatrix}.$$

1.2.1 Hessian Matrix

Definition 1.2.2. *: The Hessian of* f *is the symmetric matrix in* $\mathbb{M}_n(\mathbb{R})$ *given by*

$$H(x) = \nabla^2 f(x) = \left[\frac{\partial^2 f}{\partial x_i \partial x_j}(x)\right]_{i=1,\dots,n;\ j=1,\dots,n}.$$

Specifically,

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

Example 1.2.2. *We consider* $f(x_1,x_2,x_3) = e^{x_1} + \frac{x_2^2}{x_3} - x_1x_2x_3$. *The Hessian of* f *is given by*

$$H(x) = \begin{bmatrix} e^{x_1} + 2x_3 & -x_3 & -x_1 \\ -x_3 & 0 & -x_1 \\ -x_1 & -x_1 & 0 \end{bmatrix}.$$

Example 1.2.3. *Let us define*

$$f(x_1,x_2,x_3) = e^{x_1} + \frac{x_2^2}{x_3} - x_1x_2x_3$$

. The Hessian of f *is given by*

$$H(x) = \begin{bmatrix} e^{x_1} + 2x_3 & -x_3 & -x_1 \\ -x_3 & 0 & -x_1 \\ -x_1 & -x_1 & 0 \end{bmatrix}.$$

Definition 1.2.3. *We say that* x^* *is a stationary point of* f *if* $\nabla f(x^*) = 0$

In the following Proposition, we will give the relationship between **Gradient** and **Hessian Matrix**.

Proposition 1.2.1. 1. *The i-th row of* $\nabla^2 f(x)$ *is the Jacobian of the i-th component of* ∇f.

2. *We have*

$$\nabla^2 f(x)h = \nabla h^T \nabla f(x), \quad \forall x \in \mathbb{R}^n,\ \forall h \in \mathbb{R}^n.$$

Proof. 1. This is obvious.

2. We have that

$$\frac{\partial}{\partial x_i}\langle \nabla f(x), h\rangle = \frac{\partial}{\partial x_i}\left(\sum_{j=1}^{n}\frac{\partial f}{\partial x_j}(x)h_j\right) = \sum_{j=1}^{n}\frac{\partial^2 f}{\partial x_i \partial x_j}(x)h_j = (\nabla^2 f(x)h)_i.$$

□

Definition 1.2.4. *We say that x^* is a stationary point of f if $\nabla f(x^*) = 0$.*

.

1.2.2 Convex Sets

Definition 1.2.5. *A set $\boldsymbol{S} \subseteq \mathbb{R}^n$ is called convex if:*

$$\forall x_1, x_2 \in \boldsymbol{S},\ \forall \lambda \in [0,1],\ \lambda x_1 + (1-\lambda)x_2 \in S$$

or equivalently:

$$\forall x_1, x_2 \in \boldsymbol{S},\ \forall \lambda_1, \lambda_2 \in \mathbb{R}_+,\ \lambda_1 + \lambda_2 = 1,\ \lambda_1 x_1 + \lambda_2 x_2 \in S.$$

In other words, a geometric object **S** is said to be convex whenever, for any two points x and y in **S**, the segment $[x,y]$ joining them is entirely contained within S.

Example 1.2.4. *We define the following set: $\boldsymbol{S} = \{(x,y) \in \mathbb{R}^2; y \geq x^2\}$*
The set $\boldsymbol{S}$ is convex if:

$$\forall X_1, X_2 \in \boldsymbol{S},\ \forall \lambda \in (0,1),\ \lambda X_1 + (1-\lambda)X_2 \in \boldsymbol{S}.$$

Let X_1, X_2 be two vectors in $\boldsymbol{S}$. Then:

$$X_1 = (x_1, y_1) \in \boldsymbol{S} \implies y_1 \geq x_1^2,$$

and

$$\lambda X_1 + (1-\lambda)X_2 = (\lambda x_1 + (1-\lambda)x_2,\ \lambda y_1 + (1-\lambda)y_2) \in \mathbb{R}^2 \text{ such that}$$

$$\lambda y_1 + (1-\lambda)y_2 - \lambda x_1^2 - (1-\lambda)x_2^2 = \lambda(y_1 - x_1^2) + (1-\lambda)(y_2 - x_2^2) + 2\lambda(1-\lambda)(x_1 x_2),$$

and since

$$\lambda(y_1 - x_1^2) + (1-\lambda)(y_2 - x_2^2) \geq 0,$$

we conclude that:

$$\lambda(y_1 - x_1^2) + (1-\lambda)(y_2 - x_2^2) + 2\lambda(1-\lambda)x_1 x_2 \geq 0.$$

1.2.3 Convex Combination

Definition 1.2.6. *A convex combination of n elements $x_i \in \mathbb{R}^n$ is any element $y \in \mathbb{R}^n$ that can be written in the form:*

$$y = \sum_{i=1}^{n} \lambda_i x_i$$

with the coefficients λ_i satisfying:

$$\lambda_i \geq 0 \text{ and } \sum_{i=1}^{n} \lambda_i = 1.$$

1.2.4 Convex Hull

Definition 1.2.7. *The convex hull of a set $S \subseteq \mathbb{R}^n$ of elements $x_i \in \mathbb{R}^n$ is the set of all convex combinations of the points $x_i \in \mathbb{R}^n$.*

It is also the smallest convex set containing S, which is therefore the intersection of all convex sets containing S. It is denoted by $H(S)$ or $Conv(S)$:

$$H(S) = \left\{ \sum_{i=1}^{n} \lambda_i x_i \mid \exists x_1, x_2, \ldots, x_n \in S \text{ and } \lambda_i \in \mathbb{R}_+, \ \sum_{i=1}^{n} \lambda_i = 1 \right\}.$$

And

$$H(S) = \bigcap \{A \mid A \text{ is a convex set containing } S\}.$$

Example 1.2.5.

1. *We consider $S_1 = \{x, y\}$, then $H(S_1)$ is the segment $[x, y]$.*
2. *Let define$S_2 = \{x, y, z\}$, then $H(S_2)$ is the closed triangle with vertices x, y, z.*

1.2.5 Convex, Strictly Convex, and Strongly Convex Functions

Definition 1.2.8. *A set $U \subseteq \mathbb{R}^n$ is said to be convex if for all $x, y \in U$, we have $[x, y] \subseteq U$ (i.e., for any two points in U, the entire segment joining them is contained within U)*

Definition 1.2.9. *Let $U \subseteq \mathbb{R}^n$ be a convex set, and $f : U \to \mathbb{R}$ a function.*

1. *We say that f is convex on U if:*

$$f(ty + (1-t)x) \leq tf(y) + (1-t)f(x), \quad \forall x, y \in U, \ \forall t \in [0, 1].$$

2. *We say that f is strictly convex on U if:*

$$f(ty + (1-t)x) < tf(y) + (1-t)f(x), \quad \forall x, y \in U \text{ with } x \neq y, \ \forall t \in]0, 1[.$$

3. *We say that f is strongly convex on U if there exists $\beta > 0$ such that:*

$$f(ty + (1-t)x) \leq tf(y) + (1-t)f(x) - \beta t(1-t)\|y - x\|^2, \quad \forall x, y \in U, \ \forall t \in [0, 1].$$

Definition 1.2.10. *A function $f : \mathbb{R}^n \to \mathbb{R}$ is said to be concave if $-f$ is a convex function, that is, if:*

$$\forall x, y \in \mathbb{R}^n, \ \forall t \in [0, 1], \ f(tx + (1-t)y) \geq tf(x) + (1-t)f(y).$$

Proposition 1.2.2 (Jensen's Inequality)**.** *If $f : \mathbb{R}^n \to \mathbb{R}$ is convex, then:*

$$f\left(\sum_{i=1}^{m} t_i x_i\right) \leq \sum_{i=1}^{m} t_i f(x_i), \quad \forall m \in \mathbb{N},\ \forall t_i \geq 0,\ \textit{with}\ \sum_{i=1}^{m} t_i = 1,\ \forall x_i \in \mathbb{R}^n.$$

Proof. By induction on $m \geq 1$.

- For $m = 1$, it is obvious. For $m = 2$, $t_1 + t_2 = 1$, so $t_2 = 1 - t_1$, let $t_1 = t$, then $t_2 = 1 - t$ with $t \in [0,1]$, which is the definition.

- Suppose $P(m)$ is true. Let $x_1, x_2, \ldots, x_{m+1} \in \mathbb{R}^n$, and $t_1, t_2, \ldots, t_{m+1} \in \mathbb{R}_+$ with:

$$\sum_{i=1}^{m+1} t_i = 1,$$

we need to show that:

$$f\left(\sum_{i=1}^{m+1} t_i x_i\right) \leq \sum_{i=1}^{m+1} t_i f(x_i).$$

- If $t_{m+1} = 1$, then $t_1 = t_2 = \cdots = t_m = 0$, and the property is true.
- If $t_{m+1} \neq 1$, let $\theta = 1 - t_{m+1} > 0$, and we have:

$$t_1 x_1 + t_2 x_2 + \cdots + t_{m+1} x_{m+1} = \theta \sum_{i=1}^{m} \frac{t_i}{\theta} x_i + (1-\theta) x_{m+1}.$$

Let $x = \sum_{i=1}^{m} \frac{t_i}{\theta} x_i = \sum_{i=1}^{m} \theta_i x_i$, with $\theta_i = \frac{t_i}{\theta}$, we note that $\theta_i \geq 0$ and $\sum_{i=1}^{m} \theta_i = 1$. Therefore, we get:

$$f\left(\sum_{i=1}^{m+1} t_i x_i\right) = f\left(\theta x + (1-\theta) x_{m+1}\right) \leq \theta f(x) + (1-\theta) f(x_{m+1}),$$

and by the inductive hypothesis:

$$\theta f(x) \leq \theta \sum_{i=1}^{m} \theta_i f(x_i),$$

hence,

$$f\left(\sum_{i=1}^{m+1} t_i x_i\right) \leq \sum_{i=1}^{m+1} t_i f(x_i),$$

□

In general way it is difficult to show the convexity of a function using only the definition. The following propositions provide criteria for convexity, strict convexity, and strong convexity, which are easier to use than the respective definitions.

Proposition 1.2.3 (Characterization of Convexity)**.** *Let $\Omega \subseteq \mathbb{R}^n$ be open, $U \subseteq \Omega$ be convex, and $f : \Omega \to \mathbb{R}$ be a C^1-class function. Then:*

a) The following three statements are equivalent:

1. *f is convex on U,*
2. *$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle, \quad \forall x, y \in U$,*

3. ∇f is monotone on U, that is:

$$\langle \nabla f(y) - \nabla f(x), y - x \rangle \geq 0, \quad \forall x, y \in U.$$

b) Furthermore, if f is C^2-class on Ω, then f is convex on U if and only if:

$$\langle \nabla^2 f(x)(y-x), y-x \rangle \geq 0, \quad \forall x, y \in U.$$

Proof. a) We prove here the equivalence between 1), 2), and 3).

1) $\Rightarrow$ 2): Suppose f is convex. The definition of convexity can be written as:

$$f(x + t(y-x)) \leq (1-t)f(x) + tf(y)$$

By fixing x and y, dividing by t, and letting t approach 0 (which is possible because $t \in [0,1]$), we obtain 2).

2) $\Rightarrow$ 3): From 2), we deduce:

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle \quad \forall x, y \in U,$$

and also by swapping x and y)) :$f(x) \geq f(y) + \langle \nabla f(y), x - y \rangle \quad \forall x, y \in U$,

Adding these two inequalities, we obtain 3).

3) $\Rightarrow$ 1): We consider that x and y be fixed in U. We introduce the function $g : I \to \mathbb{R}$ defined by

$$t \in I \mapsto g(t) = f(ty + (1-t)x)$$

where I is an open interval containing $[0,1]$. It is easy to see that g is of class C^1, and we have

$$g'(t) = \langle \nabla f(ty + (1-t)x), y - x \rangle \quad \forall t \in I$$

Let $t_1, t_2 \in [0,1]$ with $t_1 < t_2$. Then

$$g'(t_2) - g'(t_1) = \langle \nabla f(x + t_2(y-x)) - \nabla f(x + t_1(y-x)), y - x \rangle = \langle \nabla f(x + t_2(y-x)) - \nabla f(x + t_1(y-x)), (t_2 - t_1$$

By assumption 3), the last term of the previous equality is ≥ 0, which shows that g' is a non-decreasing function. We then deduce that g is a convex function on $[0,1]$, which gives for all $t \in [0,1]$:

$$g(t_1 + (1-t)t_0) \leq tg(1) + (1-t)g(0)$$

which means

$$f(ty + (1-t)x) \leq tf(y) + (1-t)f(x)$$

thus f is convex.

$b)$ Suppose $f \in C^2(U)$.

$\Rightarrow$ Suppose f is convex and show (2.2). Let $h \in \mathbb{R}^n$ be fixed, and consider the function $g : U \to \mathbb{R}$ given by $g(x) = \langle \nabla f(x), h \rangle$ for all $x \in U$. Using Proposition 2.1:

$$\langle \nabla^2 f(x)h, h \rangle = \langle \nabla g(x), h \rangle = \frac{\partial g}{\partial h}(x) = \lim_{t \to 0} \frac{\langle \nabla f(x+th), h \rangle - \langle \nabla f(x), h \rangle}{t}$$

which gives:

$$\langle \nabla^2 f(x)h, h \rangle = \lim_{t \to 0} \frac{\langle \nabla f(x+th) - \nabla f(x), th \rangle}{t^2}$$

Now, consider arbitrary $x,y \in U$ and $h = y-x$. Since $x+t(y-x) \in U$ for all $t \in [0,1]$, from the previous equality and using the monotonicity of ∇f, we deduce that:

$$\langle \nabla^2 f(x)h, h\rangle \geq 0$$

which means (2.2).

$\Leftarrow$ Now, suppose (2.2) is satisfied and show that f is convex. Let $x,y \in U$ be fixed, and consider the function $g_1 : U \to \mathbb{R}$ given by $g_1(z) = \langle \nabla f(z), x-y\rangle$ for all $z \in U$. Then:

$$\langle \nabla f(x) - \nabla f(y), x-y\rangle = g_1(x) - g_1(y) = \langle \nabla g_1(y+\theta(x-y)), x-y\rangle$$

with $\theta \in (0,1)$ (using one of Taylor's formulas). On the other hand, we have:

$$\nabla g_1(z) = \nabla^2 f(z)(x-y)$$

and this allows us to deduce, using (2.2):

$$\langle \nabla f(x) - \nabla f(y), x-y\rangle = \langle \nabla^2 f(y+\theta(x-y))(x-y), x-y\rangle \geq 0$$

This gives us the monotonicity of ∇f, hence the convexity of f. □

Example 1.2.6. *1. $f(x) = ax^2+bx+c$ with $a>0$*

2. $f(x) = x^2+\sin(x)$ (since $f''(x) = 2-\sin(x) \geq 1$ for all $x \in \mathbb{R}$).

3. Let $f : \mathbb{R}^n \to \mathbb{R}$ be given by

$$f(x) = \frac{1}{2}\langle Ax, x\rangle - \langle b, x\rangle + c, \quad \forall x \in \mathbb{R}^n$$

where $A \in \mathbb{M}_n(\mathbb{R})$ is a real symmetric square matrix of size n, $b \in \mathbb{R}^n$ is a vector, and $c \in \mathbb{R}$ is a scalar (a function of this type is also called a quadratic function or form). It is easy to compute:

$$\nabla f(x) = Ax - b$$
$$\nabla^2 f(x) = A$$

(hence the Hessian of f is constant), we can deduce the following results

1. f is a convex function $\Leftrightarrow A$ is a positive semi-definite matrix.

2. f is strongly convex $\Leftrightarrow f$ is strictly convex $\Leftrightarrow A$ is a positive definite matrix (positive definite matrix).

We consider the following function:

$$f(x,y,z) = x^2+y^2+z^2-xy+xz, \quad (x,y,z) \in \mathbb{R}^3$$

The function f is of class C^2 on $\mathbb{R}^3$ (where $\mathbb{R}^3$ is an open and convex set).

The gradient of f is:

$$\nabla f(x,y,z) = \begin{pmatrix} 2x-y+z \\ 2y-x \\ 2z+x \end{pmatrix}$$

The Hessian matrix $Hf(x,y,z) = \nabla^2 f(x,y,z)$ is:

$$Hf(x,y,z) = \begin{pmatrix} 2 & -1 & 1 \\ -1 & 2 & 0 \\ 1 & 0 & 2 \end{pmatrix}$$

To determine the convexity, we calculate the eigenvalues of this matrix. We have:

$$\det(Hf - \lambda I) = \det \begin{pmatrix} 2-\lambda & -1 & 1 \\ -1 & 2-\lambda & 0 \\ 1 & 0 & 2-\lambda \end{pmatrix}$$

Expanding the determinant, we get:

$$\begin{aligned} \det(Hf - \lambda I) &= (2-\lambda)\left[(2-\lambda)^2 - 0\right] - (-1)\left[-1(2-\lambda) - 1(0)\right] + 1\left[-1(0) - 1(2-\lambda)\right] \\ &= (2-\lambda)\left[(2-\lambda)^2 - 2\right] \\ &= (2-\lambda)((2-\lambda)^2 - 2) = 0 \end{aligned}$$

Thus, the eigenvalues are:

$$\lambda_1 = 2, \quad \lambda_2 = 2+\sqrt{2}, \quad \lambda_3 = 2-\sqrt{2}$$

Since $\lambda_1, \lambda_2, \lambda_3 > 0$, the function f is strictly convex.

Remark 1.2.1. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be given by*

$$f(x) = \langle Ax, x \rangle \quad \forall x \in \mathbb{R}^n$$

where $A \in \mathbb{M}_n(\mathbb{R})$ is a real square matrix of size n (i.e., f is the quadratic form associated with the matrix A). Then, for a fixed $p \in \{1,2,\ldots,n\}$, we can write

$$f(x) = \sum_{i,j=1}^{n} A_{ij}x_i x_j = A_{pp}x_p^2 + \sum_{j=1, j\neq p}^{n} A_{pj}x_p x_j + \sum_{i=1, i\neq p}^{n} A_{ip}x_i x_p + \sum_{i,j=1, i\neq p, j\neq p}^{n} A_{ij}x_i x_j$$

which gives

$$\frac{\partial f}{\partial x_p} = 2A_{pp}x_p + \sum_{j=1, j\neq p}^{n} A_{pj}x_j + \sum_{i=1, i\neq p}^{n} A_{ip}x_i = \sum_{j=1}^{n} A_{pj}x_j + \sum_{i=1}^{n} A_{ip}x_i = (Ax)_p + (A^T x)_p$$

Thus, we obtain:

$$\nabla f(x) = (A + A^T)x \quad \forall x \in \mathbb{R}^n$$

Using the formula $\nabla^2 f = J\nabla f$, we deduce:

$$\nabla^2 f(x) = A + A^T \quad \forall x \in \mathbb{R}^n$$

(so the Hessian of f is constant).

In particular, if A is symmetric (i.e., $A = A^T$), then

$$\nabla \langle Ax, x \rangle = 2Ax \quad \forall x \in \mathbb{R}^n$$

$$\nabla^2 \langle Ax, x \rangle = 2A \quad \forall x \in \mathbb{R}^n$$

1.2.6 Taylor's Theorem

Theorem 1.2.1. *Let $U \subset \mathbb{R}^n$ be open, $f : U \to \mathbb{R}$, $a \in U$, and $h \in \mathbb{R}^n$ such that $[a, a+h] \subset U$. Then, we have that*

1. *If $f \in C^1(U)$, then*

$$f(a+h) = f(a) + \int_0^1 \langle \nabla f(a+th), h \rangle \, dt$$

(Taylor's theorem of order 1 with integral remainder).

$$f(a+h) = f(a) + \langle \nabla f(a+\theta h), h \rangle \text{ with } 0 < \theta < 1$$

(Taylor-Maclaurin theorem of order 1).

$$f(a+h) = f(a) + \langle \nabla f(a), h \rangle + o(\|h\|)$$

(Taylor-Young theorem of order 1).

2. *If $f \in C^2(U)$, then*

$$f(a+h) = f(a) + \langle \nabla f(a), h \rangle + \int_0^1 (1-t) \langle \nabla^2 f(a+th)h, h \rangle \, dt$$

(Taylor's theorem of order 2 with integral remainder).

$$f(a+h) = f(a) + \langle \nabla f(a), h \rangle + \frac{1}{2} \langle \nabla^2 f(a+\theta h)h, h \rangle \text{ with } 0 < \theta < 1$$

(Taylor-Maclaurin theorem of order 2).

$$f(a+h) = f(a) + \langle \nabla f(a), h \rangle + \frac{1}{2} \langle \nabla^2 f(a)h, h \rangle + o(\|h\|^2)$$

(Taylor-Young theorem of order 2).

Remark 1.2.2. *In the previous proposition, the notation $o(\|h\|^k)$ for $k \in \mathbb{N}$ means an expression that tends to 0 faster than $\|h\|^k$.*

1.2.7 Exercices

Exercice 1:

Which of the following sets are convex?

1. $S_1 = \{(x,y) \in \mathbb{R}^2 \mid 0 \leq x \leq 1,\ y = 0\}$
2. $S_2 = \{(x,y) \in \mathbb{R}^2 \mid x \geq 0,\ y \geq 0,\ x+y \leq 1\}$
3. $S_3 = \{x \in \mathbb{R}^n \mid A_1 x = b_1,\ A_2 x \leq b_2\}$ where A_1 and A_2 are matrices of size $m \times n$, and b_1 and b_2 are vectors in $\mathbb{R}^m$.
4. $S_4 = \{(x,y) \in \mathbb{R}^2 \mid y - x^2 \geq 0\}$
5. $S_5 = \{(x,y) \in \mathbb{R}^2 \mid xy \geq 1 \text{ and } x > 0\}$

Exercice 2:
Verify whether the following functions are convex or not on $\mathbb{R}^2$:

1. $f(x,y) = x^2 - xy + 2y^2 - 2x + e^{x+y}$
2. $f(x,y) = (x-2)^4 + (x-2)^2y^2 + (y+1)^2$
3. $f(x,y) = -x^2 - 2xy - 2y^2$

Exercice 3:
We consider the function f defined on $\mathbb{R}^2$ by

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

1. Show that there exist $(a,b) \in \mathbb{R}_+^2$ (and determine them) such that

   $$f(x,y) \geq a\|(x,y)\|^2 + b$$

   for all $(x,y) \in \mathbb{R}^2$, where $\|\cdot\|$ denotes the Euclidean norm on $\mathbb{R}^2$. Deduce that the problem

   $$\inf_{(x,y)\in\mathbb{R}^2} f(x,y) \quad (P)$$

   has at least one solution.
2. Is the function f convex on $\mathbb{R}^2$

Exercice 4:
We define the function $J : \mathbb{R}^2 \to \mathbb{R}$ by

$$J(x,y) = y^4 - 3xy^2 + x^2.$$

1. Determine the critical points of J.
2. Let $d = (d_1, d_2) \in \mathbb{R}^2$. Using the function $t \mapsto J(td_1, td_2)$, show that $(0,0)$ is a local minimum along any line passing through $(0,0)$.
3. Is the point $(0,0)$ a local minimum of the restriction of J to the parabola given by the equation $x = y^2$?
4. Compute the Hessian matrix of J. What is the nature of the critical point $(0,0)$?

1.2.8 Corrections

Exercice 1

1. Let $S_1 = \{(x,y) \in \mathbb{R}^2 \mid y - x \geq 0\}$. The set S is convex if for any $X_1, X_2 \in S_1$ and $\lambda \in [0,1]$, the point $\lambda X_1 + (1-\lambda)X_2$ also belongs to S.

   Let $X_1 = (x_1, y_1)$ and $X_2 = (x_2, y_2)$ be two vectors in S_1. Then

   $$y_1 - x_1 \geq 0 \quad \text{and} \quad y_2 - x_2 \geq 0.$$

   We need to show that

   $$\lambda X_1 + (1-\lambda)X_2 = (\lambda x_1 + (1-\lambda)x_2, \lambda y_1 + (1-\lambda)y_2) \in S_1.$$

The condition for convexity is:

$$\lambda y_1 + (1-\lambda)y_2 - (\lambda x_1 + (1-\lambda)x_2) \geq 0.$$

Simplify:

$$\lambda y_1 + (1-\lambda)y_2 - \lambda x_1 - (1-\lambda)x_2 = \lambda(y_1 - x_1) + (1-\lambda)(y_2 - x_2).$$

Since $y_1 - x_1 \geq 0$ and $y_2 - x_2 \geq 0$, it follows that:

$$\lambda(y_1 - x_1) + (1-\lambda)(y_2 - x_2) \geq 0.$$

Consequently, $\lambda X_1 + (1-\lambda)X_2 \in S_1$, proving that S_1 is convex.

2. Let $S_2 = \{(x,y) \in \mathbb{R}^2 \mid x \geq 0,\ y \geq 0,\ x + y \leq 1\}$. The set S_2 is convex if for any $X_1, X_2 \in S$ and $\lambda \in [0,1]$, the point $\lambda X_1 + (1-\lambda)X_2$ also belongs to S_2.

   Let $X_1 = (x_1, y_1)$ and $X_2 = (x_2, y_2)$ be two vectors in S_2. Then:

   $$x_1 \geq 0,\ y_1 \geq 0,\ x_1 + y_1 \leq 1$$

   and

   $$x_2 \geq 0,\ y_2 \geq 0,\ x_2 + y_2 \leq 1.$$

   We need to show that

   $$\lambda X_1 + (1-\lambda)X_2 = (\lambda x_1 + (1-\lambda)x_2, \lambda y_1 + (1-\lambda)y_2) \in S_2.$$

   We need to check the following conditions:

   - $\lambda x_1 + (1-\lambda)x_2 \geq 0.$
   - $\lambda y_1 + (1-\lambda)y_2 \geq 0.$
   - $(\lambda x_1 + (1-\lambda)x_2) + (\lambda y_1 + (1-\lambda)y_2) \leq 1.$

   We simplify,

   $$(\lambda x_1 + (1-\lambda)x_2) + (\lambda y_1 + (1-\lambda)y_2) = \lambda(x_1 + y_1) + (1-\lambda)(x_2 + y_2) \leq \lambda \cdot 1 + (1-\lambda) \cdot 1 = 1.$$

   Hence,

   $$\lambda X_1 + (1-\lambda)X_2 \in S,$$

   proving that S_2 is convex.

3. For all $x, y \in S_3$ and all $\lambda \in [0,1]$, we have:

   $$A_1(\lambda x + (1-\lambda)y) = \lambda A_1 x + (1-\lambda)A_1 y = b_1,$$

   and

   $$A_2(\lambda x + (1-\lambda)y) \leq \lambda A_2 x + (1-\lambda)A_2 y = b_2,$$

   which implies that $\lambda x + (1-\lambda)y \in E$. Therefore, S_3 is a convex set.

4. For all $X,Y \in S_4$ and all $\lambda \in [0,1]$, we have:

$$\lambda X + (1-\lambda)Y = (\lambda x_1 + (1-\lambda)x_2,\ \lambda y_1 + (1-\lambda)y_2),$$

where $y_1 - x_1^2 \geq 0$ and $y_2 - x_2^2 \geq 0$. Thus, $y_1 \geq x_1^2$ and $y_2 \geq x_2^2$. On the other hand, we have:

$$\lambda y_1 + (1-\lambda)y_2 - \lambda^2 x_1^2 - (1-\lambda)^2 x_2^2 - 2\lambda(1-\lambda)x_1x_2 \geq \lambda(1-\lambda)(x_1-x_2)^2 \geq 0.$$

Therefore, S_4 is a convex set.

5. For all $X,Y \in S_5$ and all $\lambda \in [0,1]$, we have:

$$\lambda X + (1-\lambda)Y = (\lambda x_1 + (1-\lambda)x_2,\ \lambda y_1 + (1-\lambda)y_2),$$

where $x_1, x_2 > 0$. Thus, $\lambda x_1 + (1-\lambda)x_2 > 0$. On the other hand, we have:

$$(\lambda x_1 + (1-\lambda)x_2)(\lambda y_1 + (1-\lambda)y_2) = \lambda^2 x_1y_1 + \lambda(1-\lambda)x_1y_2 + \lambda(1-\lambda)x_2y_1 + (1-\lambda)^2 x_2y_2$$
$$> \lambda^2 + (1-\lambda)^2 + \lambda(1-\lambda)\left(\frac{x_1}{x_2} + \frac{x_2}{x_1}\right) \geq 0,$$

which implies that $\lambda X + (1-\lambda)Y \in S_5$. Therefore, S_5 is not a convex set.

Exercise 2:

1. The gradient of $f(x,y)$ is:

$$\nabla f(x,y) = \begin{pmatrix} 2x - y + e^{x+y} - 2 \\ -x + xy + e^{x+y} \end{pmatrix},$$

and the Hessian matrix of $f(x,y)$ is:

$$Hf(x,y) = \begin{pmatrix} e^{x+y} + 2e^{x+y} - 1 & -1 + e^{x+y} \\ -1 + e^{x+y} & 4 + e^{x+y} \end{pmatrix}.$$

The eigenvalues are $\lambda_1 = e^{x+y} + 2 > 0$ and $\lambda_2 = 8e^{x+y} + 7 > 0$, which implies that f is strictly convex.

2. The Hessian matrix of $f(x,y)$ is:

$$Hf(x,y) = \begin{pmatrix} 12(x-2)^2 + 2y^2 & 4y(x-2) \\ 4y(x-2) & 2(x-2)^2 + 2 \end{pmatrix}.$$

We have:

$$\lambda_1 = 12(x-2)^2 + 2y^2 \geq 0,$$

and

$$\lambda_2 = 24(x-2)^4 + (x-2)^2(24 - 12y^2) + 4y^2.$$

For $x = 0$ and $y = 5$, $\lambda_2 < 0$; for $x = y = 0$, $\lambda_2 > 0$, so f is neither convex nor concave.

3. The Hessian matrix of $f(x,y)$ is:

$$Hf(x,y) = \begin{pmatrix} -2 & -2 \\ -2 & -4 \end{pmatrix}.$$

The eigenvalues are $\lambda_1 < 0$ and $\lambda_2 > 0$, which implies that f is strictly concave.

Exercise 3:

1. The function f is polynomial and hence of class $C^\infty(\mathbb{R}^2)$. Using the fact that $xy \leq -\frac{1}{2}(x^2+y^2)$, we can write:

$$f(x,y) \leq x^4+y^4-2x^2-2y^2+4xy \leq x^4+y^4-4x^2-4y^2,$$

for all $(x,y) \in \mathbb{R}^2$. Using the fact that for all $(X,\#) \in \mathbb{R}^2$, $X^4+\#^4-2\#X^2 \geq 0$, we get:

$$f(x,y) \leq (2\#-4)x^2+(2\#-4)y^2-2\#^4.$$

For example, if we choose $\# = 3$, we deduce:

$$f(x,y) \leq 2(x^2+y^2)-162-\|(x,y)\|_\infty - \infty.$$

This proves that f is coercive on $\mathbb{R}^2$, which is closed and finite-dimensional. According to the theorem covered in class, problem (P) has at least one solution.

2. To study the convexity of f (which is of class C^2 on $\mathbb{R}^2$), we calculate its Hessian matrix at any point (x,y) in $\mathbb{R}^2$. We have:

$$\text{Hess } f(x,y) = \begin{pmatrix} 12x^2-4 & 0 \\ 0 & 12y^2-4 \end{pmatrix}.$$

Recall that f is convex on $\mathbb{R}^2$ if and only if its Hessian matrix is positive semi-definite at all points. However, it is easy to verify that the eigenvalues of Hess $f(0,0)$ are 0 and -2. Therefore, f is not convex.

3. The critical points of f are given by the solutions to $\nabla f(x,y) = (0,0)$, i.e., the critical points are solutions to the system:

$$\begin{cases} x^3-(x-y)=0 \\ y^3+(x-y)=0 \end{cases},$$

which simplifies to:

$$\begin{cases} x^3+y^3=0 \\ y^3+(x-y)=0 \end{cases},$$

or:

$$\begin{cases} y=-x \\ x^3-2x=0 \end{cases}.$$

We deduce that f has three critical points: $O(0,0)$, $A\left(\sqrt{2},-\sqrt{2}\right)$, and $B\left(-\sqrt{2},\sqrt{2}\right)$.

Chapter 2

Unconstrained Minimization

We consider the following function as follows $f : \mathbb{R}^n \to \mathbb{R}$. The problem of unconstrained minimization is defined as follows:

$$(P) \quad \text{minimize } f(x) \text{ for } x \in \mathbb{R}^n.$$

The analysis of these problems is useful for various domains. Many constrained optimization problems are transformed into sequences of unconstrained optimization problems (e.g., Lagrange multipliers, penalty methods, etc.). The study of unconstrained optimization problems also gets applications in solving nonlinear systems. A large class of algorithms that we will consider for unconstrained optimization problems have the following general form:

Given x_0, calculate $x_{k+1} = x_k + \alpha_k d_k$;

where d_k is called the descent direction and α_k is the step size at the k-th iteration. In practice, we almost always ensure the following inequality:

$$f(x_{k+1}) \leq f(x_k),$$

which ensures the desirable decrease of the objective function f. Such algorithms are often called descent methods. Essentially, the difference between these algorithms lies in the choice of the descent direction d_k. Once the direction is chosen, we are more or less reduced to a one-dimensional problem to determine α_k. To approach the optimal solution of problem (P) (in general, it is a point where the necessary conditions for optimality of f may hold with some precision), we naturally move from the point x_k in the direction of the decrease of the function f. Unconstrained optimization has the following properties:

- All methods require a starting point x_0.
- Deterministic methods converge to the nearest local minimum.
- The more you know about the function (gradient, Hessian), the more efficient the minimization will be.

Let the unconstrained optimization problem (P).

Definition 2.0.1. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a continuously differentiable function.*

a) A point $\hat{x} \in \mathbb{R}^n$ is called a global optimal solution of (P) if and only if:

$$\forall x \in \mathbb{R}^n, \quad f(\hat{x}) \leq f(x).$$

b) *A point $\hat{x} \in \mathbb{R}^n$ is called a local optimal solution of (P) if and only if there exists a neighborhood $V_\varepsilon(\hat{x})$ of $\hat{x}$ such that:*

$$f(\hat{x}) \leq f(x), \quad \forall x \in V_\varepsilon(\hat{x}).$$

c) *A point $\hat{x} \in \mathbb{R}^n$ is called a strict optimal solution of (P) if and only if there exists a neighborhood $V_\varepsilon(\hat{x})$ of $\hat{x}$ such that:*

$$f(\hat{x}) < f(x), \quad \forall x \in V_\varepsilon(\hat{x}) \text{ and } x \neq \hat{x}.$$

2.1 Existence and Uniqueness Results

Before studying the properties of the solution(s) of (P), we must ensure their existence. We will then provide results on uniqueness.

Definition 2.1.1. *Let $f : \mathbb{R}^n \to \mathbb{R}$. We say f is **coercive** if:*

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

Here, $\|\cdot\|$ denotes any norm on $\mathbb{R}^n$. We denote $\|\cdot\|_p$ (with $p \in \mathbb{N}$) the l_p norm on $\mathbb{R}^n$:

$$\forall x = (x_1, \ldots, x_n) \in \mathbb{R}^n, \quad \|x\|_p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{\frac{1}{p}}.$$

The infinity norm on $\mathbb{R}^n$ is:

$$\forall x = (x_1, \ldots, x_n) \in \mathbb{R}^n, \quad \|x\|_\infty = \max_{1 \leq i \leq n} |x_i|.$$

Theorem 2.1.1 (Existence)**.** *Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be proper, continuous, and coercive. Then (P) has at least one solution.*

Proof. Let $R = \inf_{x \in \mathbb{R}^n} f(x) < +\infty$, and let $\{x_k\}_{k \in \mathbb{N}}$ be a minimizing sequence, i.e.,

$$\lim_{k \to +\infty} f(x_k) = R < +\infty. \tag{1}$$

Assume that $\{x_k\}_{k \in \mathbb{N}}$ is unbounded. Then, there exists a subsequence $\{x_{k_j}\}_{j \in \mathbb{N}}$ such that:

$$\lim_{j \to +\infty} \|x_{k_j}\| = +\infty.$$

By the coercivity of f, we have:

$$\lim_{j \to +\infty} f(x_{k_j}) = +\infty,$$

which contradicts equation (1). Therefore, the sequence $\{x_k\}_{k \in \mathbb{N}}$ must be bounded.

Since $\{x_k\}_{k \in \mathbb{N}}$ is bounded, by the Bolzano-Weierstrass theorem, there exists a subsequence $\{x_{k_j}\}_{j \in \mathbb{N}}$ that converges to some point $x^* \in \mathbb{R}^n$. Using the continuity of f, we have:

$$f(x^*) = \lim_{j \to +\infty} f(x_{k_j}) = R.$$

Thus, x^* is a solution to problem (P), and $R > -\infty$.

□

Theorem 2.1.2 (Uniqueness)**.** *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a strictly convex function. Then, there exists at most one $x^* \in \mathbb{R}^n$ such that:*

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

Proof. Let f be strictly convex. Suppose there exist x_1^* and $x_2^* \in \mathbb{R}^n$ such that $f(x_1^*) = f(x_2^*) = \min_{x\in\mathbb{R}^n} f(x)$. Assume that $x_1^* \neq x_2^*$. Since f is strictly convex, we have:

$$f\left(\frac{1}{2}x_1^* + \frac{1}{2}x_2^*\right) < \frac{1}{2}f(x_1^*) + \frac{1}{2}f(x_2^*) = \min_{x\in\mathbb{R}^n} f(x),$$

which is a contradiction. Thus, $x_1^* = x_2^*$. □

Theorem 2.1.3 (Existence and Uniqueness)**.** *Let $f : \mathbb{R}^n \to \mathbb{R}$, satisfying:*

1. *f is continuous;*
2. *f is coercive;*
3. *f is strictly convex.*

Then, there exists a unique $x^* \in \mathbb{R}^n$ such that $f(x^*) = \min_{x\in\mathbb{R}^n} f(x)$.

Theorem 2.1.4. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a continuously differentiable function (class C^1). Suppose there exists $\beta > 0$ such that:*

$$\forall x, y \in \mathbb{R}^n, \quad \langle \nabla f(x) - \nabla f(y), x - y\rangle \geq \beta \|x - y\|^2. \tag{2.1}$$

Then f is strictly convex and coercive, and in particular, the problem (P) admits a unique solution.

Proof. The condition (2.1) implies that ∇f is monotone and that f is convex. Additionally, f is strictly convex. Finally, f is coercive. Indeed, let us apply Taylor's formula with an integral remainder:

$$f(y) = f(x) + \int_0^1 \frac{d}{dt} f(x + t(y - x))\, dt = f(x) + \int_0^1 (\nabla f(x + t(y - x)), y - x)\, dt.$$

Thus,

$$f(y) = f(x) + (\nabla f(x), y - x) + \int_0^1 (\nabla f(x + t(y - x)) - \nabla f(x), y - x)\, dt. \tag{2.3}$$

From (2.2), we obtain

$$f(y) \geq f(x) + (\nabla f(x), y - x) + \int_0^1 t\beta \|x - y\|^2\, dt.$$

Finally,

$$f(y) \geq f(x) - \|\nabla f(x)\| \|y - x\| + \frac{\beta}{2}\|x - y\|^2.$$

Fixing $x = 0$, for instance, it becomes clear that f is coercive.
Consequently, f admits a unique minimum $x^\star$ on $\mathbb{R}^n$, characterized by

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

□

2.2 Optimality conditions

Optimality conditions are equations, inequalities, or properties that the solutions of (P) satisfy (necessary conditions) or that guarantee a point to be a solution of (P) (sufficient condition).

We refer to first-order conditions when they involve only the first derivatives of f. Second-order conditions, on the other hand, involve both the first and second derivatives of f.

2.2.1 Necessary Optimality Conditions

Given a point $\hat{x}$, the continuous differentiability of the function f provides a primary way to characterize an optimal solution.

First-Order Necessary Optimality Conditions

Theorem 2.2.1. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be differentiable at a point $x \in \mathbb{R}^n$. Let $d \in \mathbb{R}^n$ be such that $\nabla f(x)^T d < 0$. Then, there exists $\varepsilon > 0$ such that $f(x+\alpha d) < f(x)$ for all $\alpha \in (0,\varepsilon)$. In this case, the direction d is called a descent direction.*

Proof. Since f is differentiable at x, we have:

$$f(x+\alpha d) = f(x) + \alpha \nabla f(x)^T d + \alpha \|d\| \phi(x,\alpha d),$$

where $\phi(x,\alpha d) \to 0$ as $\alpha \to 0$. This implies:

$$\frac{f(x+\alpha d) - f(x)}{\alpha} = \nabla f(x)^T d + \|d\| \phi(x,\alpha d), \quad \alpha \neq 0.$$

Since $\nabla f(x)^T d < 0$ and $\phi(x,\alpha d) \to 0$ as $\alpha \to 0$, there exists $\varepsilon > 0$ such that:

$$\nabla f(x)^T d + \|d\| \phi(x,\alpha d) < 0 \quad \text{for all} \quad \alpha \in (0,\varepsilon).$$

Consequently, we obtain:

$$f(x+\alpha d) < f(x) \quad \text{for all} \quad \alpha \in (0,\varepsilon).$$

□

Theorem 2.2.2. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be differentiable at a point $\hat{x} \in \mathbb{R}^n$. If $\hat{x}$ is a local minimum of the problem (P), then $\nabla f(\hat{x}) = 0$.*

Proof. We prove by contradiction. Assume that $\nabla f(\hat{x}) \neq 0$.

Let $d = -\nabla f(\hat{x})$. We then obtain:

$$\nabla f(\hat{x})^T d = -\|\nabla f(\hat{x})\|^2 < 0.$$

By Theorem 2.2.1, there exists $\varepsilon > 0$ such that:

$$f(\hat{x}+\beta d) < f(\hat{x}), \text{ for all } \beta \in (0,\varepsilon).$$

This leads to a contradiction with the fact that $\hat{x}$ is a local minimum. Therefore, $\nabla f(\hat{x}) = 0$. □

Second-Order Necessary Conditions for Optimality

Definition 2.2.1.
- *A symmetric matrix A is said to be positive semi-definite if:*

$$\forall d \in \mathbb{R}^n,\ d^T A d \geq 0.$$

- *A symmetric matrix A is said to be positive definite if:*

$$\forall d \in \mathbb{R}^n,\ d \neq 0,\ d^T A d > 0.$$

Theorem 2.2.3. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be twice differentiable at the point $\hat{x} \in \mathbb{R}^n$. If $\hat{x}$ is a local minimum of (P), then $\nabla f(\hat{x}) = 0$ and the Hessian matrix of f at $\hat{x}$, denoted $H(\hat{x})$, is positive semi-definite.*

Proof. Let $d \in \mathbb{R}^n$ be arbitrary. Since f is twice differentiable at $\hat{x}$, we have:

$$f(\hat{x} + \beta d) = f(\hat{x}) + \frac{1}{2}\beta^2 d^T H(\hat{x}) d + \beta^2 \|d\|^2 \psi(\hat{x}, \beta d),$$

where $\psi(\hat{x}, \beta d) \to 0$ as $\beta \to 0$.
This implies:

$$\frac{f(\hat{x} + \beta d) - f(\hat{x})}{\beta^2} = \frac{1}{2} d^T H(\hat{x}) d + \psi(\hat{x}, \beta d).$$

Since $\hat{x}$ is a local minimum, there exists $\varepsilon > 0$ such that:

$$\frac{f(\hat{x} + \beta d) - f(\hat{x})}{\beta^2} \geq 0, \text{ for all } \beta \in (0, \varepsilon).$$

Thus, we have:

$$\frac{1}{2} d^T H(\hat{x}) d \geq 0, \text{ for all } d \in \mathbb{R}^n.$$

Hence, $H(\hat{x})$ is positive semi-definite.

□

2.2.2 Sufficient Conditions for Optimality

The conditions given previously are necessary (if f is not convex), meaning they must be satisfied for any local minimum. However, a point satisfying these conditions is not necessarily a local minimum.

Theorem 2.2.4. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be twice differentiable at the point $\hat{x} \in \mathbb{R}^n$. If $\nabla f(\hat{x}) = 0$ and $H(\hat{x})$ is positive definite, then $\hat{x}$ is a strict local minimum of (P).*

Proof. Since f is twice differentiable at $\hat{x}$, we have for all $x \in \mathbb{R}^n$:

$$f(x) = f(\hat{x}) + \frac{1}{2}(x - \hat{x})^T H(\hat{x})(x - \hat{x}) + \|x - \hat{x}\|^2 \psi(\hat{x}, x - \hat{x}),$$

where $\psi(\hat{x}, x - \hat{x}) \to 0$ as $x \to \hat{x}$ (since $\nabla f(\hat{x}) = 0$).

Suppose $\hat{x}$ is not a strict local minimum. Then there exists a sequence $\{x_k\}_{k \in \mathbb{N}}$ such that $x_k \neq \hat{x}$ for all k, and:

$$x_k \to \hat{x} \text{ as } k \to \infty \text{ and } f(x_k) \leq f(\hat{x}).$$

Let $x = x_k$, divide everything by $\|x-\hat{x}\|^2$, and denote:

$$d_k = \frac{x-\hat{x}}{\|x-\hat{x}\|}, \text{ with } \|d_k\| = 1.$$

We obtain:

$$\frac{f(x_k)-f(\hat{x})}{\|x_k-\hat{x}\|^2} = \frac{1}{2}d_k^T H(\hat{x})d_k + \frac{\psi(\hat{x},x_k-\hat{x})}{\|x-\hat{x}\|^2},$$

where $\psi(\hat{x},x_k-\hat{x}) \to 0$ as $x_k \to \hat{x}$.
Therefore:

$$\frac{1}{2}d_k^T H(\hat{x})d_k + \frac{\psi(\hat{x},x_k-\hat{x})}{\|x_k-\hat{x}\|^2} \leq 0, \text{ for all } k.$$

Since $\psi(\hat{x},x_k-\hat{x}) \to 0$, it follows that:

$$\frac{1}{2}d_k^T H(\hat{x})d_k \leq 0, \text{ for all } k.$$

This implies $H(\hat{x})$ is not positive definite, contradicting the assumption. Hence, $\hat{x}$ must be a strict local minimum. □

2.2.3 Necessary and Sufficient Conditions

Theorem 2.2.5. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex and differentiable function. A necessary and sufficient condition for $\bar{x} \in \mathbb{R}^n$ to be a global minimum of f on $\mathbb{R}^n$ is that:*

$$\nabla f(\bar{x}) = 0.$$

Proof. Proof. 1. **Necessity:**
Suppose $\bar{x}$ is a global minimum of f on $\mathbb{R}^n$. By the definition of a global minimum, for all $x \in \mathbb{R}^n$, we have:

$$f(x) \geq f(\bar{x}).$$

Since f is differentiable, we can examine the behavior of f around $\bar{x}$ using a directional derivative. For any direction $d \in \mathbb{R}^n$, the first-order condition for optimality implies:

$$\nabla f(\bar{x}) \cdot d \geq 0 \quad \text{and} \quad \nabla f(\bar{x}) \cdot (-d) \geq 0.$$

These inequalities imply that $\nabla f(\bar{x}) \cdot d = 0$ for all directions $d \in \mathbb{R}^n$, which is only possible if $\nabla f(\bar{x}) = 0$.
2. **Sufficiency:**
Now, suppose $\nabla f(\bar{x}) = 0$. Since f is convex, we have the property:

$$f(x) \geq f(\bar{x}) + \nabla f(\bar{x}) \cdot (x-\bar{x}).$$

Substituting $\nabla f(\bar{x}) = 0$ into this inequality yields:

$$f(x) \geq f(\bar{x}),$$

which means $f(\bar{x}) \leq f(x)$

□

Example 2.2.1. *Find the extrema of the function analytically:*

$$f(x,y) = x^3 + y^3 + 3xy$$

and determine whether they are minima or not.

Solution

We calculate the Gradient, we have

$$\nabla f(x,y) = \begin{pmatrix} 3x^2 + 3y \\ 3y^2 + 3x \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

This yields:

$$\begin{cases} 3x^2 + 3y = 0 \\ 3y^2 + 3x = 0 \end{cases}$$

which simplifies to:

$$\begin{cases} x^2 + y = 0 \\ y^2 + x = 0 \end{cases}$$

Solving these equations, we find the critical points:

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \text{ or } \begin{pmatrix} -1 \\ -1 \end{pmatrix}$$

The Hessian matrix is:

$$H_f(x,y) = \begin{pmatrix} 6x & 3 \\ 3 & 6y \end{pmatrix}$$

Thus,

$$\nabla^2 f \begin{pmatrix} 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 & 3 \\ 3 & 0 \end{pmatrix}$$

From which we get:

$$\nabla^2 f \begin{pmatrix} 0 \\ 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \cdot \begin{pmatrix} x \\ y \end{pmatrix} = 6xy$$

Thus:

$$\nabla^2 f \begin{pmatrix} 0 \\ 0 \end{pmatrix} \text{ is not positive definite}$$

Therefore:

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix} \text{ is not an extremum.}$$

For the point:

$$\nabla^2 f \begin{pmatrix} -1 \\ -1 \end{pmatrix} = \begin{pmatrix} -6 & 3 \\ 3 & -6 \end{pmatrix}$$

The eigenvalues are the roots of:

$$\lambda^2 + 12\lambda + 27 = 0 \implies \lambda_1 = -9 \text{ and } \lambda_2 = -3$$

Since:

$$\lambda_1, \lambda_2 < 0 \implies \nabla^2 f \begin{pmatrix} -1 \\ -1 \end{pmatrix} \text{ is negative definite.}$$

In this case, the point:

$$\begin{pmatrix} -1 \\ -1 \end{pmatrix}$$

is a local maximum for f and:

$$f \begin{pmatrix} -1 \\ -1 \end{pmatrix} = 1.$$

2.2.4 Exercises

Exercise 1:
Consider the function $g : \mathbb{R}^2 \to \mathbb{R}$ defined by:

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

1. Show that the function f is coercive and strictly convex on $\mathbb{R}^2$.
2. Deduce that f has a unique minimum on $\mathbb{R}^2$.
3. Provide the optimality condition.

Exercise 2:
For each of the following functions:
$f_1(x_1,x_2,x_3) = 3x_1^2 + 3x_2^2 + 3x_3^2 - 2x_1 - 10x_3 - 2x_1x_3$, $f_2(x_1,x_2) = x_1^3 + x_2^3 - 3x_1 - 12x_2 + 20$
$f_3(x_1,x_2) = x_1^4 + x_2^4 - 2(x_1 - x_2)^2$.

1. Study the existence of extrema points.
2. Using a first-order optimality condition, determine the critical points.
3. Specify their nature each time (minimum or maximum? local or global?).

Exercise 3:
We consider the problem $\min_{x \in \mathbb{R}^2} f(x)$, where $f(x) = \frac{1}{2}\langle Ax, x\rangle$ with

$$A = \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}$$

1. Provide the solution to this problem (denoted $\bar{x}$). Let $\alpha > 0$, and let the sequence $(x_k)_{k\in\mathbb{N}}$ be defined by $x_{k+1} = x_k - \alpha A x_k$.
2. For which values of α does the sequence (x_k) converge to $\bar{x}$ for any $x_0 \in \mathbb{R}^2$?

3. What is the optimal step size $\bar{\alpha}$?

Exercise 4:

Determine and specify the nature of the critical points of the following functions:

$$f(x,y) = x^3 + y^3 - 3axy, \quad a \in \mathbb{R}$$
$$g(x,y) = x^2 - \cos(y)$$
$$h(x,y) = y^2 + xy\ln(x)$$

2.2.5 Corrections

Exercise 1:

We consider the function g from $\mathbb{R}^2$ to $\mathbb{R}$, defined by the relation:

$$g(x,y) = x^2 + y^2 + xy$$

1. Show that the function f is coercive.

   Since:

$$f(x,y) = x^2 + y^2 + xy \quad \text{and} \quad \lim_{\|(x,y)\| \to \infty} f(x,y) \to \infty,$$

   then, the function f is coercive.

2. Show that the function f is strictly convex on $\mathbb{R}^2$.

   The Hessian matrix of f at any point $(x,y) \in \mathbb{R}^2$ is:

$$H(x,y) = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$$

   This matrix is positive definite, and thus f is strictly convex.

3. Since f is continuous and coercive, it has at least one minimum on $\mathbb{R}^2$. Moreover, g is strictly convex, so this minimum is unique.

   To find the critical points, we solve the system:

$$\begin{cases} 2x + y = 0 \\ 2y + x = 0 \end{cases}$$

   Thus, $x = y = 0$. Therefore, $(0,0)$ is the unique minimum of f.

Exercise 2

1.

$$f_1(x_1,x_2,x_3) = 3x_1^2 + 3x_2^2 + 3x_3^2 - 2x_1 - 10x_3 - 2x_1x_3$$

$$= 2\left(x_1 - \frac{1}{2}\right)^2 + 2\left(x_3 - \frac{5}{2}\right)^2 + 3x_2^2 - 13$$

It is clear that:

$$\lim_{\|(x_1,x_2,x_3)\|\to+\infty} f_1(x_1,x_2,x_3) = +\infty,$$

which shows that f_1 is a coercive function. Consequently, f admits at least one global minimum.

$$\nabla f_1(x_1,x_2,x_3) = \begin{pmatrix} 6x_1 - 2x_3 - 2 \\ 6x_2 \\ 6x_3 - 2x_1 - 10 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

This gives the critical point:

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix}$$

The Hessian matrix is:

$$\nabla^2 f_1(x_1,x_2,x_3) = \begin{pmatrix} 6 & 0 & -2 \\ 0 & 6 & 0 \\ -2 & 0 & 6 \end{pmatrix}$$

This is a positive definite matrix since its eigenvalues are $\lambda_1 = 4$, $\lambda_2 = 6$, and $\lambda_3 = 8$, all greater than zero. Therefore, the critical point is a global minimum.

2. For the function:

$$f_2(x_1,x_2) = x_1^3 + x_2^3 + 3x_1 - 12x_2 + 20$$

We have:

$$\lim_{x\to+\infty} f_2(x,0) = +\infty, \qquad \lim_{x\to-\infty} f_2(x,0) = -\infty$$

Thus, f is not **coercive**. If a minimum exists, it will not be global.

We compute the gradient,

$$\nabla f_2(x_1,x_2) = \begin{pmatrix} 3x_1^2 - 3 \\ 3x_2^2 - 12 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

This gives the following critical points:

$$\begin{pmatrix} 1 \\ 2 \end{pmatrix}, \quad \begin{pmatrix} 1 \\ -2 \end{pmatrix}, \quad \begin{pmatrix} -1 \\ 2 \end{pmatrix}, \quad \begin{pmatrix} -1 \\ -2 \end{pmatrix}$$

The Hessian matrix is written as follows

$$\nabla^2 f_2(x_1,x_2) = \begin{pmatrix} 6x_1 & 0 \\ 0 & 6x_2 \end{pmatrix}$$

For the critical point $\begin{pmatrix} 1 \\ 2 \end{pmatrix}$, we have:

$$\nabla^2 f_2(1,2) = \begin{pmatrix} 6 & 0 \\ 0 & 12 \end{pmatrix}$$

This is a positive definite matrix, so the point $\begin{pmatrix} 1 \\ 2 \end{pmatrix}$ is a minimizer of f.

For $\nabla^2 f_2(1,-2)$, we have:

$$\nabla^2 f_2(1,-2) = \begin{pmatrix} 6 & 0 \\ 0 & -12 \end{pmatrix}$$

This is an indefinite matrix, so $\begin{pmatrix} 1 \\ -2 \end{pmatrix}$ is a saddle point.

For $\nabla^2 f_2(-1,2)$:

$$\nabla^2 f_2(-1,2) = \begin{pmatrix} -6 & 0 \\ 0 & 12 \end{pmatrix}$$

This is also an indefinite matrix, so $\begin{pmatrix} -1 \\ 2 \end{pmatrix}$ is a saddle point.

For $\nabla^2 f_2(-1,-2)$:

$$\nabla^2 f_2(-1,-2) = \begin{pmatrix} -6 & 0 \\ 0 & -12 \end{pmatrix}$$

This is a negative definite matrix, so $\begin{pmatrix} -1 \\ -2 \end{pmatrix}$ is a maximizer of f.

3. For the function,

$$f_3(x_1,x_2) = x_1^4 + x_2^4 - 2(x_1 - x_2)^2$$

Using the identity:

$$2(x_1^2 + x_2^2) - (x_1 + x_2)^2 = (x_1 - x_2)^2$$

We obtain:

$$x_1^4 + x_2^4 \leq \frac{1}{2}(x_1^2 + x_2^2)^2 \quad (1)$$

$$-(x_1 - x_2)^2 \leq -2(x_1^2 + x_2^2) \quad (2)$$

From (1) and (2), we have:

$$x_1^4 + x_2^4 - 2(x_1 - x_2)^2 \leq \frac{1}{2}\left(-(x_1^2 + x_2^2)^2 - 16\right)$$

Thus:

$$\lim_{\|(x_1,x_2)\|\to+\infty} f(x_1,x_2) = +\infty$$

So, f is a coercive function, and therefore f admits at least one global minimizer.

The critical points are:

$$\begin{pmatrix}0\\0\end{pmatrix}, \quad \begin{pmatrix}\sqrt{2}\\-\sqrt{2}\end{pmatrix}, \quad \begin{pmatrix}-\sqrt{2}\\\sqrt{2}\end{pmatrix}$$

The point $\begin{pmatrix}0\\0\end{pmatrix}$ is a saddle point, but the points:

$$\begin{pmatrix}\sqrt{2}\\-\sqrt{2}\end{pmatrix}, \quad \begin{pmatrix}-\sqrt{2}\\\sqrt{2}\end{pmatrix}$$

are global minima, with:

$$f\left(\begin{pmatrix}\sqrt{2}\\-\sqrt{2}\end{pmatrix}\right) = f\left(\begin{pmatrix}-\sqrt{2}\\\sqrt{2}\end{pmatrix}\right) = -8.$$

Exercise 3:

- The eigenvalues of matrix A are $\lambda_1 = 1$ and $\lambda_2 = 3$. The quadratic form f is strictly convex and coercive, thus it admits a unique minimum, which is the solution of

$$\nabla f(x) = 0 \iff Ax = 0.$$

  Since A is invertible, we obtain $\bar{x} = \begin{pmatrix}0\\0\end{pmatrix}^T$.

- This corresponds to the sequence generated by the fixed-step gradient method, which converges for any value of $\alpha \in \left]0, \frac{2}{3}\right[$.

- The optimal step size for this method is given by:

$$\bar{\alpha} = \frac{2}{\lambda_1 + \lambda_2} = \frac{2}{4} = \frac{1}{2}.$$

Exercise 4:

1. The gradient and Hessian of f are given by:

$$\nabla f(x,y) = \begin{pmatrix}3x^2 - 3ay\\3y^2 - 3ax\end{pmatrix}$$

$$\nabla^2 f(x,y) = \begin{pmatrix}6x & -3a\\-3a & 6y\end{pmatrix}$$

For $a \neq 0$, the function f has two critical points: $p_0 = (0,0)$ and $p_a = (a,a)$.

The Hessian matrix of f at the point p_0 has eigenvalues of different signs, meaning p_0 is a saddle point. The Hessian calculated at the point p_a has two eigenvalues: $\lambda_1 = 3a$ and $\lambda_2 = 9a$. Thus, if $a > 0$, p_a is a minimum, and if $a < 0$, p_a is a maximum.

2. We show the gradient and Hessian of g are given by:

$$\nabla g(x,y) = \begin{pmatrix} 2x \\ \sin(y) \end{pmatrix}$$

$$\nabla^2 g(x,y) = \begin{pmatrix} 2 & 0 \\ 0 & \cos(y) \end{pmatrix}$$

The critical points of g are of the form $p_k = (0, k\pi)$, where $k \in \mathbb{Z}$. The Hessian at the point p_k is:

$$\nabla^2 g(p_k) = \begin{pmatrix} 2 & 0 \\ 0 & (-1)^k \end{pmatrix}$$

Hence, p_k is a minimum if k is even, and a saddle point if k is odd.

3. The function h is defined for $x > 0$ and is of class C^2 on its domain of definition. The gradient and Hessian of h are given by:

$$\nabla h(x,y) = \begin{pmatrix} y\ln(x) + y \\ 2y + x\ln(x) \end{pmatrix}$$

$$\nabla^2 h(x,y) = \begin{pmatrix} \frac{y}{x} & \ln(x) + 1 \\ \ln(x) + 1 & 2 \end{pmatrix}$$

There are two critical points: $(1,0)$ and $\left(\frac{1}{e}, \frac{1}{2e}\right)$. The first is a saddle point, and the second is a minimum.

Chapter 3

Algorithms

In this chapter, we will give several algorithms that allow us to show approximately the solutions of the problem (P). Indeed, we will introduce the most classical and fundamental methods. Most of these algorithms, however, rely on optimality conditions (chapter 2), which we have seen are useful for determining local minima. The challenge of finding global minima is more complex . Nonetheless, in the following section, we will describe a probabilistic algorithm that can help "identify" a global minimum for the problem (P).

It is important to assume **the differentiability of the function f.**

Definition 3.0.1 (Algorithms)**.** *An algorithm is defined by a mapping m from $\mathbb{R}^n$ to $\mathbb{R}^n$, allowing the generation of a sequence of elements in $\mathbb{R}^n$ using the formula:*

$$\begin{cases} x_0 \in \mathbb{R}^n \textit{ is given}, \\ x_{i+1} = m(x_i), \end{cases}$$

with $k = 0$ is initialization step and $i = i + 1$ is the iteration step.

We will begin by the first algorithm which represent the Gradient method

3.0.1 Gradient Method

The Gradient method belongs to a hige class of numerical methods called **descent methods**. The main objective is to minimize a function f. To do this, we start with an arbitrary point x_0. To construct the next iterate x_1, we aim to move closer to the minimum of f; hence, we want $f(x_1) < f(x_0)$. We then seek x_1 in the form $x_1 = x_0 + \rho_1 d_1$, witj d_1 is a non-zero vector in $\mathbb{R}^n$, and ρ_1 is a strictly positive real number. In general way, we obtain d_1 and ρ_1 such that $f(x_0 + \rho_1 d_1) < f(x_0)$. It is not always possible to find d_1. When d_1 exists, it is called a **descent direction**, and ρ_1 is the step size. The direction and the step size can either be fixed or vary at each iteration. The general scheme of a descent method is as follows:

$$\begin{cases} x_0 \in \mathbb{R}^n \text{ is given}, \\ x_{k+1} = x_k + \rho_k d_k, \end{cases}$$

with $d_k \in \mathbb{R}^n \setminus \{0\}$, and $\rho_k \in \mathbb{R}^+$, we choose ρ_k and d_k such that $f(x_k + \rho_k d_k) \leq f(x_k)$.

A natural way to determine a descent direction is to express a second-order Taylor expansion of the function f between two iterates x_k and $x_{k+1} = x_k + \rho_k d_k$:

$$f(x_k + \rho_k d_k) = f(x_k) + \rho_k \langle \nabla f(x_k), d_k \rangle + o(\rho_k d_k).$$

Since we want $f(x_k + \alpha_k d_k) < f(x_k)$, we can choose, as a first approximation, $d_k = -\nabla f(x_k)$. The method obtained this way is called the Gradient algorithm. The step size α_k can either be constant or variable.

3.0.2 Principle of the Method

The iteration of the sequence from x_i to x_{i+1} occurs in two steps:

1. At the point x_i, we choose a direction of descent d_i.
2. We find $\rho_i > 0$ such that $f(x_i + \rho_k d_i) < f(x_i)$, for all $i \in \mathbb{N}$.

Therefore, the principle of a descent method consists of performing the following iterations:

$$x_{i+1} = x_i + \rho_i d_i$$

3.0.3 Convergence Tests

We consider x^* be a local minimum of the objective function f to be optimized. Moreover, a stopping criterion must be chosen to guarantee that the algorithm always halts after a finite number of iterations, and that the last computed point is sufficiently close to x^*.

Let $\varepsilon > 0$ be the required precision. Several criteria are at our disposal:

First, an optimality criterion based on **the first-order necessary conditions for optimality** is tested as follows:

$$\|\nabla f(x_i)\| < \varepsilon,$$

in which case the algorithm stops and returns the current iterate x_k as the solution.

In practice, the optimality test is not always satisfied, and we may need to rely on other criteria (based on numerical experience):

- **Solution stagnation:**

$$\|x_{i+1} - x_i\| < \varepsilon$$

- **Current value stagnation:**

$$|f(x_{i+1}) - f(x_i)| < \varepsilon$$

- **Number of iterations exceeding a predefined threshold:**

$$i = i + 1.$$

The descent direction chosen at each iteration will be:

$$d^{(i)} = -\nabla f(x^{(i)}),$$

thus, the points are successively generated by this method as follows:

$$\begin{cases} x^{(i+1)} = x^{(i)} + \rho_i d^{(i)}, \\ \rho_i > 0, \end{cases}$$

The Gradient algorithm is given by:

1. **Initialization:** Set $k = 0$, choose x_0, $\rho > 0$, and $\varepsilon > 0$.

2. **Iteration** i:
$$x_{i+1} = x_i - \rho_i \nabla f(x_i)$$

3. **Stopping criterion:** Stop if $\|x_{i+1} - x_i\| < \varepsilon$ or $\|\nabla f(x_i)\| < \varepsilon$. Otherwise, set $i = i+1$ and return to step 2.

In this case, ε represents a small positive real number that show the desired precision.

3.0.4 Gradient Methods with Constant Step Size

If instead of using the optimal step size, we take a fixed step size ρ, the algorithm, called the gradient method with constant (or fixed) step size, is simply the algorithm applied to the find for a fixed point for the function $x - \alpha \nabla f(x)$:

$$x_{i+1} = x_i - \rho \nabla f(x_i)$$

with $f \in C^1$, this method converges if ρ is chosen sufficiently small.
The choice of ρ step size:

- A well-chosen step gives results similar to those obtained by the steepest descent.
- A smaller step reduces the zigzags of the iterates but significantly increases the number of iterations.
- A step size that is too large causes the method to diverge.

3.0.5 Optimal Step Gradient Method

This method consists of showing the following iterations:

$$\begin{cases} x_{i+1} = x_i - \rho_i \nabla f(x_i) \\ \alpha_k > 0 \end{cases}$$

with ρ_i is chosen by the minimization rule. It involves selecting, at each iteration i, ρ_i as the optimal solution to the one-dimensional minimization problem of f along the half-line defined by the point x_i and the direction d_i. Therefore, ρ_i is chosen so that:

$$f(x_i + \rho_i d_i) < f(x_i) \text{ for all } \rho > 0$$

In this case, the descent directions d_k satisfy:

$$\nabla f(x_i + \rho_i d_i) \cdot d_i = 0$$

because if we introduce the function:

$$g(\rho) = f(x_i + \rho d_i)$$

we have:

$$g'(\rho) = \nabla f(x_i + \rho d_i)^t \cdot d_i$$

and since g is differentiable, we necessarily have $g'(\rho_i) = 0$, which implies:

$$\nabla f(x_i + \rho_i d_i)^t \cdot d_i = 0$$

This computation of determining the step size is called **line search**.

Example 3.0.1. *Consider the following quadratic function:*

$$f(x) = \frac{1}{2} x^T A x - b^T x$$

where $A > 0$ (i.e., A is a positive definite matrix). Let ρ_k^* denote the optimal step size, characterized by $g'(\rho_k) = 0$. We then have:

$$\nabla g(\rho_k) = \nabla f(x_k + \rho_k d_k) = A(x_k + \rho_k d_k) - b$$

To find ρ_k^*, we solve:

$$\nabla f(x_k + \rho_k d_k) \cdot d_k = (A(x_k + \rho_k d_k) - b) \cdot d_k = (Ax_k - b) \cdot d_k + \rho_k (d_k^T A d_k)$$

Setting this to zero gives:

$$(Ax_k - b) \cdot d_k + \rho_k (d_k^T A d_k) = 0$$

Solving for ρ_k, we obtain:

$$\rho_k = \frac{-(Ax_k - b) \cdot d_k}{d_k^T A d_k}$$

where d_k is a descent direction and $\rho_k > 0$.

Theorem 3.0.1 (Convergence Theorem)**.** *Let f be a C^1 (continuously differentiable) function from $\mathbb{R}^n$ to $\mathbb{R}$, which is coercive and strictly convex. Suppose there exists a constant $M > 0$ such that, for all $x, y \in \mathbb{R}^n$,*

$$\|\nabla f(x) - \nabla f(y)\| \leq M \|x - y\|. \tag{3.1}$$

Then, if the step size α_k is chosen in an interval $[\alpha_1, \alpha_2]$, where $0 < \alpha_1 < \alpha_2 < \frac{2}{M}$, the gradient descent method converges to the unique minimum of f.

Proof. Since f is strictly convex, it admits a **unique minimum** x^* in $\mathbb{R}^n$, characterized by $\nabla f(x^*) = 0$. We aim to show that the sequence x_k, generated by the gradient descent algorithm, converges to x^*.

Using the Taylor expansion of f, applied to $x = x_k$ and $y = x_{k+1}$, we write:

$$f(x_{k+1}) = f(x_k) + \langle \nabla f(x_k), x_{k+1} - x_k \rangle + \int_0^1 \langle \nabla f(x_k + t(x_{k+1} - x_k)) - \nabla f(x_k), x_{k+1} - x_k \rangle \, dt.$$

Substituting $x_{k+1} = x_k - \alpha_k \nabla f(x_k)$ into the above, we obtain:

$$f(x_{k+1}) - f(x_k) \leq -\frac{1}{\alpha_k} \|x_{k+1} - x_k\|^2 + \int_0^1 \|\nabla f(x_k + t(x_{k+1} - x_k)) - \nabla f(x_k)\| \cdot \|x_{k+1} - x_k\| \, dt.$$

From the Lipschitz condition on ∇f (Equation (3.1)), it follows:

$$\|\nabla f(x_k + t(x_{k+1} - x_k)) - \nabla f(x_k)\| \leq M\|x_{k+1} - x_k\|.$$

Thus,

$$f(x_{k+1}) - f(x_k) \leq -\frac{1}{\alpha_k}\|x_{k+1} - x_k\|^2 + \frac{M}{2}\|x_{k+1} - x_k\|^2.$$

Rewriting this, we have:

$$f(x_{k+1}) - f(x_k) \leq \left(\frac{M}{2} - \frac{1}{\alpha_k}\right)\|x_{k+1} - x_k\|^2.$$

Since α_k is chosen in $[\alpha_1, \alpha_2]$ with $0 < \alpha_1 < \alpha_2 < \frac{2}{M}$, it follows that $\frac{M}{2} - \frac{1}{\alpha_k} < 0$. Hence,

$$f(x_{k+1}) - f(x_k) < 0.$$

This shows that $f(x_k)$ is strictly decreasing.

Since $f(x_k)$ is coercive and strictly decreasing, it is also bounded below. Hence, $f(x_k)$ converges to some limit.

Then, the difference $f(x_{k+1}) - f(x_k)$ tends to 0. By coercivity, $\|x_k\|$ is bounded.

Relation Between x_{k+1} and x_k:** Using the inequality:

$$\|x_{k+1} - x_k\|^2 \leq \frac{1}{\alpha_2 - \frac{M}{2}}[f(x_k) - f(x_{k+1})],$$

and since $f(x_{k+1}) - f(x_k) \to 0$, it follows that $\|x_{k+1} - x_k\| \to 0$.

Behavior of $\nabla f(x_k)$: Since $x_{k+1} = x_k - \alpha_k \nabla f(x_k)$, we have:

$$\nabla f(x_k) = \frac{x_k - x_{k+1}}{\alpha_k}.$$

As $\|x_{k+1} - x_k\| \to 0$, it follows that $\nabla f(x_k) \to 0$.

6. **Continuity of ∇f:** By continuity of ∇f, we deduce $\nabla f(x) = 0$, where x is the unique minimum x^*.

since f is strictly convex, this holds for every limit point of the sequence x_k, proving that the entire sequence x_k converges to x^*.

□

3.1 Conjugate Gradient Method

In this section, we will describe Conjugate gradient methods. But before delving into them, we will first outline the general principle of a conjugate direction method.Conjugate gradient methods are used to solve nonlinear optimization problems, and they are also employed to solve large linear systems. These methods rely on the concept of conjugate directions, where successive gradients are orthogonal to each other and to previous directions. The conjugate gradient method is an optimal step descent method that allows for minimizing a quadratic function from $\mathbb{R}^n$ to $\mathbb{R}$ in at most n iterations.

The initial idea was to find a sequence of descent directions that allows solving the problem:

$$\min_{x \in \mathbb{R}^n} f(x) \quad \text{(P)}$$

where f is continuously differentiable function.
We will define conjugate vectors in the following definition.

Definition 3.1.1. *Let A be a symmetric positive definite matrix of size $n \times n$. Two vectors x and y in $\mathbb{R}^n$ are said to be A-conjugate (or conjugate with respect to A) if they satisfy:*

$$x^T A y = 0$$

We Consider the following quadratic unconstrained problem:

$$\min_{x \in \mathbb{R}^n} f(x) = \frac{1}{2} x^T A x + b^T x + c$$

where $A \in \mathbb{R}^{n \times n}$ is a symmetric positive definite matrix, $b \in \mathbb{R}^n$, and $c \in \mathbb{R}$.

3.1.1 The Principle of the method

The method involves minimizing f starting from a point x_0, following n directions: $d_0, d_1, \ldots, d_{n-1}$, which are mutually conjugate with respect to A.

- The first fundamental idea of the conjugate gradient algorithm is to choose each descent direction to be conjugate to the previous descent direction with respect to A.
- The second fundamental idea is to search for d_i as a linear combination of d_{i-1} and the gradient at x_i, namely:

  $$d_i = -\nabla f(x_i) + \beta_{i-1} d_{i-1}$$

  where β_{i-1} is chosen so that the successive directions are conjugate with respect to A.
- We construct the sequence:

  $$x_{i+1} = x_i + \rho_i d_i,$$

  where

  $$\rho_i \in \arg\min_{\alpha > 0} f(x_i + \rho d_i)$$

Calculation of the Step Size:
Since ρ_i minimizes f in the direction d_i, we have:

$$\forall k, \quad d_i^t \nabla f(x_{i+1}) = 0$$

which implies:

$$d_i^t (A(x_i + \rho_i d_i) + b) = 0$$

then,

$$\rho_i = -\frac{d_i^t (A x_i + b)}{d_k^t A d_k} = -\frac{d_k^t \nabla f(x_i)}{d_k^t A d_i}$$

(As A is positive definite and d_i are mutually conjugate, we have $d_i^t A d_i \neq 0$ for all i.)
Calculation of β_i:
We have that

$$d_{i+1} = -\nabla f(x_{i+1}) + \beta_i d_i$$

it follows that:

$$d_i^t A d_{i+1} = 0$$

which implies that

$$-\nabla f(x_{i+1})^t A d_i + \beta_k d_i^t A d_i = 0$$

Since $d_i \neq 0$, we obtain:

$$\beta_i = \frac{\nabla f(x_{i+1})^t A d_i}{d_k^t A d_i}$$

Lemma 3.1.1. *For all $i < k$, we have $d_i^T \nabla f(x_i) = 0$.*

Proof.

$$\begin{aligned} d_i^T \nabla f(x_k) = d_i^T (Ax_k + b) \\ &= d_i^t (A(x_i + \sum_{j=i}^{k-1} \beta_j d_j) + b) \\ &= d_i^t (Ax_i + b) + \beta_i d_i^t A d_i \\ &= d_i^t (Ax_i + b) - \frac{d_i^T (Ax_i + b)}{d_i^t A d_i} d_i^T A d_i \quad = 0. \end{aligned}$$

□

Proposition 3.1.1 ((Fletcher-Reeves)). *The real number β_k is calculated using the following formula*

$$\beta_k = \frac{\|\nabla f(x_{k+1})\|^2}{\|\nabla f(x_k)\|^2}$$

Proof. We have:

$$\nabla f(x_{i+1}) - \nabla f(x_i) = \rho_i A d_i$$

Thus:

$$(\nabla f(x_{i+1}))^t A d_i = \frac{1}{\rho_i} (\nabla f(x_{i+1}))^t (\nabla f(x_{i+1}) - \nabla f(x_i))$$

Since:

$$\nabla f(x_i) = -d_i + \beta_{i-1} d_{i-1}$$

By Lemma (3.1.1), we obtain

$$(\nabla f(x_{i+1}))^t (\nabla f(x_i)) = 0$$

Therefore, we get

$$\beta_i = \frac{(\nabla f(x_{k+1}))^t A d_i}{d_i^t A d_i} = \frac{1}{\rho_i} \cdot \frac{(\nabla f(x_{i+1}))^t \nabla f(x_{k+1})}{d_i^t A d_i},$$

substituting ρ_i yields that

$$\beta_i = -\frac{\|\nabla f(x_{i+1})\|^2}{d_i^t \nabla f(x_i)},$$

but we know that

$$(\nabla f(x_i))^t d_i = (\nabla f(x_i))^t (-\nabla f(x_i) + \beta_{i-1} d_{i-1}),$$

then, we conclude that

$$\beta_i = \frac{\|\nabla f(x_{i+1})\|^2}{\|\nabla f(x_i)\|^2}.$$

3.1.2 Algorithm

1. We Choose x_0 and set $i = 0$, $d_0 = -\nabla f(x_0)$. 2. While $\|\nabla f(x_i)\| > \varepsilon$, do :

- $\rho_i = -\frac{d_k^t \nabla f(x_i)}{d_i^t A d_i}$;
- $x_{i+1} = x_i + \rho_i d_i$;
- $\beta_i = \frac{\|\nabla f(x_{i+1})\|^2}{\|\nabla f(x_i)\|^2}$;
- $d_{i+1} = -\nabla f(x_{i+1}) + \beta_i d_i$.

3.1.3 The Advantages of the Conjugate Gradient Method for Quadratic Problems

- Minimal Memory Consumption: The algorithm requires only a minimal amount of memory; we need to store four vectors: x_i, $\nabla f(x_i$, d_i, and Ad_ialong with the scalars ρ_i and β_{i+1}).
- Efficiency for Large Sparse Systems: The conjugate gradient method is particularly useful for solving large sparse systems, as it is sufficient to know how to apply the matrix A to a vector.
- Rapid Convergence: Convergence can be quite fast.

3.1.4 Different Formulas for β_{i+1} in the Quadratic function

The various values attributed to β_k define the different forms of conjugate gradient methods. If we denote $y_{k-1} = \nabla f(x_i) - \nabla f(x_{i-1}$ and $s_k = x_{k+1} - x_k$, we have the following formulas:

- **Hestenes-Stiefel (HS) Conjugate Gradient:**

$$\beta_i^{\mathrm{HS}} = \frac{\nabla f(x_{i+1})^t y_k}{d_k^t y_k}.$$

- **Fletcher-Reeves (FR) Conjugate Gradient:**

$$\beta_i^{\text{FR}} = \frac{\|\nabla f(x_{i+1})\|^2}{\|\nabla f(x_i)\|^2}.$$

- **Liu-Storey (LS) Conjugate Gradient:**

$$\beta_i^{\text{LS}} = -\frac{{}^T y_k}{d_k^T g_k}.$$

- **Dai-Yuan (DY) Conjugate Gradient:**

$$\beta_i^{\text{DY}} = \frac{\|\nabla f(x_i\|^2}{d_k^T \nabla f(x_{i-1}}.$$

3.2 Newton's method

Definition 3.2.1. *A Newtonian method refers to any descent algorithm where the descent direction d_i at each iteration . The direction d_i defined this way is called the Newton direction.*

We set that

$$d_i = -H_i^{-1} \nabla f(x_i),$$

where $H_i = \nabla^2 f(x_i)$. It is clear that d_i is indeed a descent direction:

$$d_i^T \nabla f(x_i) = -\nabla f(x_i)^t H_i^{-1} \nabla f(x_i) < 0,$$

so:

$$x_{i+1} = x_i + d_i,$$

in other words, $\alpha_i = 1$ for all i.

□

3.2.1 Algorithm

- Initial Step: Let $\varepsilon > 0$ be the stopping criterion. Choose an initial point x_1, set $i = 1$, and go to the main step.
- Main Step: If $\|\nabla f(x_i)\| \leq \varepsilon$, stop. Otherwise, update:

$$x_{i+1} = x_i - [H(x_i)]^{-1} \nabla f(x_i)$$

Replace i with $i+1$ and return to the main step.

3.3 Exercises

Exercise 1:

Let $f(x)$ be a continuously differentiable (C^1) function on $\mathbb{R}^n$. It is known that in the neighborhood of a point $a \in \mathbb{R}^n$, f decreases most rapidly in the direction of the negative gradient of f at a, i.e., the direction $-\nabla f(a)$.

We begin with an estimate x_0 for a local minimum of f and consider the sequence $(x_0, x_1, \cdots)$, where for all $i \in \mathbb{N}$, we have:

$$x_{i+1} = x_i - \rho \nabla f(x_i)$$

such that:

$$f(x_0) \geq f(x_1) \geq f(x_2) \geq \dots$$

The function $f(x)$ is defined as follows:

$$f(x) = \frac{1}{2}(x_1 - 1)^2 + \frac{1}{6}(x_2 - 2)^2$$

- What is the unique global minimum $\hat{x}$ of f?
- Starting with $x_0 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ and $\rho = 0.1$, calculate the next two iterates x_1 and x_2.
- Find the maximum step size ρ such that the method converges to $\hat{x}$ regardless of the starting point.

Exercise 2:

We define the function $J : \mathbb{R}^n \to \mathbb{R}$ by:

$$J(x) := \frac{1}{2}\langle Ax, x\rangle - \langle b, x\rangle,$$

where $A \in M_n(\mathbb{R})$ is a symmetric positive definite matrix and $b \in \mathbb{R}^n$. We consider the gradient method with optimal step size for minimizing J.

- Give the expression for x_{k+1} obtained by this method. - Show that the step size θ_k is written as:

$$\theta_k = \frac{||d_k||^2}{\langle Ad_k, d_k\rangle}.$$

- Recall that the minimization problem of J has a unique solution $\bar{x} \in \mathbb{R}^n$, characterized as the unique solution of the Euler equation $\nabla J(x) = 0$. Show that:

$$\langle A^{-1}d_k, d_k\rangle = 2(J(x_k) - J(\bar{x})).$$

Exercise 3:

Consider the following mathematical program:

$$(P) \quad \begin{array}{l} \min f(x) = \frac{1}{2}x^T Ax - b^T x, \\ \text{where} \quad x \in \mathbb{R}^2, \end{array}$$

where:

$$A = \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix}, \quad b = \begin{pmatrix} -1 \\ 1 \end{pmatrix}.$$

- Solve (P) using the conjugate gradient algorithm with a precision of $\varepsilon = 10^{-6}$, starting from the initial point $x_0 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$.
- Deduce the number of iterations.

3.4 Corrections

Exercise 1

Let $f(x)$ be a C^1 function on $\mathbb{R}^n$. It is known that in the neighborhood of a point $a \in \mathbb{R}^n$, f decreases most rapidly in the direction of the negative gradient of f at a, that is, in the direction $-\nabla f(a)$.

We start with an estimate x_0 for a local minimum of f and consider the sequence $(x_0, x_1, \dots)$ where for each $i \in \mathbb{N}$, we have:

$$x_{i+1} = x_i - \rho \nabla f(x_i)$$

such that $f(x_0) \geq f(x_1) \geq f(x_2) \geq \dots$, and the function $f(x)$ is defined as:

$$f(x) = \left(x - \begin{pmatrix} 1 \\ 1 \end{pmatrix}\right)^T \begin{pmatrix} 6 & 2 \\ 2 & 6 \end{pmatrix} \left(x - \begin{pmatrix} 1 \\ 1 \end{pmatrix}\right).$$

What is the unique (global) minimum $\hat{x}$ of f?

The function f is quadratic and can be written as $f(x) = \frac{1}{2}x^T A x - b^T x$, where:

$$A = \begin{pmatrix} 6 & 2 \\ 2 & 6 \end{pmatrix}, \quad b = \begin{pmatrix} 4 \\ 4 \end{pmatrix}.$$

Thus, f is continuous, and moreover, it is coercive and strictly convex (its Hessian matrix is positive definite since its eigenvalues, 4 and 8, are strictly positive). Therefore, f has a unique global minimum $\hat{x}$, characterized by:

$$\hat{x} = \begin{pmatrix} 0.25 \\ 0.25 \end{pmatrix},$$

where:

$$\nabla f(\hat{x}) = 0 \quad \text{and} \quad \hat{x} = A^{-1} b.$$

Using the fixed-step gradient method to compute the next two iterations x_1 and x_2. Starting with $x_0 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ and $\rho = 0.1$,

We have:

$$x_{i+1} = x_i - \rho \nabla f(x_i)$$

with:

$$\nabla f(x) = \begin{pmatrix} 6 & 2 \\ 2 & 6 \end{pmatrix} \left(x - \begin{pmatrix} 1 \\ 1 \end{pmatrix}\right) - \begin{pmatrix} 4 \\ 4 \end{pmatrix}$$

Thus:

$$\nabla f(x) = \begin{pmatrix} 6 & 2 \\ 2 & 6 \end{pmatrix} \left(x - \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right) - \begin{pmatrix} 4 \\ 4 \end{pmatrix}$$

For $x_0 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$, we compute:

$$\nabla f(x_0) = \begin{pmatrix} 6 & 2 \\ 2 & 6 \end{pmatrix} \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix} - \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right) - \begin{pmatrix} 4 \\ 4 \end{pmatrix} = \begin{pmatrix} -8 \\ -8 \end{pmatrix}$$

Then:

$$x_1 = x_0 - \rho \nabla f(x_0) = \begin{pmatrix} 0 \\ 0 \end{pmatrix} - 0.1 \cdot \begin{pmatrix} -8 \\ -8 \end{pmatrix} = \begin{pmatrix} 0.8 \\ 0.8 \end{pmatrix}$$

Now for $x_1 = \begin{pmatrix} 0.8 \\ 0.8 \end{pmatrix}$:

$$\nabla f(x_1) = \begin{pmatrix} 6 & 2 \\ 2 & 6 \end{pmatrix} \left(\begin{pmatrix} 0.8 \\ 0.8 \end{pmatrix} - \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right) - \begin{pmatrix} 4 \\ 4 \end{pmatrix} = \begin{pmatrix} -0.8 \\ -0.8 \end{pmatrix}$$

Then:

$$x_2 = x_1 - \rho \nabla f(x_1) = \begin{pmatrix} 0.8 \\ 0.8 \end{pmatrix} - 0.1 \cdot \begin{pmatrix} -0.8 \\ -0.8 \end{pmatrix} = \begin{pmatrix} 0.88 \\ 0.88 \end{pmatrix}$$

Find the maximum step size ρ for which the method converges to $\hat{x}$ for any initial point x_0.

we have that

$$0 < \rho < \frac{2}{\lambda_{\max}}$$

where $\lambda_{\max}$ is the largest eigenvalue of A. In this case, $\lambda_{\max} = 8$, so:

$$0 < \rho < \frac{2}{8} = 0.25$$

Thus, the step size ρ should be in the interval $(0, 0.25)$.

Exercise 2:

The algorithm of the gradient method with optimal step size is:

$$\begin{cases} x_{k+1} = x_k + \theta_k d_k \\ d_k = -Ax_k + b \\ \theta_k \text{ is such that } J(x_k + \theta_k d_k) = \min_{\theta \in \mathbb{R}} J(x_k + \theta d_k) \end{cases}$$

- In the optimal step size gradient algorithm, two successive directions are orthogonal:

$$\langle d_{k+1}, d_k \rangle = 0$$

Using the formulas for x_{k+1} and d_k, we get:

$$\begin{aligned} 0 &= \langle -Ax_{k+1} + b, d_k \rangle \\ &= \langle -A(x_k + \theta_k d_k) + b, d_k \rangle \\ &= \langle -Ax_k + b - \theta_k A d_k, d_k \rangle \\ &= \langle d_k - \theta_k A d_k, d_k \rangle \\ &= \|d_k\|^2 - \theta_k \langle A d_k, d_k \rangle \end{aligned}$$

Assuming $d_k \neq 0$ and thus $\langle Ad_k, d_k\rangle \neq 0$ (since A is positive definite), we get:

$$\theta_k = \frac{\|d_k\|^2}{\langle Ad_k, d_k\rangle}$$

- The solution to the minimization problem of J is $\bar{x} = A^{-1}b$. A simple calculation shows that:

$$\begin{aligned}
\langle A^{-1}d_k, d_k\rangle &= \langle A^{-1}(Ax_k - b), Ax_k - b\rangle \\
&= \langle x_k - A^{-1}b, Ax_k - b\rangle \\
&= \langle Ax_k, x_k\rangle - \langle b, x_k\rangle - \langle A^{-1}b, Ax_k\rangle + \langle A^{-1}b, b\rangle \\
&= 2\left(\frac{1}{2}\langle Ax_k, x_k\rangle - \langle b, x_k\rangle + \frac{1}{2}\langle A^{-1}b, b\rangle\right) \\
&= 2\left(J(x_k) - J(\bar{x})\right)
\end{aligned}$$

since

$$J(\bar{x}) = \frac{1}{2}\langle A\bar{x}, \bar{x}\rangle - \langle b, \bar{x}\rangle = \frac{1}{2}\langle b, \bar{x}\rangle - \langle b, \bar{x}\rangle = -\frac{1}{2}\langle b, \bar{x}\rangle = -\frac{1}{2}\langle b, A^{-1}b\rangle$$

Exercise 3:

- We consider that $d_0 = -\nabla f(x_0) = \begin{pmatrix} 0 \\ 0 \\ -2 \\ 1 \end{pmatrix}$ with $\|\nabla f(x_0)\| = 2$. Calculate the step size:

$$\alpha_0 = \frac{-d_0^T \nabla f(x_0)}{d_0^T Q d_0} = \frac{1}{2}$$

  Therefore,

$$x_1 = x_0 + \alpha_0 d_0 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

  and

$$\phi_0 = \frac{\|\nabla f(x_1)\|^2}{\|\nabla f(x_0)\|^2} = \frac{1}{4}$$

  Thus,

$$d_1 = -\nabla f(x_1) + \frac{1}{4}d_0 = \begin{pmatrix} -1 \\ -\frac{1}{2} \\ 1 \end{pmatrix}$$

- $$\|\nabla f(x_1)\| = 1 > \varepsilon, \quad \beta_1 = 2, \quad x_2 = \begin{pmatrix} -1 \\ 0 \end{pmatrix}, \quad \lambda_1 = 0, \quad d_2 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix},$$

  so:

$$\|\nabla f(x_2)\| = 0 < \varepsilon,$$

  hence the solution to the problem (P) is

$$x^* = x_2 = \begin{pmatrix} -1 \\ 0 \end{pmatrix}.$$

- The number of iterations is $k = 2$.

3.5 Conclusion

In this document, we provide a comprehensive course on unconstrained optimization and introduce the essential tools needed to understand the core concepts of optimization. Indeed, optimization plays a crucial role in solving minimization and maximization problems, particularly in economics, where reducing economic losses and maximizing profits is paramount.

The applications of optimization are incredibly diverse, spanning various fields such as route planning, object shape optimization, pricing strategies, chemical reaction optimization, air traffic control, device efficiency, engine performance, railway management, investment selection, and shipbuilding. Optimizing these systems allows for the discovery of ideal configurations, leading to gains in effort, time, cost, energy, raw materials, and even customer satisfaction.

Bibliography

[1] M.Belloufi, *Cours Optimisation sans contraintes* , Cours et exercices corrigées, Université Mouhamed Chérif Mesaadia-Souk Ahras- , 2015.

[2] A.Berhail, *Optimisation sans contraintes,* Cours et exercices corrigées, Université 08 Mai Guelma, 2016.

[3] N.Boudiaf, *Optimisation sans contraintes aspect théorique et alghorithmique,* Cours et exercices corrigées, université de Batna 2, 2017.

[4] J.M. Borwein and A.S. Lewis. *Convex Analysis and Nonlinear Optimization*. Springer-Verlag, New York, 2000.

[5] F. Bonans. *Continuous Optimization*. Dunod, Paris, 2006.

[6] G. Cohen. *Convexity and Optimization*. École nationale des ponts et chaussées, 2000.

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

[8] A. Xavier, P. Dreyfuss, and Y. Privat. *Introduction to Optimization: Theoretical, Numerical, and Algorithmic Aspects*. http://math.unice.fr/~dreyfuss/D4, (2006-2007).